e-ISSN: 2602-3563

ACTA INFOLOGICA (ACIN)

JUNE, 2023
Volume: 7 | Issue: 1

ISTANBUL UNIVERSITY, INFORMATICS DEPARTMENT

acin.istanbul.edu.tr
https://dergipark.org.tr/tr/pub/acin

İSTANBUL UNIVERSITY PRESS

Indexing and Abstracting
TÜBİTAK-ULAKBİM TR Dizin
EBSCO Applied Sciences Source Ultimate
Erih Plus
DOAJ
Bielefeld Academic Search Engine (BASE)
OpenAIRE
ResearchBib
ASOS Index

Acta INFOLOGICA (ACIN) | ISSN: 2602-3563

June 2023
Volume: 7 | Issue: 1

Owner
Prof. Dr. Sevinç GÜLSEÇEN
Istanbul University, Informatics Department, Istanbul, Turkiye

Responsible Manager
Prof. Dr. Sevinç GÜLSEÇEN
Istanbul University, Informatics Department, Istanbul, Turkiye

Correspondence Address
Istanbul University, Informatics Department
Kalenderhane Mahallesi, 16 Mart Şehitleri Caddesi, No: 8
Vezneciler, Fatih, İstanbul, Türkiye
Phone: +90 212 440 00 00/10037
E-mail: acin@istanbul.edu.tr
http://iupress.istanbul.edu.tr/tr/journal/acin/home
https://dergipark.org.tr/tr/pub/acin

Publisher
Istanbul University Press
İstanbul Üniversitesi Merkez Kampüsü, 34452 Beyazıt,
Fatih, İstanbul, Türkiye
Phone: +90 212 440 00 00

The publication language of the journal is English

This is a scholarly, international, peer-reviewed and open-access journal
published biannually in June and December.

Publication Type
Periodical

Acta INFOLOGICA (ACIN) | ISSN: 2602-3563
June 2023
Volume: 7 | Issue: 1

EDITORIAL MANAGEMENT BOARD

Acta INFOLOGICA (ACIN) | ISSN: 2602-3563
June 2023
Volume: 7 | Issue: 1

EDITORIAL BOARD

Acta INFOLOGICA (ACIN) | ISSN: 2602-3563
June 2023
Volume: 7 | Issue: 1

CONTENTS

Research Articles / Araştırma Makaleleri

Acta INFOLOGICA (ACIN) | ISSN: 2602-3563
June 2023
Volume: 7 | Issue: 1

CONTENTS

ACTA INFOLOGICA

ACTA INFOLOGICA 2023;7(1):1-16
dergipark.org.tr/acin

İSTANBUL UNIVERSITY PRESS

DOI: 10.26650/acin.1134979

RESEARCH ARTICLE

A Novel Data Encryption Algorithm To Ensure Database Security

Veritabanı Güvenliğini Sağlamak için Yeni Bir Veri Şifreleme Algoritması

Sivan Ibrahim[1], Ahmet Zengin[2], Selman Hızal[3], A. F. M. Suaib Akhter[4], Cevat Altunkaya[5]

[1]Sakarya University, Faculty of Computer and Information Sciences, Department of Computer Engineering, Sakarya, Turkiye

[2](Prof. Dr.), Sakarya University, Faculty of Computer and Information Sciences, Department of Computer Engineering, Sakarya, Turkiye

[3](Assist. Prof.), Sakarya University of Applied Sciences, Department of Computer Engineering, Sakarya, Turkiye

[4](Dr.), Sakarya University of Applied Sciences, Department of Computer Engineering, Sakarya, Turkiye

[5](Dr.), Sakarya University, Faculty of Computer and Information Sciences, Department of Computer Engineering, Sakarya, Turkiye

ORCID: S.S.I. 0000-0003-4707-0578;
A.Z. 0000-0003-0384-4148;
S.H. 0000-0001-6345-0066;
A.F.M.S.A. 0000-0002-2675-1684;
C.A. 0000-0001-5118-2711

Corresponding author:
Selman HIZAL
Sakarya University of Applied Sciences, Department of Computer Engineering, Sakarya, Turkiye
E-mail address: selmanhizal@subu.edu.tr

Submitted: 24.06.2022
Revision Requested: 26.10.2022
Last Revision Received: 10.02.2023
Accepted: 11.02.2023
Published Online: 09.05.2023

Citation: Ibrahim, S.S., Zengin, A., Hizal, S., Suaib Akhter, A.F.M., Altunkaya, C. (2023). A novel data encryption algorithm to ensure database security. *Acta Infologica, 7*(1), 1-16. https://doi.org/10.26650/acin.1134979

ABSTRACT

Many people apply technologies gradually, moving towards new progress and development. Moreover, many of us utilize database systems to assist our work management; the database contains our work or personal information, which increases the risk of losing our data due to disruptive electronic attacks. As a result, protecting databases from electronic attacks and data seizures is crucial. One method of identifying data is through several algorithms so that no benefit is taken from our data during electronic attacks. In this paper, we explain a formula we created for data encryption. The data is encrypted using ASCII Code. Also, we used three keys in the main formula. Because of that formula, each data we save in the database will be encrypted. The data can be Text or numbers. And through using another coordinator with the three previous keys, we can render the data to our original style. The formula focuses on data size and recording speed so that we get the same data size as when the data is encrypted at a reasonable speed.

Keywords: Cryptography, encryption, decryption, cyber, security, database, ASCII code

ÖZ

Günümüzde, özellikle internet üzerinde birçok bulut tabanlı uygulama geliştirilmekte ve insanların kullanımına sunulmaktadır. Bu uygulamaların çoğu veri tabanı sistemlerini kullanır ve özellikle çalışma hayatımızdaki iş veya kişisel bilgilerimizi içerebilir. Kritik bilgilerin bu veri depolarında tutulması yıkıcı elektronik saldırılar nedeniyle veri kaybı riskini artırır. Sonuç olarak, veri tabanlarını elektronik saldırılardan ve veri hırsızlığından korumak oldukça önemlidir ve veri güvenliği sistemlerinin kullanılmasını zorunlu hale getirir. Bu kapsamda çeşitli güvenlik algoritmalarına ihtiyaç vardır. Bu makalede, verilerin şifrelenmesi için yeni bir algoritma önerilmiştir. Üç anahtar ile çalışan algoritmada veriler ASCII Kodu kullanılarak şifrelenir. Çeşitli formüller sayesinde veri tabanına kaydettiğimiz metin ve rakam formundaki her veri şifrelenecektir. Geliştirilen algoritma, verilerin yüksek bir hızda şifrelenmesiyle kayıpsız aynı veriyi tekrar elde etmemizi sağlar. Yapılan testler algoritmanın yüksek veri boyutunda oldukça hızlı çalıştığını ortaya koymuştur.

Anahtar Kelimeler: Kriptografi, şifreleme, şifre çözme, siber, güvenlik, veri tabanı

1. INTRODUCTION

A database is a collection of data that is organized and stored in a structured way so that it can be accessed and modified efficiently. Databases store and manage large amounts of data in a systematic and organized manner. They are an essential component of many modern information systems, allowing organizations to store and retrieve data quickly and accurately (Silberschatz et al., 2002).

Many databases exist, including relational, object-oriented, and NoSQL databases. They can be used for many purposes, such as storing customer information, managing inventory, and tracking financial transactions.

The importance of databases lies in their ability to store and manage large amounts of data in an organized, efficient, and reliable way. They enable organizations to store, retrieve, and manipulate data quickly and accurately, which is essential for various business and scientific applications. Additionally, databases allow for data integration from different sources and enable data sharing and collaboration among other users and systems.

Database security refers to the various measures taken to protect databases from unauthorized access, misuse, or damage. It is an essential aspect of information security that ensures that sensitive data stored in databases is kept safe and secure. Overall, database security is vital to protecting data integrity and confidentiality and helps organizations meet legal and regulatory requirements for data protection (Bertino et al., 2005).

Several articles have been written in the literature. For instance, the authors of (Shelly et al., 2013) concluded that Databases were a preferred target for hackers because of their sensitive and valuable information. A database can be hacked in a variety of ways. A database should be protected from various types of threats and risks. This paper identifies solutions to most attacks, although some are beneficial while others are only momentary. In (Aamer et al., 2005), the writer measured cipher algorithms (AES, DES, 3-DES, and Blowfish) to various data dimensions and encryption periods for two separate devices, a Pentium-4, 2.4 GHz, and a Pentium-II 266 MHz in EBC and CFB Mode. The author decided Blowfish is the quickest, afterward came DES and Triple DES; and CFB needed longer than ECB cipher block mode. Mathur (2012) proposed an encryption algorithm that depended on the ASCII value of the message to be encoded. This algorithm required a key that had the same length as the letter. This key was encrypted and utilized in the encryption and decryption processes. This device required the user to input the key. If the message was longer than the device allowed, the recipient was asked to input a key identical to the range of the letter. This made it difficult for the user to input a large key. Another disadvantage of this algorithm was that it took longer to execute. So, these are two deficiencies of the current algorithm. On the other hand, Nadeem (2006) explored well-known unique vital algorithms, for example, DES, 3DES, AES, and Blowfish. These algorithms were tested, and their efficiency was measured by encrypting input folders of various points and dimensions. Compared to other algorithms, the results revealed that Blowfish performed exceptionally well. It was also discovered that AES outperformed 3DES and DES.

Despite the advancements in database security algorithms, some deficiencies still need to be addressed. Some of the most notable ones include insufficient encryption, lack of access control, lack of data audit trails, and human error. Overall, while current database security algorithms have made significant progress in protecting sensitive data, there is still room for improvement to ensure that databases are more secure and better able to withstand attacks.

This paper's primary goal is to develop a new high-performance and secure data encryption algorithm to overcome the abovementioned issues. In our algorithm, several performance characteristics were considered, such as keys' strength, data size, and the speed of encrypting and decrypting data. In developing the encryption approach, we created the formula composed of four other sub-formulas and three keys so that each is encrypted through those sub-formulas. Then these are all used in the leading formula with an ASCII code to strengthen the protected side of the formula. Finally, we reached the result that the size of the original data and the encrypted data are equal. We showed the developed approach's performance compared with other standard cryptographic algorithms.

The paper's organization is the following: in Section 2, we discussed the important information related to the study. In Section 3, studies related to the proposed system are presented. In Section 4, we gave the proposed system structure with the

implementation details. The experimental result and achievements are discussed in Section 5, and finally, the paper is concluded in section 6 with potential future research directions.

2. BACKGROUND

2.1. Database Security

DBSs were created in the 1960s and have become crucial to different firms because this allowed the data to be better coordinated and accessible to consumers. A database (DB) is the collection of data that are connected in pairs and serve in place of information that can be recorded and contain indirect inferences. Databases are categorized based on their architecture: concentrated DB (CDB) and alternatively disseminated DB (DDB). A primary distinction between CDB and DDB is that the CDB saves whole statistics and information in a mere place; however, the DDB protects various parts of DB inside many corporeal places (Emad et al., 2017).

In CDB interference in a site leaves the complete structure inaccessible for any consumer; however, with DDB consumers can visit other DB locations. DBS can be built using different data structures, including comparative patterns, categorized patterns, and patterns toward the object. These DB patterns contain lots of personal information, like credit card use and history, medical records, and pupil history that must be kept secure against unauthorized use. As the risks for DBS were revealed, the necessity for maintaining data security coupled with confidentiality has arisen as a needed safeguard against any threats.

Many approaches have appeared to secure data, and three requirements must be fulfilled to achieve this goal. They are the following: confidentiality, integrity, and availability. Confidentiality is using rules to prohibit an unlicensed person from accessing records. The term “integrity” refers to the assurance that the data is not subjected to any alteration or degradation (Emad et al., 2017). Availability ensures that the customer has consistent and timely access to the information.

The lack of any of these requirements endangers the database. Since the mid-1970s, DB system security has gained a great deal of interest, beginning with entry restriction patterns to DB order that is regarded as one of the early safety approaches for DB safety. Entry restriction is a system that checks a consumer’s privileges in the face of a catalog of authorization to ensure data integrity and confidentiality. Authorization defines which database operations a consumer can carry out and what data the consumer can enter. One more tool for verifying the consumer’s persona is verification which is the first step in accessing the database (Emad et al., 2017).

Also, after verification of the consumer within the system, the database management system (DBMS) contained several mechanisms, for example, inspection inquiry coupled with a display, which kept the data safe from unauthorized interactions. An inspection inquiry is records of activities performed by a particular person in the database. If an illegal procedure is recorded, the database administrator (DBA) investigated the account number that carried it out. A sight approach is a digital table generated by performing comparative exercises on the base table (Kaur et al., 2014).

This approach allowed a user to access a portion of the data while the user cannot directly reach the database. Data secrecy may be protected by the encryption methods used in the figure. Encryption data is done by applying a cipher which makes it indecipherable for other consumers apart from the person having the security clearance to decode the data (Basharat et al., 2012).

2.2. Threats To Database Security

Risks are either problems or activities which may negatively impact DB protection, and they can be deliberate or unintentional (Kaur et al., 2014). The following are some of the most widespread threats to database security:

1. Privilege Abuse: We have two types of Privilege Abuse: Excessive Privilege Abuse (EPA) and Legitimate Privilege Abuse (LPA). EPA occurs when consumers illegally get control of entry rights for a database, which is above their task responsibilities; this excessive privilege could be abused for harmful purposes (Rohilla et al., 2013). The permitted user’s use of legal DB privilege for destructive purposes is called LPA (Rohilla et al., 2013).

2. Privilege Elevation: If the database has a loophole, an attacker might be able to manipulate it to change the entry requirements for all consumers from regular customers to company management (Rohilla et al., 2013).
3. SQL Injection: SQL Injection occurs when an assailant enters sequences of illegal SQL statements into a vulnerable SQL data means. By applying SQL injection, assailants can receive full entry for the whole database (Singh et al., 2014).
4. Platform Vulnerabilities: Vulnerabilities in performing systems and whatever external assistance is enabled on a DB server may cause DB harm, for instance, disallowed entry, rejection of service, and, ultimately, data exploitation (Singh et al., 2014).
5. Weak Audit Trail: Audit trails are designed to save a consumer's actions inside the database. As a result, the lack of an audit trail endangers the organization's databases (Basharat et al., 2012).
6. Rejection of Assistance: This threat inhibits permitted consumers from entering the DB. It threatens every company (Basharat et al., 2012).
7. Weak Authentication: Weak verification can enable hackers to employ techniques, for example (community engineering and violent strength) for hacking legitimate consumers' usernames and passwords to subsequently enter the database (Singh et al., 2014).
8. Backup Data Exposure: Since backup DB storage media is rarely secured from any threat, most often safety violations involve stealing hard disks as well as backup tapes (Singh et al., 2014).

2.3. Database Security Measures

Risks are either problems or activities which may negatively impact DB protection, and they can be deliberate or unintentional (Kaur et al., 2014). The following are some of the most widespread threats to database security:

Security methods that protect the database against unauthorized individuals, intentional attacks, data leakage, and hackers (Kaur et al., 2014). It covers many problems, including lawful, moral, regulation, and order-connected concerns. Database safety is a complex step that every company should strengthen to perform its actions smoothly and effectively. Any good company demands that the security and privacy of its database be protected from unauthorized entry, malicious intent, and unintended alteration (Basharat et al., 2012).

Data security is achieved through various sides of a data control scheme (DBMS). DBMS is a collection of implementations that handle the data stored inside a database and aids in data organization for better performance (Kulkarni et al., 2012).

To minimize risks, all DBMSs have security strategies developed for these purposes (Patil et al., 2012). Many security mechanisms have been developed to secure databases.

The four main security mechanisms that are applied to protect DBs from attacks are as follows: entry restriction, disruption management, flow management, and figure encryption.

Figure 1 depicts these monitoring steps. As previously said, entry management is a procedure that ensures data security by comparing the user's privileges to a list of licenses.

The mentioned licenses are managed using a Discretionary Access Control (DAC) approach, a Mandatory Access Control (MAC) approach, a Role-Based Access Control (RBAC) approach, or an Attribute-Based Access Control (ABAC) approach (ABAC). When individuals are allowed entry to analytical or synopsis data, the access control system prohibits them from gathering sensitive information that is not part of their user privileges.

Flow control guarantees that no unauthorized users can access the records. Encryption (as previously stated) is considered an act of transforming data by applying a cipher to cause it to be illegible for all consumers but the person having a password and user privilege to decode the figure (Kaur et al., 2014).

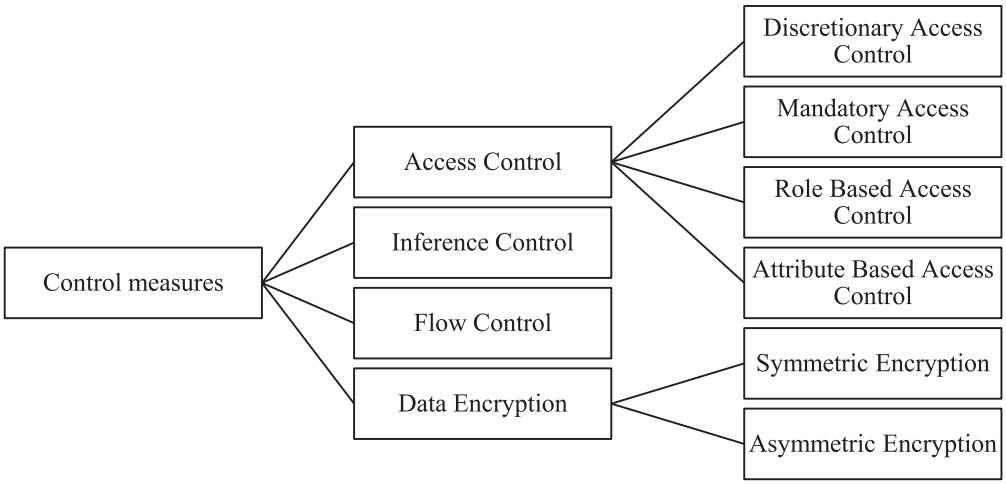


Figure 1. Control measures to protect DBs

2.4. Data Encryption

As data is encrypted with a cipher, it becomes unreadable to most consumers apart from people having the proper password to decode it (Basharat et al., 2012). Although hackers could compromise entry restriction procedures, encryption entries are still required for decoding the data (Thuraisingham et al., 2015).

Encryption stages rely on the algorithm and the entrance applied to encrypt the figure. There are two kinds of encryption: ordered encryption as well as deformed encryption. Encryption can be performed at three different stages, as seen below. Fig. 2 depicts them:

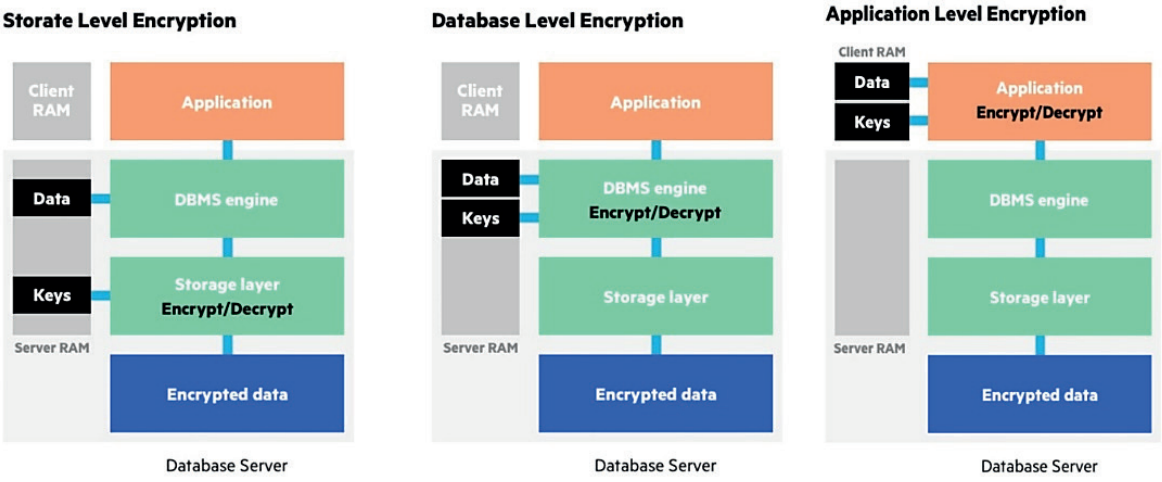


Figure 2. Levels of encryption

Storage-Level Encryption: storage-level encryption encrypts data inside the storehouse component, protecting it, for instance, against television stealing). It is ideal for encrypting folders, and whole folders within a performing order circumstance. Storage-level encryption has the benefit of being transparent from a database standpoint, preventing any updates to the current application. However, as the storehouse component is unaware of objects of the database as well as order, the encryption technique is not likely to be linked to consumer rights, for example, applying different encryption entries for different consumers) or data privacy.

As a result, choosy encryption – that is, encrypting database parts for reducing encryption elevated –is restricted to the folder granularity. Furthermore, discriminatingly encrypting folders is unsafe as no copy of confidential data can be left unencrypted, for instance, in log folders, momentary files, and so on. (Luc et al., 2010).

Database Encryption (DLE): database-level encryption secures data since it is being put into alternatively extracted relative to a database. Thus, the encryption technique could get integrated into database structures and linked to data privacy and consumer rights. Using discriminating encryption is feasible and could be conducted at different levels of granularity, including rows, columns, and tables. It could also get linked to logical requirements (for example, encrypt wages greater than 10K€ per month).

For both methods, data is decrypted on the database server at duration. In this way, on the server side, the encryption entries must be transferred and held, coupled with encrypted files, offering a poor level of security to stop an unprincipled server administrator or an attacker impersonating the administrator. Hackers could also follow the history and find encryption keys or plain text data (Luc et al., 2010).

Application-Level Encryption: Application-level encryption transfers the encryption/decryption mechanism toward the data-generating devices. Encryption is conducted in the program, bringing the data to the network. The data is transmitted encrypted, customarily kept, and recovered encrypted, to be eventually decrypted inside the application.

Since the keys should never depart the application side, this technique has the advantage of separating encryption keys from the encrypted data contained in the database. But, to implement this approach, programs must be modified. Furthermore, relying on the encryption granularity, the program could be required to retrieve a significant collection of data than that allowed to the user, thereby creating a security leak. Indeed, the user (or other hacker obtaining entry to the computer that the program works) can access the program and receive illegal entry to data. Finally, such a technique results in output overheads (indexes on encrypted information are purposeless) and also prevents the employment of progressive database workability like saved formulas (i.e., code saved within the DBMS that could be exchanged and used via multiple programs) as well as triggers on encrypted data (i.e., code fired while information within the database is added).

Application-level encryption has the most flexibility regarding granularity and essential control since the encryption granularity, and encryption keys can be selected based on application logic (Luc et al., 2010).

2.5. ASCII Code

American Standard Code for Information Interchange (ASCII) is a symbol-representation code that applies data. Each letter is a designated number between 0 and 127. Separate data is designated for each capital as well as a little letter.

As seen in the ASCII below, the capital letter A was designated the digit 65. However, the little letter was assigned the decimal number 97. ASCII code returns to the teletype's times and mechanic-like printers, because it precedes the Internet. Management codes for ASCII decimal data are limited to numbers between 0 to 31 and are not widely available anymore.

However, if you want to attempt to play alongside association procedures, you can view the restriction codes in the application. The restriction codes are explained in the ASCII Control Codes table. The entries you click and the correspondences are obtained like a sequence of data while a device delivers the results. Letters that you wrote or made are symbolized through this data. Because ASCII is limited to numbers between 0 and 127, 7 bits or 1 byte of information is required. To give an ASCII example of string coding, it could be 99 97 99 116 117 115 46 105 111. These coding bits, as well as bytes, are understood by microprocessors. All is a series of pieces to it.

2.6. Cryptography

Cryptography is a method of protecting data from unauthorized entry. It is made up of two major parts:

Nowadays, cryptography has turned into a requirement to all firms. Data safety is a critical element in a firm for maintaining and protecting their firm's data from rival firms. It also assists in ensuring a consumer's confidentiality. Currently, codes

could be more dependable regarding this duty as it is effortless to find codes because of the limited scope. Furthermore, if the passcode is not strong, it is relatively easy for a hacker to break the it. (Kakkar et al., 2010).

Accordingly, to preserve data, different algorithms have been developed. It assists in carefully preventing the unauthorized use of and maintaining the privacy and security of bank accounts, electronic repositioning of financial supports, and a lot of daily life consequences.

3. PREVIOUS WORKS

As new technology is developed, ensuring the security of databases is an area that has received long term and ongoing research.. Some of the works related to that research work are listed in this section.

Authors in (Diaa et al., 2010) investigated the Symmetric Encryption Algorithms' performance. This study evaluated six exceptional, widespread encryption algorithms: AES (Rijndael), DES, 3DES, RC2, Blowfish, and RC6. A differentiation was implemented at different setups per algorithm, like different dimensions of data blocks, various data types, battery power consumption, various key dimensions, and consequently, encryption/decryption quickness. The empirical model indicated the succeeding consequences.

No crucial difference existed in the exhibition of the results, whether in hexadecimal base encoding or foundation 64 encodings. When modifying packet dimension, it was discovered that RC6 demands fewer durations when compared to the total algorithms excluding Blowfish. When it came to changing data categories like images in the position of text, it was discovered that RC2, RC6, and Blowfish contained drawbacks as compared to other algorithms concerning spending time. Furthermore, 3DES, up to now, had weaker operations than algorithm DES. Lastly, when modifying the key dimension (feasible just in AES and RC6 algorithms), it showed that a more significant key dimension caused the evident difference in the battery and time-wasting.

The authors of (Hamdan et al., 2010) conducted a relative dissection of three encryption algorithms (DES, 3DES, as well as AES) based on nine criteria, including key breadth, cipher kind, block dimensions, safety, possible keys, feasible ASCII can be printed letter keys, as well as duration, taken for searching the total feasible keys at 50 billion keys each moment, and so on. According to research, AES was superior to DES and 3DES.

Authors in (Agarwal et al., 2010) provided a thorough review of common symmetric key encryption algorithms such as DES, TRIPLE DES, AES, and Blowfish. Symmetric Key algorithms, such as RSA, were quicker than Asymmetric Key algorithms. Furthermore, Symmetric algorithms needed less memory than Asymmetric encryption algorithms. Moreover, Symmetric-key encryption was more secure in comparison to Asymmetric key encryption.

Authors in (Singh et al., 2013), showed that the AES algorithm was efficient in quickness, duration, quantity, and outpouring impact.

The author in (Mittal, 2012) discovered that the algorithm's capability depended on the key breadth. Key breadth was a straight comparable for safety and conversely a comparable for Implementation. When the key breadth rises, the algorithm's security expanded; but, the operation decreased.

The authors in (Sesay et al., 2005) examined the importance of cryptography in database security. Confidential data stored in the clear in database systems was subject to attack. No matter how many security mechanisms were implemented, there were certain security flaws that attackers could exploit to access the database. However, information leaks could be avoided by encrypting confidential data before storing it in the archive. And the whole database security problem could be reduced to protecting a few cryptographic keys.

This method is based on ASCII values in (Arya et al., 2017). For encryption and decryption, ASCII characters utilized with string length accompanied by numerical calculations. To crack the procedure, the attacker needs a lot of information about the plain text; a single piece of information, such as string length, is insufficient. The use of different string lengths strengthens the technique. Further operations apply and are dependent on the size of the string. Consequently, that technique does not depend on any specific key or key generation method. It is the strength of the method.

Authors in (Shinge et al., 2014) suggested a symmetric encryption algorithm applying ASCII prices of the figure. The algorithm provides positive leads to little performing duration. This method produces a key for encrypting the letter. The spontaneously produced key is changed to a distinct string, and the identical key is applied to encryption and decryption. Consequently, this algorithm is identified as the symmetric-key algorithm.

Authors in (Seth et al., 2011) performed the dissection of three algorithms: RSA, DES, and AES; while considering some parameters, for example, the length of duration, memory consumption as well as output byte. The frameworks are a primary problem to worry about in all Encryption Algorithm. Empirical results indicate that the DES algorithm requires minimum encryption duration and the AES algorithm contains minimum memory application; however, the encryption duration variety is a small problem when it comes to AES and the DES algorithm. The RSA algorithm requires the largest encryption duration and memory utilization, which is extremely big; however, the output byte is minimal.

The authors (Kulkarni et al., 2012) deduced that databases were the primary factor to whatever varieties of implications. The database includes extremely significant and private data so hacks are very likely. Different types of risks on databases were explored in this study. The revision of effective database safety methods, such as entry control, methods relative to SQLIA, encryption, and data deforming, was explored.

The authors of (Elminaam et al., 2009) provided an implementation assessment of various symmetric encryption algorithms. The selected algorithms were AES, DES, 3DES, RC6, Blowfish, and RC2. They discovered that when the packet size was changed, Blowfish outperformed other encryption algorithms.

4. PROPOSED METHOD

The formula is clarified using a working style chart seen in Figure 1. We used these in our formula:

1. 3 keys are created by the user, one by the programmer, and the other in a variable way.
2. Two texts to create key1 and key2.
3. Ascii Code.
4. Four Sub formulas

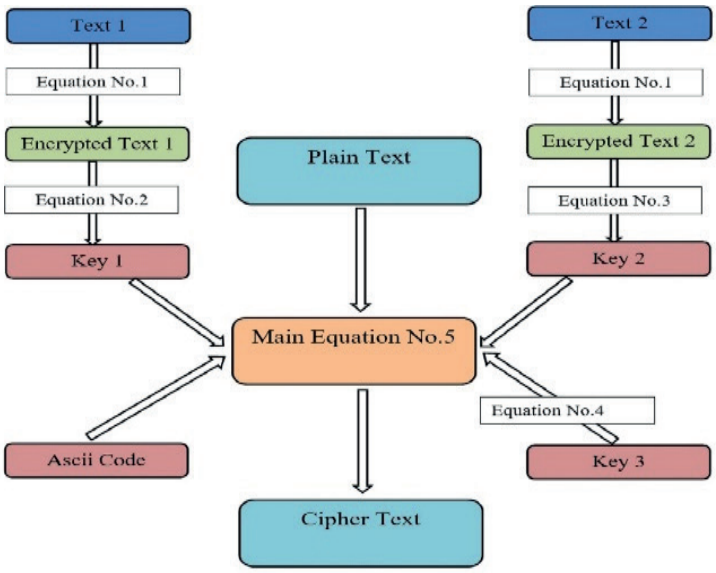


Figure 3. Equation diagram

In general, our formula was applied in the following steps:

In the first step, we needed two texts, one written by the user and the other by the programmer. The number of characters in the text had to be between 8 and 50. The character of the text was one of these characters (a-z, A-Z, 0-9, Keyboard Symbols). We encrypted both texts through formula number 1.

For example, Text1 = Ah5$Z2t*K7 and Text2 = @mQ4s#D7$L

So: Eq (1) = (Ascii code for character + sum of character number in Text)

Table 1
Examples of encrypting text (1) in equation (1)

Character	**A**	**h**	**5**	**$**	**Z**	**2**	**t**	*****	**K**	**7**
Ascii code	65	104	53	36	90	50	116	42	75	55
Eq(1)	65+10	104+10	53+10	36+10	90+10	50+10	116+10	42+10	75+10	55+10
New Ascii code	75	114	63	46	100	60	126	52	85	65
New Character	K	r	?	.	d	<	~	4	U	A

Table 2
Examples of encrypting text (2) in equation (1)

Character	**@**	**M**	**Q**	**4**	**s**	**#**	**D**	**7**	**$**	**L**
Ascii code	64	109	81	52	115	35	68	55	36	76
Eq(1)	64+10	109+10	81+10	52+10	115+10	35+10	68+10	55+10	36+10	76+10
New Ascii code	74	119	91	62	125	45	78	65	46	86
New Character	J	W	[	>	}	-	N	A	.	V

Text1 changed to Kr?.d<~4UA, and Text2 changed to Jw[>}-NA.V

In the second step, we got both Key1 and Key2 by applying two different formulas (formula number 2 on text 1 and formula number 3 on text 2).

Eq (2) = ((total number of all encrypted character ascii code number in Text1) * 2) / (sum of character number in Text1 -2) = Key 1

Table 3
Key (1) generating equation

Character	**K**	**R**	**?**	**.**	**d**	**<**	**~**	**4**	**U**	**A**
Ascii code	75	114	63	46	100	60	126	52	85	65
Eq(2)	((75+114+63+46+100+60+126+52+85+65) * 2) / (10 - 2) = (786 * 2) / 8 = 196.5									

Key1 = 196 Because Key1 should be integer number

Eq (3) = ((total number of all encrypted character ascii code number in Text2) - Key1) / (sum of character number in Text2 -1) = Key 2

Table 4
Key (2) generating equation

Character	**J**	**w**	**[**	**>**	**}**	**-**	**N**	**A**	**.**	**V**
Ascii code	74	119	91	62	125	45	78	65	46	86
Eq (2)	((74+119+91+62+125+45+78+65+46+86) - 196) / (10 - 1) = (791 - 196) / 9 = 66.11									

Key2 = 66, because Key2 should be an integer number like Key1.

In the third step, by applying a sub-formula number 4, we got Key 3, which was a variable key and shifted according to the number of characters of the text we encrypted.

Key3= sum of the numbers in the ASCII code of the letter

In the fourth step, we got our encrypted letter using all three keys (1, 2, 3), with ASCII code for characters in the main formula.

We applied the formula to the plain text (Sakarya) when it was known that Key1 =196 and Key2 =66

Eq (5) = (Ascii code for character + Key1) + (Key2 - sum of character number in Key2) + Key3 = New Ascii code. For Encryption:

Table 5
Encrypting procedure

Plain Character	Ascii code	Key1	Key2	Key3	Encryption Equation: (Ascii code + K1) + (K2 - Length of K2) +K3	New Ascii code	Cipher Character
S	83	196	66	11	(83 + 196) + (66 - 10) + 11	346	Ś
A	97	196	66	16	(97 + 196) + (66 - 10) + 16	365	ŭ
K	107	196	66	8	(107 + 196) + (66 - 10) + 8	367	ů
A	97	196	66	16	(97 + 196) + (66 - 10) + 16	365	ŭ
R	114	196	66	6	(114 + 196) + (66 - 10) + 6	372	Ŵ
Y	121	196	66	4	(121 + 196) + (66 - 10) + 4	377	Ź
A	97	196	66	16	(97 + 196) + (66 - 10) + 16	365	ŭ

Eq (5) = (New Ascii code + Key1) - (Key2 - sum of character number in Key2) - Key3 = Ascii code

For Decryption:

Table 6
Decrypting procedure

Cipher Character	New Ascii code	Key1	Key2	Key3	Decryption Equation (New Ascii code + K1) - (K2 - Length of K2) - K3	Ascii code	Plain Character
Ś	346	196	66	11	(346 - 196) - (66 - 10) - 11	83	S
ŭ	380	196	66	16	(380 - 196) - (66 - 10) - 16	97	A
ů	372	196	66	8	(372 - 196) - (66 - 10) – 8	107	K
ŭ	379	196	66	16	(379 - 196) - (66 - 10) – 16	97	A
Ŵ	381	196	66	6	(381 - 196) - (66 - 10) – 6	114	R
Ź	401	196	66	4	(401 - 196) - (66 - 10) – 4	121	Y
ŭ	367	196	66	16	(367 - 196) - (66 - 10) – 16	97	A

If the data type is number or currency, then: Eq (5) = (value of (number or currency) / 5) + 5

For example: How to encrypt 550, Encrypted value = (550 / 5) + 5 = 110 + 5 = 115

Eq (5) = ((Encrypted value) - 5) * 5

Now, to decrypt 115, decrypted value = (115 – 5) * 5 = 110 * 5 = 550

4.1. Implementation

To test the formula, we designed a database from Microsoft access 2016 that consisted of two parts; one was to insert some data into the database table, and the other was to test our formula's speed. To access our database, a user must be created, as seen in Figure 4. Then, as seen in Figure 5. we logged into the database using the user and password that were created.

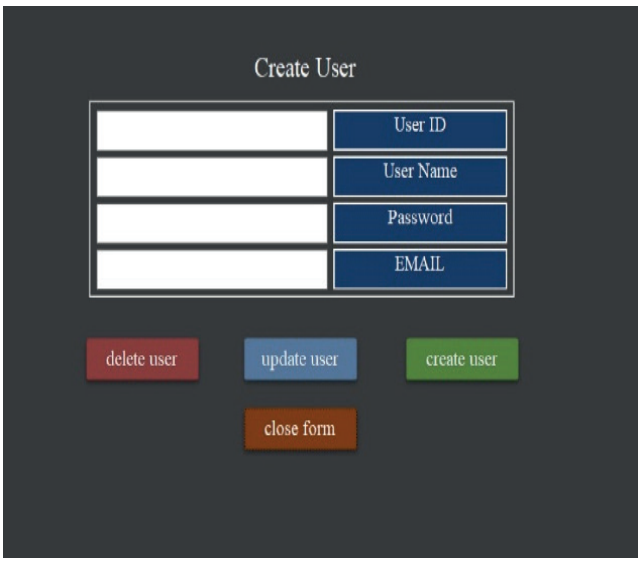


Figure 4. Create User

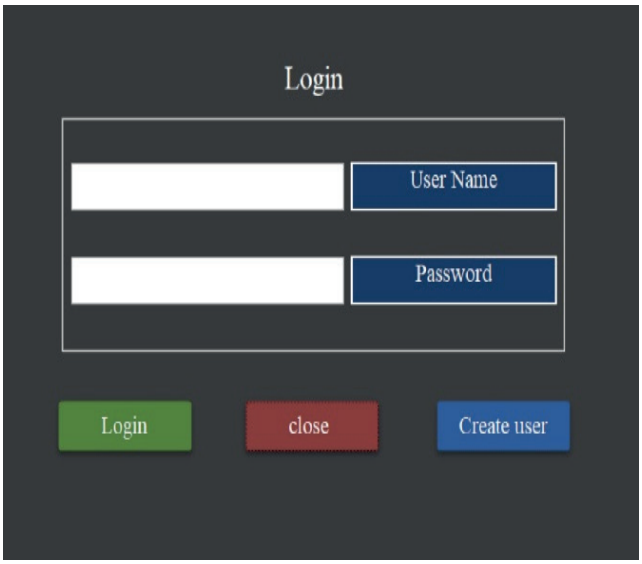


Figure 5. Login form

Next, we opened a form with two sections, one of which was for data entry and the other for data recording, where Text 1 and Text 2 were used to enter information such as (Id, Full name, Age, Gender, Email) after which we saw that the data was encrypted as seen in Figure 6.

In the second section, we inserted Text 1 and Text 2, and then we inserted our plain text, which, after we pressed the encrypt command, produced ciphertext in one text box and the time to encryption in another text box. For decryption, we followed the same procedure as seen in Figure 7.

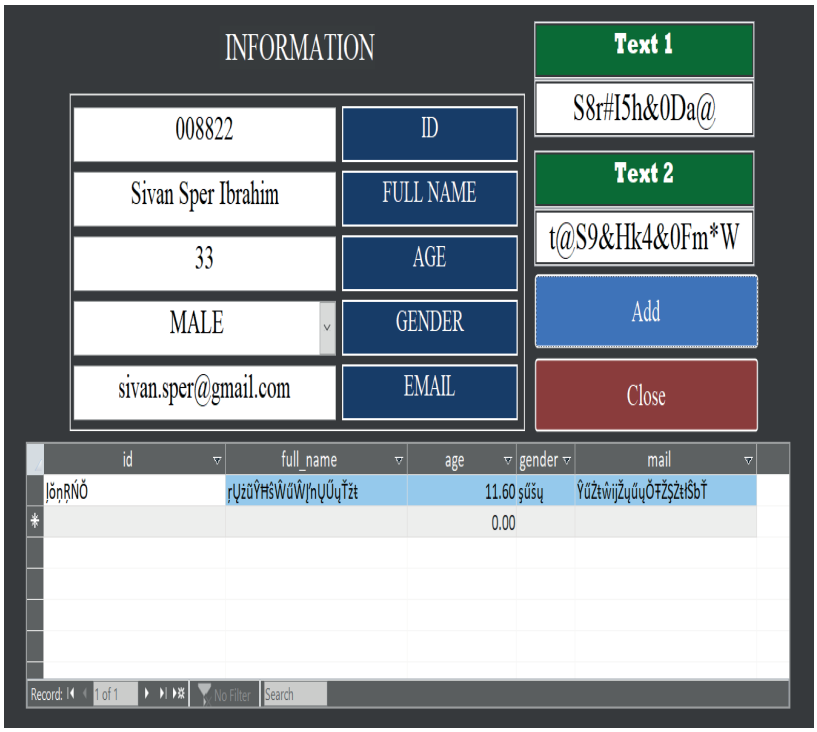


Figure 6. Encryption form

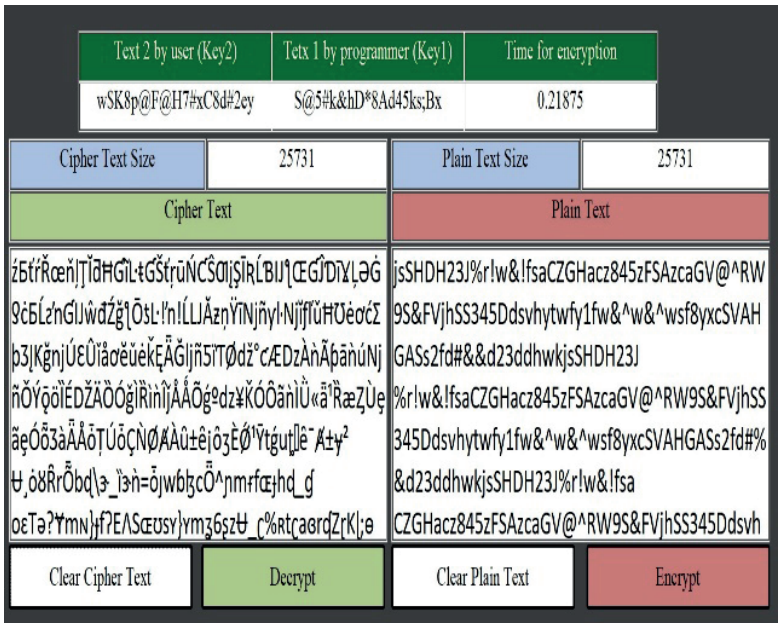


Figure 7. Speed test form

5. RESULTS AND DISCUSSION

This section compared our algorithm to other algorithms (Blowfish, DES, 3DES, RSE, and RSA). For instance, if we have data as shown in Table 7, we reached the results using the following::

Key1 = S8r#I5h&0Da@

Key2 = t@S9&Hk4&0Fm*W

1. When we encrypted plain text, it provided us with the same size as plain text.

2. If a character appeared in the text more than once, it was encrypted with a new character each time.

3. When we encrypted text-type data, it gave us text-type data, but if we encrypted number-type data, we got number-type data, which was another significant point for the algorithm speed when searching for specific data.

4. Another essential point in our algorithm was that when we encrypted different characters in the plain text, the matched character appeared in the ciphertext. This was significant for data security because it would be impossible to convert the encrypted character back to the original if an attacker took over the data.

5. Another point was the length of our Keys, which were between 8 and 50 characters, which was a vital factor in protecting our algorithm because breaking the Keys would require more time and effort.

To clarify the above-mentioned points in our experimental results, we provided two charts for both encrypted and decrypted texts.

Figure 7 represents the original plain text which holds 52 frequencies the same size as the text, while Figure 7. represents the secured text which holds 1039 frequencies. This proves that the proposed method works well on the suggested text.

Table 8
Data encryption in our system

Field name	Id	Full name	Age	Gender
plain text	008821	Sivan Sper Ibrahim	33	MALE
cipher text	ļõņŖŃō	ŗŲżũŶĦŝŴűŴĮnŲŰųŤžŧ	11.6	şűšų

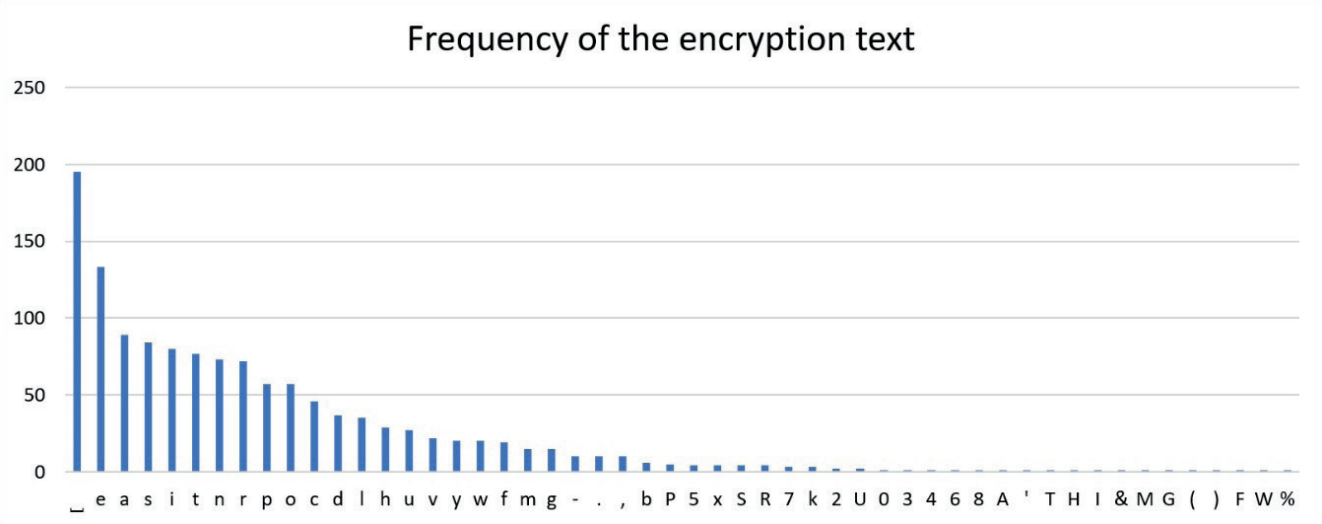


Figure 8. Frequency of the encryption text

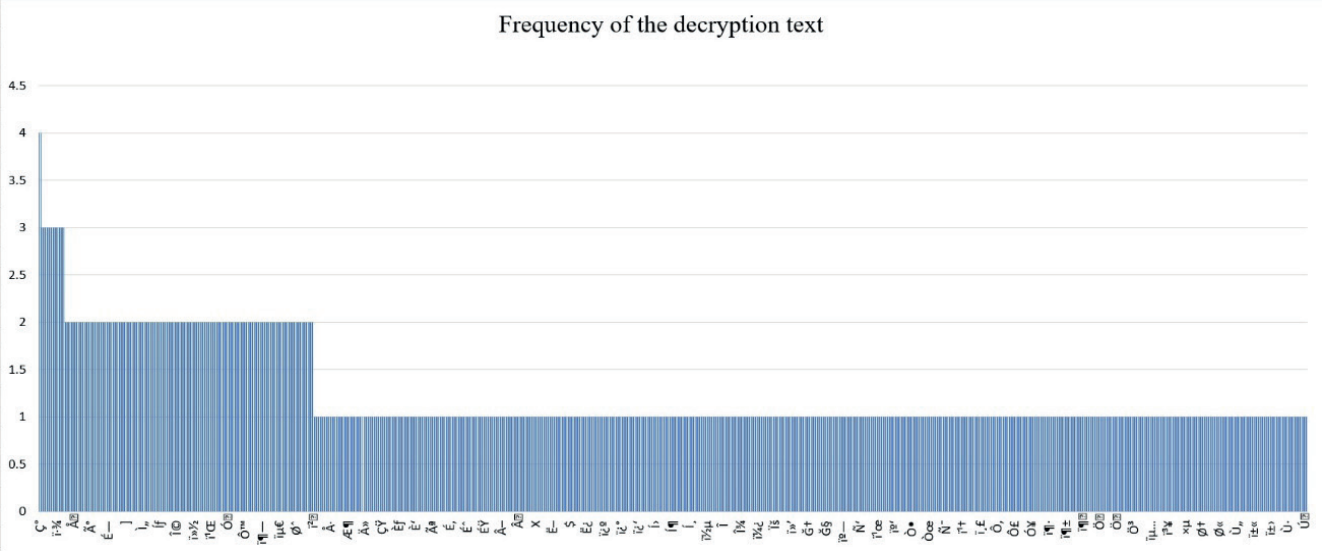


Figure 9. Frequency of the decryption text

To compare the speed of our algorithm, we used the previous research (Patil et al., 2016). The author showed that when a 25KB file (RSA) was encrypted, it took more time than other algorithms, but (Blowfish) took the least amount of time, as shown in Figure 8.

When we conducted similar research on a decrypted 25kb file (RSA), it took more time than other algorithms, but (Blowfish) took the least time, as shown in Figure 9.

To compare algorithm speeds for a 25 KB file, we tested our method. We found that it required 256ms for encryption and 272ms for decryption after five repetitions, indicating that our algorithm was quicker than all of the (RSA, DES, TDES, and RSA) encryption as decryption speeds, as illustrated in Figure 10.

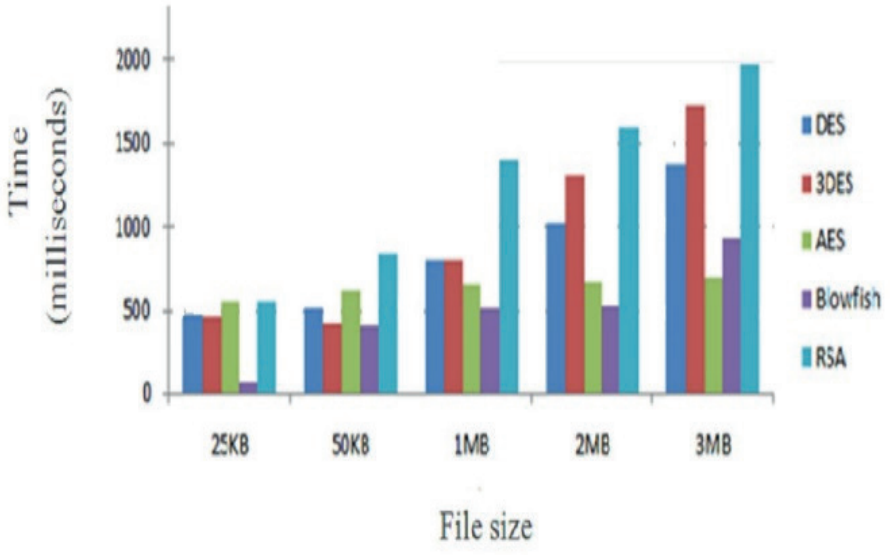


Figure 10. Encryption time vs. file size for DES, 3DES, AES, Blowfish, and RSA (Patil et al., 2016)

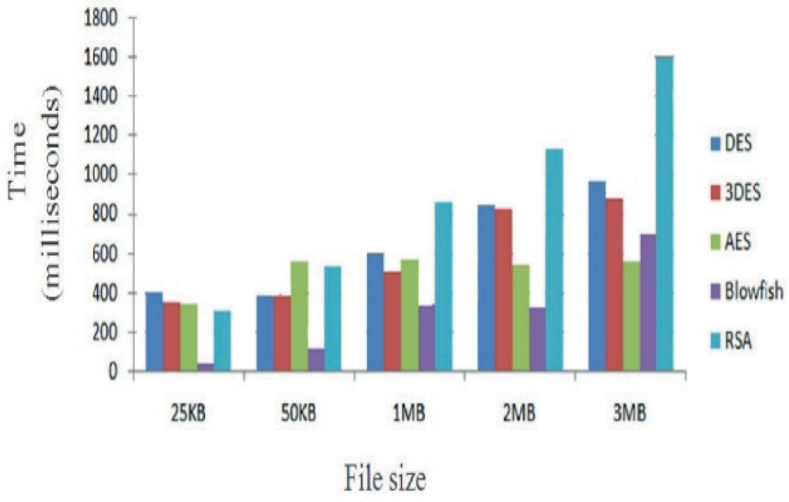


Figure 11. Decryption time vs. file size for DES, 3DES, AES, Blowfish, and RSA (Patil et al., 2016)

To compare the algorithm speeds for a 25 KB file, we tested our method. It required 256ms for encryption and 272ms for decryption after five repetitions, indicating that our algorithm was quicker than all the (RSA, DES, TDES, and RSA) encryption as decryption speeds, as illustrated in Figure 11.

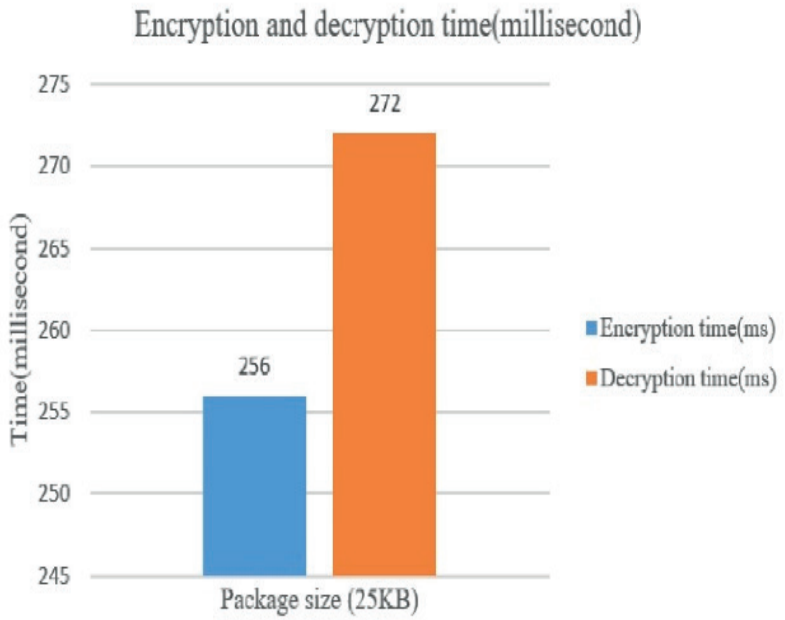


Figure 12. Encryption and decryption time

As a result, we will use the algorithm to create databases and encrypt data because it has large and fast data encryption security, which will secure our data from attacks.

6. CONCLUSIONS AND FUTURE WORK

Cryptographic algorithms protect the confidentiality, integrity, and authenticity of data stored in databases. Confidentiality is the property of ensuring that sensitive information is not disclosed to unauthorized individuals. Cryptographic algorithms can encrypt data stored in a database so that even if an attacker gains access to the database, they will not be able to read the data without the decryption key. Integrity ensures that unauthorized individuals have not modified or tampered with data. Cryptographic algorithms create a message digest of the data, which can then be stored in the database and the data itself. The message digest is a unique representation of the data produced by applying a mathematical function to the data. If the data is modified, the message digest will also be different, allowing for the detection of any changes to the data.

Authenticity ensures that data coming from the source it claims to come from has not been tampered with in transit. Cryptographic algorithms can be used to sign data using a private key so that the recipient can verify the authenticity of the data using the corresponding public key.

Overall, cryptographic algorithms play a critical role in ensuring the security and integrity of data stored in databases. This paper is a comparative study between created algorithms for data encryption used when creating databases against several different cryptographic algorithms, including (DES, 3DES, AES, Blowfish, and RSA) in terms of security and speed of encryption and decryption. The results revealed that the created algorithm was faster than the (DES, 3DES, AES, and RSA) algorithms but slower than the (Blowfish) algorithm. Also, in the created algorithm, the plain text and the ciphertext were the same size. When we encrypted text-type data, we got text-type data. But if we encrypted number type-data, we got number-type data. The developed algorithm is secure. If there is more than one character in the text, each will encrypt with a different character. Also, when we encrypted various characters in plain text, the matched character appeared in the ciphertext. In the future, we will focus on the created algorithm, which can be compared with other cryptographic algorithms.

Peer-review: Externally peer-reviewed.
Author Contributions: Conception/Design of Study- S.H., A.F.M.S.A., C.A., S.S.İ., A.Z.; Data Acquisition- S.H., A.F.M.S.A., C.A., S.S.İ., A.Z.; Data Analysis/Interpretation- S.H., A.F.M.S.A., C.A., S.S.İ., A.Z.; Drafting Manuscript- S.H., A.F.M.S.A., C.A., S.S.İ., A.Z.; Critical Revision of Manuscript- S.H., A.F.M.S.A., C.A., S.S.İ., A.Z.; Final Approval and Accountability- S.H., A.F.M.S.A., C.A., S.S.İ., A.Z.; Material and Technical Support- S.H., A.F.M.S.A., C.A., S.S.İ., A.Z.; Supervision- S.H., A.F.M.S.A., C.A., S.S.İ., A.Z.
Conflict of Interest: The authors have no conflict of interest to declare.
Grant Support: The authors declared that this study has received no financial support.

REFERENCES

Aamer N., Younus J., "A Performance Comparison of Data Encryption Algorithms", *First International Conference on IEEE Information and Communication Technologies (ICICT),* Vol 1, Issue 6, 27-28 Aug. 2005, pp 84-89.

Agrawal M., Mishra P., "A Comparative Survey on Symmetric Key Encryption Techniques" *International Journal on Computer Science and Engineering (IJCSE)*, Vol. 4 No. 05 May 2012, ISSN: 0975-3397.

Arya, E. S., & Kumar, A. (2017). Ascıı Based Encryptıon Decryptıon Technıque For Informatıon Securıty And Communıcatıon. In 3rd international conference on innovative trends in science, *Engineering and Management*, YMCA connaught place, New Delhi, 7th January.

Basharat, I., Azam, F. and Muzaffar, A., Database Security and Encryption: A Survey Study, *International Journal of Computer Applications*, 47(12): 28-34 (2012).

Bertino, E., & Sandhu, R. (2005). Database security-concepts, approaches, and challenges. *IEEE Transactions on Dependable and secure computing*, 2(1), 2-19.

Diaa E., Hatem K. and Mohie H., "Performance Evaluation of Symmetric Encryption Algorithms", *IJCSNS International Journal of Computer Science and Network Security,* VOL.8 No.12, pp. 280-286, December 2008.

Elminaam, D. S. A., Kader, H. M. A., & Hadhoud, M. M. (2009). Energy efficiency of encryption schemes for wireless devices. *International Journal of Computer Theory and Engineering,* 1(3), 302.

Emad F. Khalaf and Mustafa M. Kadi., "A Survey of Access Control and Data Encryption for Database Security", *JKAU: Eng. Sci*., Vol. 28 No. 1, pp: 19 - 30 (2017) Doi: 10.4197/Eng. 28-1.2.

Eskicioğlu, Ö. C. & Isık, A. H. (2022). Mobil Uyumlu Çoklu Dil Destekli Hibrit Şifreleme Algoritması . Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi , 24 (72) , 1007-1019 . DOI: 10.21205/deufmd.2022247228

Hamdan.O. Alanazi, B.B. Zaidan, A.A. Zaidan, Hamid A. Jalab, M. Shabbir and Y. Al-Nabhani, "New Comparative Study Between DES, 3DES and AES within Nine Factors" *Journal of Computing*, Volume 2, ISSUE 3, March 2010, ISSN 2151-9617.

Kakkar A. and Bansal P. K., "Reliable Encryption Algorithm used for Communication", *M. E. Thesis*, Thapar University, 2004.

Kaur, R., Kiranpreet and Verma, P., Survey on Database Security, *International Journal of Computer Applications*, 105(10): 27–31 (2014).

Kulkarni, S. and Urolagin, S., Review of Attacks on Databases and Database Security Techniques, *International Journal of Emerging Technology and Advanced Engineering*, 2(11): 253-263 (2012).

Luc B., Yanli GUO, "Database Encryption", *Article* · January 2010 DOI: 10.1007/978-1-4419-5906-5_677.

Mathur A., "A Research paper: An ASCII value based data encryption algorithm and its comparison with other symmetric data encryption algorithms", *International Journal on Computer Science and Engineering (IJCSE),* Vol. 4, pp. 1650-1657, Sep 2012 ISSN: 0975- 3397

Mittal H., "Optimized Encryption Algorithm using Dynamic Keys", *Electronics and Communication Engineering Department*, THAPAR UNIVERSITY PATIALA-147004, 2012

Nadeem A., "A performance comparison of data encryption algorithms," *IEEE Information and Communication Technologies*, pp. 84-89, 2006.

Patil, P., Narayankar, P., Narayan, D. G., & Meena, S. M. (2016). A comprehensive evaluation of cryptographic algorithms: DES, 3DES, AES, RSA and Blowfish. *Procedia Computer Science*, 78, 617-624.

Patil, A. and Meshram, B. B., Database Access Control Policies, *International Journal of Engineering Research and Applications,* 2(3): 3150–3154 (2012).

Rohilla, S. and Mittal, P. K., Database Security: Threats and Challenges, *International Journal of Advanced Research in Computer Science and Software Engineering,* 3(5): 810–813 (2013).

Sesay, S., Yang, Z., Chen, J., & Xu, D. (2005, January). A secure database encryption scheme. In Second *IEEE Consumer Communications and Networking Conference*, 2005. CCNC. 2005 (pp. 49-53). IEEE.

Seth, S. M., & Mishra, R. (2011). *Comparative analysis of encryption algorithms for data communication. IJCST Vol. 2, Issue 2, June 2, 2011.*

Shelly R., Pradeep Kumar M., Database Security: Threats and Challenges, International Journal of *Advanced Research in Computer Science and Software Engineering*, Volume 3, Issue 5, May 2013.

Shinge, S. R., & Patil, R. (2014). An encryption algorithm based on ASCII value of data. *International Journal of Computer Science and Information Technologies*, 5(6), 7232-7234.

Silberschatz, A., Korth, H. F., & Sudarshan, S. (2002). *Database system concepts (Vol. 5)*. New York: McGraw-Hill.

Singh, S. and Rai, R. K., A Review Report on Security Threats on Database*, International Journal of Computer Science and Information Technologies*, 5(3): 3215–3219 (2014).

Singh G., Supriya, "A Study of Encryption Algorithms (RSA, DES, 3DES and AES) for Information Security" *International Journal of Computer Applications* (0975 – 8887) Volume 67– No.19, April 2013.

Thuraisingham, B., Database Security: Past, Present, and Future, *IEEE International Congress on Big Data*, 772–774 (2015).

ACTA INFOLOGICA 2023;7(1):17-28
dergipark.org.tr/acin

DOI: 10.26650/acin.1070261

RESEARCH ARTICLE

Plant Classification Method Using Histogram and Machine Learning for Smart Agriculture Applications

Akıllı Tarım Uygulamaları için Histogram ve Makine Öğrenimi Kullanan Bitki Sınıflandırma Yöntemi

Orhan Yaman[1], **Türker Tuncer**[1]

[1] (Assoc. Prof.), Fırat University, Faculty of Technology, Department of Digital Forensics Engineering, Elazig, Turkiye

ORCID: O.Y. 0000-0001-9623-2284; T.T. 0000-0002-5126-6445

Corresponding author:
Orhan YAMAN
Fırat University, Faculty of Technology, Department of Digital Forensics Engineering, Elazig, Turkiye
E-mail address: orhanyaman@firat.edu.tr

Submitted: 08.02.2022
Revision Requested: 11.12.2022
Last Revision Received: 25.12.2022
Accepted: 05.01.2023
Published Online: 13.02.2023

Citation: Yaman, O., Tuncer, T. (2023). Plant classification method using histogram and machine learning for smart agriculture applications. *Acta Infologica, 7*(1), 17-28.
https://doi.org/10.26650/acin.1070261

ABSTRACT

Due to its high potential and value, the Internet of things (IoT) has been used in various areas such as information security, industry 4.0, and smart agriculture. IoT is used in agriculture through the use of sensors, unmanned aerial vehicles (UAV), satellite technologies, robots, image processing, and artificial intelligence technologies. These smart agricultural practices increase production and quality and lead to savings in irrigation, thereby reducing environmental pollution during production. This study proposes an ultra-lightweight automated plant species classification method for smart agriculture applications. A UAV is used to acquire a new image dataset. An ultra-lightweight classification method is then used to classify the acquired plant species images. Our proposed ultra-lightweight computer vision model presents a histogram-based simple feature extraction function. The presented feature extractor uses histogram extraction and median filter in conjunction. The generated features are fed to two shallow classifiers, which are the support vector machine (SVM), and k nearest neighbor (KNN). The utilized SVM and KNN classifiers have attained 96.45% and 94.11% accuracies consecutively. The results demonstrate that this model is very capable of plant image classification and is ready for use in a physical agriculture environment.

Keywords: Plant classification, smart agriculture, histogram, feature extraction, machine learning

ÖZ

Nesnelerin interneti (IoT) insanlık için çok değerli bir teknolojidir, dolayısıyla IoT bilgi güvenliği, endüstri 4.0, akıllı tarım gibi çeşitli alanlarda kullanılmaya başlanmıştır. Akıllı tarım uygulamaları sensörler, insansız hava araçları (İHA), uydu teknolojileri, robotlar, görüntü işleme ve yapay zekâ teknolojileri kullanılarak geliştirilmektedir. Akıllı tarım uygulamaları ile sulama alanında tasarruf sağlanmakta ve üretim sırasında çevre kirliliği azaltılmaktadır. Aynı zamanda üretimi ve kaliteyi arttırır. Bu çalışmada, akıllı tarım uygulamaları için ultra hafif otomatik bitki türleri sınıflandırma yöntemi geliştirilmiştir. Bir İHA kullanılarak yeni bir görüntü veri seti elde edilmiştir. Elde edilen bitki türleri görüntüsünü sınıflandırmak için ultra hafif bir sınıflandırma yöntemi önerilmiştir. Önerilen ultra hafif bilgisayarlı görü modelimizde, histogram tabanlı basit bir özellik çıkarma işlevi sunulmuştur. Sunulan öznitelik çıkarıcı, histogram çıkarımı ve medyan filtresi birlikte kullanılmıştır. Oluşturulan öznitelikler, destek vektör makinesi (SVM) ve k en yakın komşu (KNN) olan iki sığ sınıflandırıcıya beslenir. Kullanılan SVM ve KNN sınıflandırıcıları arka arkaya %96,45 ve %94,11 doğruluk elde etmiştir. Sonuçlar, bu modelin bitki görüntü sınıflandırması için oldukça başarılı olduğunu ve fiziksel tarım ortamında kullanıma hazır olduğunu göstermektedir.

Anahtar Kelimeler: Bitki sınıflandırması, akıllı tarım, histogram, özellik çıkarma, makine öğrenmesi

1. INTRODUCTION

Many factors influence agricultural productivity, including soil structure, irrigation type, spraying technique, and meteorological conditions. To enhance production output, it is important to continually monitor the plants and make the required modifications. Agricultural production has grown in recent years with the advancement of technology and smart agricultural systems. Smart agriculture can enable the development of applications such as water monitoring, soil monitoring, and disease monitoring. Temperature/humidity sensors, soil sensors, and cameras can all be used for these applications (Babayigit & Büyükpatpat, 2019), thereby automatically computing irrigation, fertilization, agricultural spraying, and yield analysis.

Plant species and disease detection are significant for smart agricultural applications. It is important to know the type of plant to assess its production, development, and disease. Furthermore, determining the plant species is required to detect diseases that may afflict them. In the literature, techniques for classifying plant species based on fruits and leaves have been developed (Atila, Uçar, Akyol, & Uçar, 2021; Hameed, Chai, & Rassau, 2018; Yalcin & Razavi, 2016; Zhu, Zhu, & Ren, 2018). Plants' fruits (Hameed et al., 2018), flowers (Pawara, Okafor, Schomaker, & Wiering, 2017), and leaves (Keivani, Mazloum, Sedaghatfar, & Tavakoli, 2020) are generally used to classify them (Yalcin & Razavi, 2016). Moreover, this also results in disease detection in the images as well (El, Es-saady, El, Mammass, & Benazoun, 2017; Wang, Yang, Yu, Dong, & Wang, 2021; Xie, Yang, & He, 2017). Table 1 summarizes the research and methodologies utilized in plant classification and disease detection in the literature.

Table 1
Studies and methods used in plant classification and disease detection in the literature

Studies	Year	Dataset	Number of species/groups	Explanation	Method
Atila et al. (Atila et al., 2021)	2021	PlantVillage dataset	38	Leaf disease classification	EfficientNet
Goyal et al. (Goyal, Kumar, & Gupta, 2021)	2021	Swedish leaf dataset	15	Plant recognition	1DPCA+ MTSVM
		Flavia dataset	-		1DPCA+ MTSVM
		Folio dataset	-		LDA+ MTSVM
Selvam et al. (Selvam & Kavitha, 2020)	2020	Self-collected dataset	3	Leaf disease classification	CNN
Murtaza et al. (Murtaza, Saba, Haroon Yousaf, & Viriri, 2020)	2020	Flavia dataset	3	Plant Species Identification	SURF + Bag of Words
		ImageClef dataset	3		SURF + Bag of Words
Adak (Adak, 2020)	2020	Self-collected dataset	43	Identification of Plant Species	CNN
Saleem et al. (Saleem, Akhtar, Ahmed, & Qureshi, 2019)	2019	Flavia dataset	25	Plant classification	CNN KNN Decision Tree Naïve Bayes Multi SVM
Yacin et al. (Yalcin & Razavi, 2016)	2016	TARBIL dataset	16	Plant Classification	CNN SVM
Dyrmann et al. (Dyrmann, Karstoft, & Midtiby, 2016)	2016	BBCH 12e16 dataset	22	Plant species classification	CNN

* CNN: Convolutional neural network, SVM: Support vector machine, KNN: k nearest neighbor

As can be seen in Table 1, many image processing, machine learning, and deep learning-based methods have been developed in the literature. There are many datasets of fruits, vegetables, and other plants in the literature. PlantVillage is a dataset containing leaf diseases in plants. The Swedish, Flavia, Folio, TARBIL, and BBCH 12e16 datasets are used for plant classification. Deep learning-based methods have high computational complexity necessitating methodical work and computers with high processors.

2. MATERIAL METHOD

2.1. The collected dataset

In the first step of this research, a new plant image dataset has been collected using a UAV. This dataset contains 17 plant species and consists of 6170 color images. The size of these images are fixed to 1280 × 720 × 3 since they are RGB images. They recorded as JPEG. This dataset contains common garden plants. The used species are watermelon, corn, pepper, eggplant, tomato, wild grass, bean, cucumber, zucchini, melon, white cabbage, black cabbage, basil, parsley, carrot, mint, and strawberry plants. The sample plant images and counts used in the collected dataset are shown in Figure 1.

Figure 1. Plant images and numbers in the collected dataset

As illustrated in Figure 1, plant images have been collected with a UAV camera.

2.2. Preprocessing

The dataset collected in this study was carried out on agricultural land of approximately 3000 square meters. There are thousands of roots belonging to each plant species here. Video recordings were collected by drone for each plant species. One image was extracted from the videos at 300 millisecond intervals. Thus, approximately 3 images were obtained in 1 second. The images were resized and used. The number of images obtained for each plant category is given in Figure 1 and Figure 5.

2.3. The proposed ultra-lightweight plant species classification model using images

This research presents an ultra-lightweight image classification model for plan image classification. This model consists of two basic phases: feature extraction using histogram extraction and median filter and classification. The block diagram of the proposed method in this study is summarized in Figure 2.

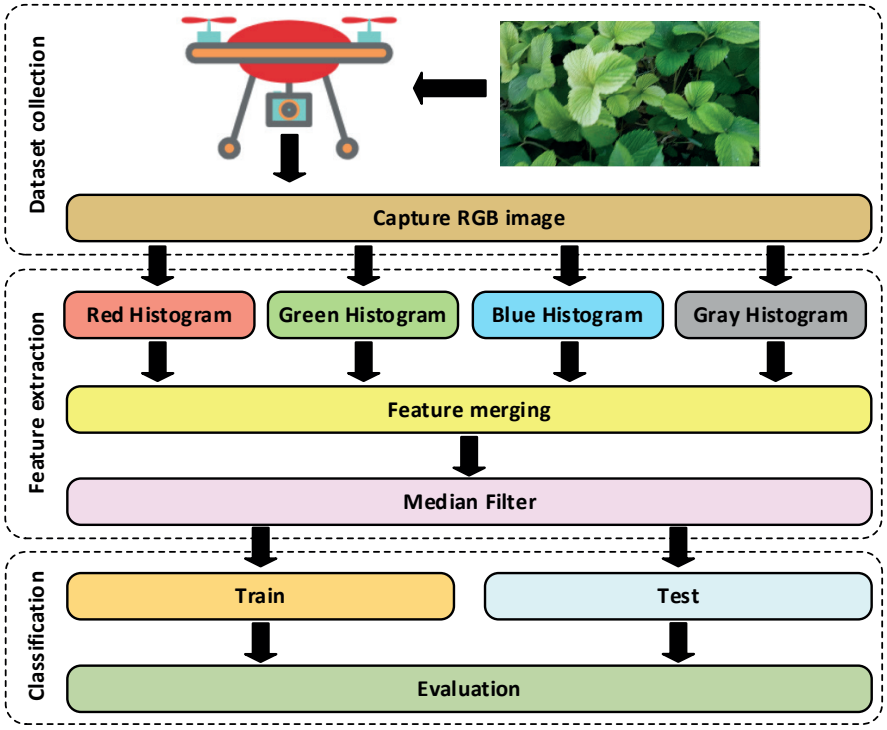


Figure 2. Block diagram of the proposed method

As shown in Figures 2, the collected images are RGB. RGB images consist of Red, Green, and Blue channels. In the proposed method, the histogram is used to extract features from the image. Separate histograms have been extracted from the Red, Green, and Blue channels. In addition, the RGB image has been converted to a gray image. Thus, four images have been obtained from an RGB image. The Red, Green, Blue, and Gray channels obtained from a sample RGB image are illustrated in Figure 3.

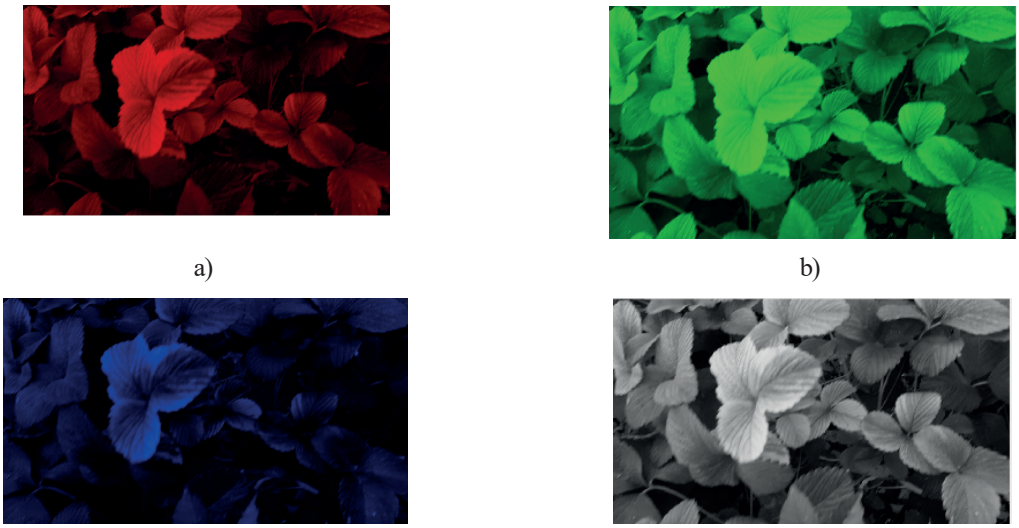


Figure 3. Red, Green, Blue and Gray channels from a sample RGB image
a) Red image b) Green image c) Blue image d) Gray image

In the proposed method, a histogram is used to extract features from the image. Separate histograms have been taken for the Red, Green, and Blue channels. A histogram is a graph that shows the number of color values in an image. In the histogram graph, the horizontal axis represents the color values between 0-255. The vertical axis represents the pixel numbers of these color values. The color distribution in the image is obtained by histogram extraction. Thus, feature extraction can be done from an image. Histogram graphics of the sample Red, Green, Blue, and Gray channels given in Figure 3 are illustrated in Figure 4.

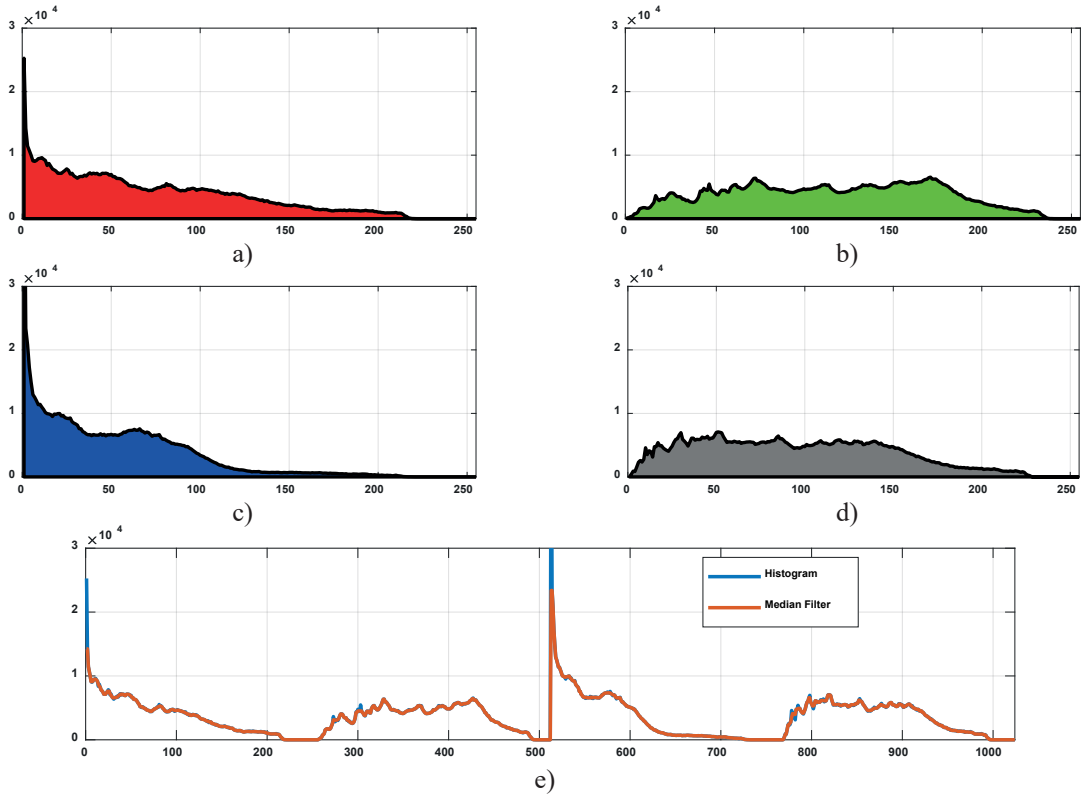


Figure 4. Example Red, Green, Blue and Gray channels histogram graphics
a) Red channel histogram b) Green channel histogram c) Blue channel histogram d) Gray channel histogram e) Merging histograms of Red, Green, Blue and Gray channels

As illustrated in Figure 4, the histogram graphs of Red, Green, Blue, and Gray channels have been obtained and concatenated. Thus, the length of the created feature is 256 × 4=1024. A median filter has been applied to the obtained histogram graphics. By applying the median filter, the noise in the extracted features is removed. The proposed feature extraction results for images of 17 plant species are demonstrated in Figure 5.

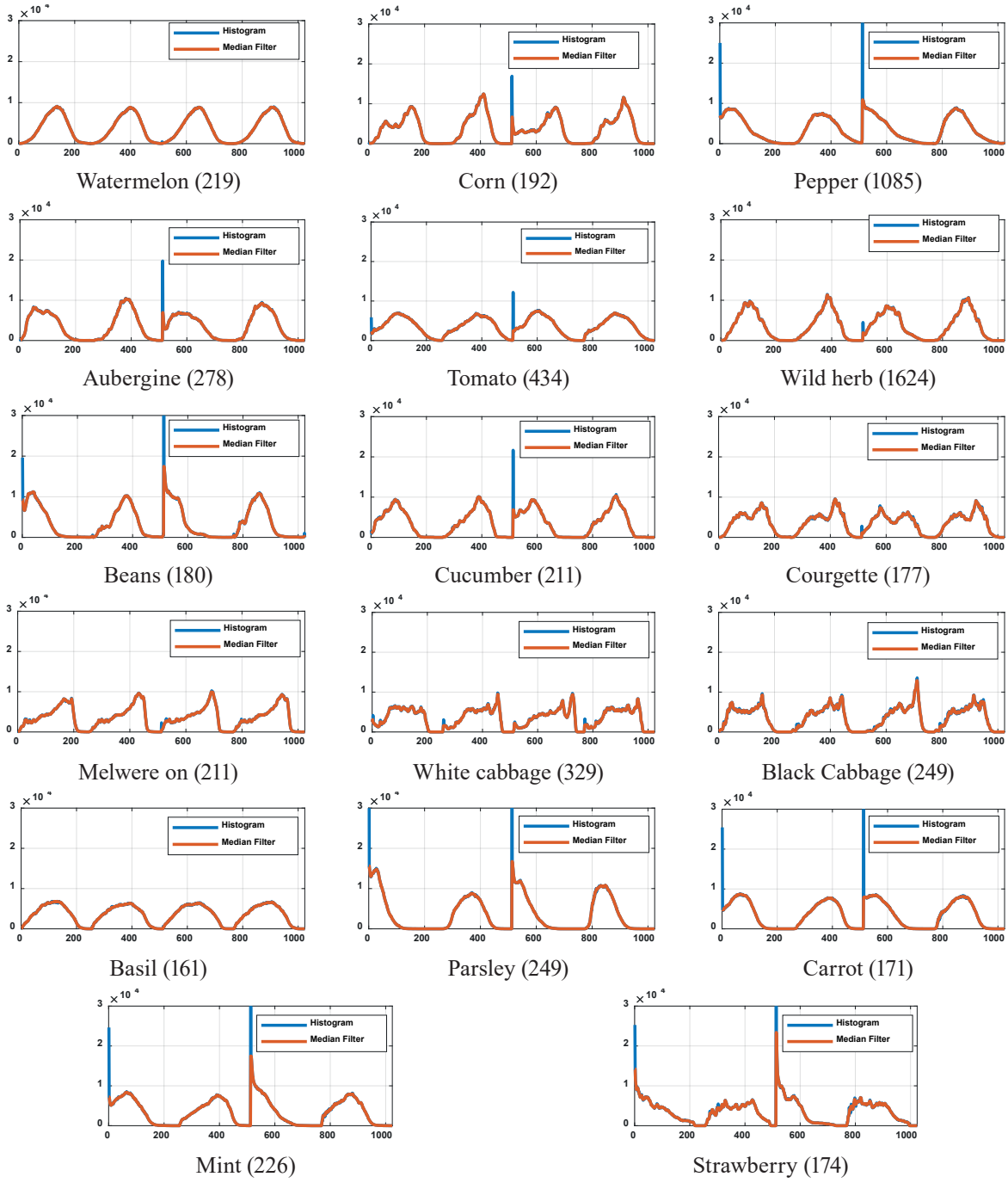


Figure 5. Proposed feature extraction results for images of 17 plant species

As shown in Figure 5, features have been extracted from RGB images through the histogram method. SVM and KNN algorithms have been used to classify the extracted features. SVM and KNN algorithms are commonly preferred machine learning algorithms in the literature (Pooja, Das, & Kanchana, 2018; Yaman, Ertam, Tuncer, & Firat Kilincer, 2020; Yaman & Tuncer, 2021). In order to choose these classifiers (KNN and SVM), five shallow classifiers have been used to test: Decision Tree (DT), Linear Discriminant (LD), Cubic SVM, Fine KNN, and Ensemble Boosted Trees (EBT). The calculated accuracies for these classifiers are depicted in Figure 6.

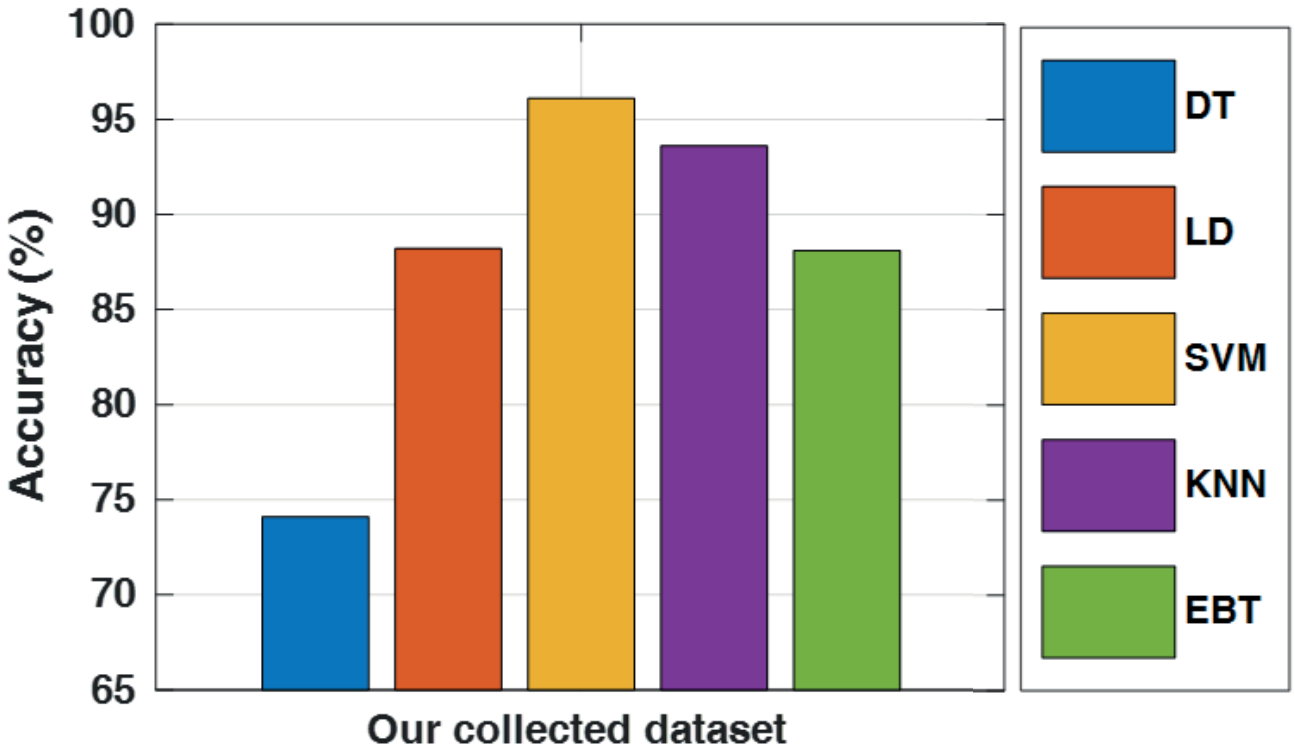
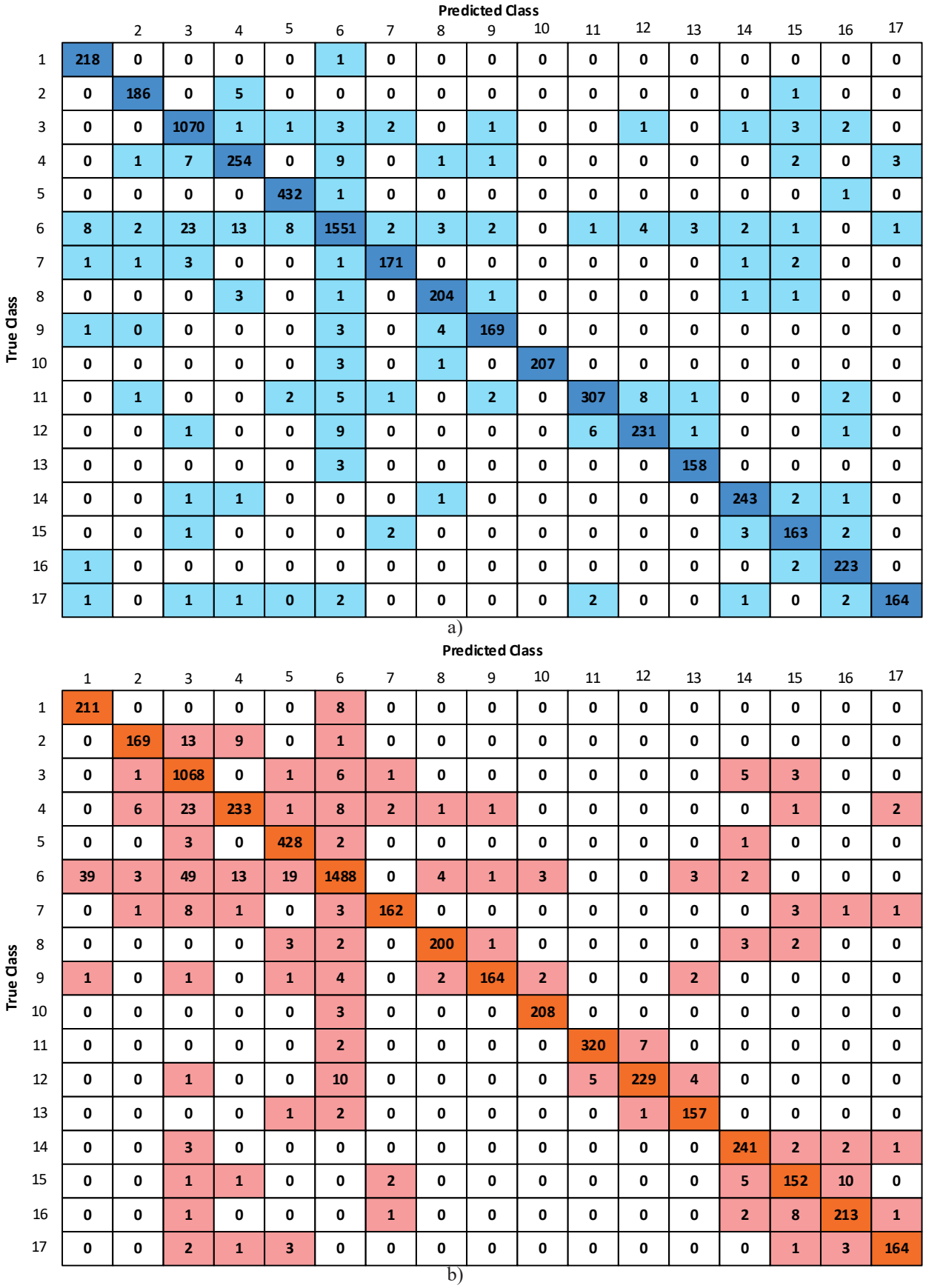


Figure 6. The accuracy results calculated for DT, LD, SVM, KNN, and EBT

As demonstrated in Figure 6, the highest accuracy has been computed with Cubic SVM and Fine KNN for our dataset. For this reason, the features extracted with the histogram have been classified using Cubic SVM and Fine KNN algorithms. 'Box constraint level=1', 'Kernel scale mode=Auto', 'Kernel scale=1' and 'Multiclass method= One-vs-All' are selected in Cubic SVM parameters. 'Number of neighbors=1', 'Distance metric=City block' and 'Distance weight=Equal' are selected in Fine KNN parameters.

3. EXPERIMENTAL RESULTS

In this study, the proposed method has been applied on a personal computer with i7-9700 CPU 3.00 GHz, 32GB RAM, and a 64-bit Windows 10 operating system. In the proposed method, preprocessing and feature extraction steps are developed in the MATLAB 2020a program as m-file. The classification process has been calculated using the MATLAB Classification Learner Toolbox. Confusion matrices have been computed using a 10-fold cross-validation. The confusion matrices are shown in Figure 7.



True Class \ Predicted Class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	218	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
2	0	186	0	5	0	0	0	0	0	0	0	0	0	0	1	0	0
3	0	0	1070	1	1	3	2	0	1	0	0	1	0	1	3	2	0
4	0	1	7	254	0	9	0	1	1	0	0	0	0	0	2	0	3
5	0	0	0	0	432	1	0	0	0	0	0	0	0	0	0	1	0
6	8	2	23	13	8	1551	2	3	2	0	1	4	3	2	1	0	1
7	1	1	3	0	0	1	171	0	0	0	0	0	0	1	2	0	0
8	0	0	0	3	0	1	0	204	1	0	0	0	0	1	1	0	0
9	1	0	0	0	0	3	0	4	169	0	0	0	0	0	0	0	0
10	0	0	0	0	0	3	0	1	0	207	0	0	0	0	0	0	0
11	0	1	0	0	2	5	1	0	2	0	307	8	1	0	0	2	0
12	0	0	1	0	0	9	0	0	0	0	6	231	1	0	0	1	0
13	0	0	0	0	0	3	0	0	0	0	0	0	158	0	0	0	0
14	0	0	1	1	0	0	0	1	0	0	0	0	0	243	2	1	0
15	0	0	1	0	0	0	2	0	0	0	0	0	0	3	163	2	0
16	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	223	0
17	1	0	1	1	0	2	0	0	0	0	2	0	0	1	0	2	164

a)

True Class \ Predicted Class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	211	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0
2	0	169	13	9	0	1	0	0	0	0	0	0	0	0	0	0	0
3	0	1	1068	0	1	6	1	0	0	0	0	0	0	5	3	0	0
4	0	6	23	233	1	8	2	1	1	0	0	0	0	0	1	0	2
5	0	0	3	0	428	2	0	0	0	0	0	0	0	1	0	0	0
6	39	3	49	13	19	1488	0	4	1	3	0	0	3	2	0	0	0
7	0	1	8	1	0	3	162	0	0	0	0	0	0	0	3	1	1
8	0	0	0	0	3	2	0	200	1	0	0	0	0	3	2	0	0
9	1	0	1	0	1	4	0	2	164	2	0	0	2	0	0	0	0
10	0	0	0	0	0	3	0	0	0	208	0	0	0	0	0	0	0
11	0	0	0	0	0	2	0	0	0	0	320	7	0	0	0	0	0
12	0	0	1	0	0	10	0	0	0	0	5	229	4	0	0	0	0
13	0	0	0	0	1	2	0	0	0	0	0	1	157	0	0	0	0
14	0	0	3	0	0	0	0	0	0	0	0	0	0	241	2	2	1
15	0	0	1	1	0	0	2	0	0	0	0	0	0	5	152	10	0
16	0	0	1	0	0	0	1	0	0	0	0	0	0	2	8	213	1
17	0	0	2	1	3	0	0	0	0	0	0	0	0	0	1	3	164

b)

Figure 7. Confusion matrices of the proposed method a) Cubic SVM b) Fine KNN

The class-wise computed with confusion matrices can be seen in Figure 8.

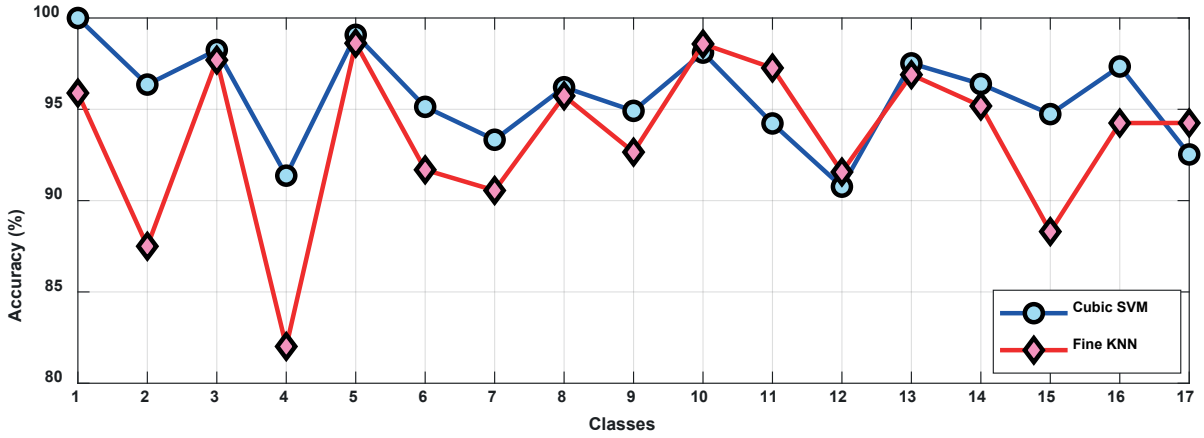


Figure 8. Class wise calculated with our collected dataset

The proposed method ran 100 iterations to calculate Accuracy, Precision, Recall, Geometric mean, and F1-Score parameters. At the end of 100 iterations, the maximum, minimum, mean, and standard deviation results of Accuracy, Precision, Recall, Geometric mean, and F1-Score parameters were computed. The Accuracy, Precision, Recall, Geometric mean, and F1-Score results computed with 100 iterations are tabulated in Table 2.

Table 2
Accuracy, Precision, Recall, Geometric mean, and F1-Score results calculated with 100 iterations

Dataset	Classifiers	Statistics	Accuracy	Precision	Recall	Geometric mean	F1-Score
Our collected dataset	**Cubic SVM**	Maximum	**96.45**	**96.04**	**96.28**	**96.25**	**96.16**
		Minimum	95.78	95.28	95.2	95.16	95.34
		Mean	96.09	95.7	95.82	95.79	95.76
		Standard deviation	0.12	0.16	0.17	0.18	0.15
	Fine KNN	Maximum	**94.11**	**94.09**	**93.83**	**93.75**	**93.92**
		Minimum	93.48	93.11	92.97	92.87	93.04
		Mean	93.82	93.67	93.44	93.33	93.56
		Standard deviation	0.12	0.16	0.17	0.18	0.15

As can be seen in Table 2, a maximum accuracy of 96.45% has been computed with Cubic SVM for our collected dataset. With Fine KNN, accuracy, precision, recall, geometric mean, and F1-Score results have been reckoned as 94.11%, 94.09%, 93.83%, 93.75%, and 93.92% for our collected dataset, respectively. Fold-wise results calculated in the proposed method can be seen in Figure 9.

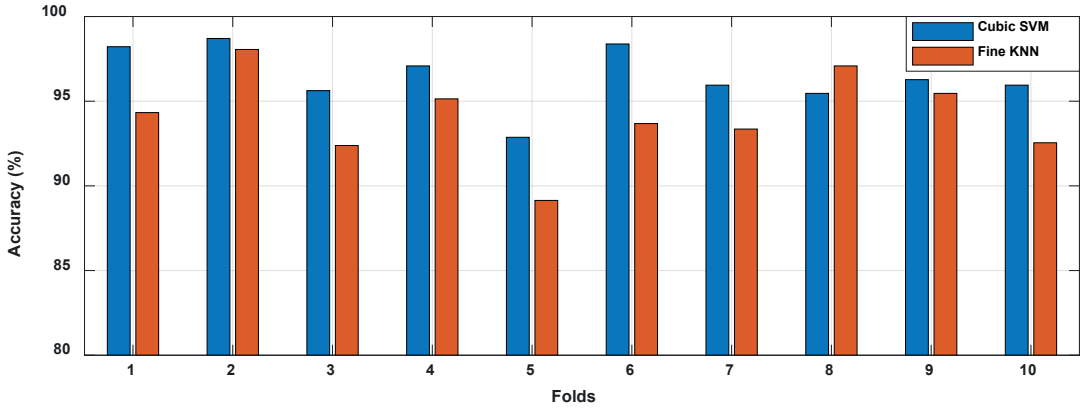


Figure 9. Fold-wise calculated with our collected dataset

The comparison of the proposed method and the literature are summarized in Table 3.

Table 3
Comparison of the proposed method and the literature

Studies	Dataset	Number of species/groups	Method	Accuracy	Precision	Recall	F1-Score
Atila et al. (Atila et al., 2021)	PlantVillage dataset	38	EfficientNet	99.97%	99.39%	-	-
Goyal et al. (Goyal, Kumar, & Gupta, 2021)	Swedish leaf dataset	15	1DPCA+ MTSVM	97.50%	-	-	-
	Flavia dataset	-	1DPCA+ MTSVM	98.15%	-	-	-
	Folio dataset	-	LDA+ MTSVM	89.5%	-	-	-
Selvam et al. (Selvam & Kavitha, 2020)	Self-collected dataset	3	CNN	94%	92%	95%	93%
Murtaza et al. (Murtaza, Saba, Haroon Yousaf, & Viriri, 2020)	Flavia dataset	3	SURF + Bag of Words	98%	-	-	-
	ImageClef dataset	3	SURF + Bag of Words	94%	-	-	-
Adak (Adak, 2020)	Self-collected dataset	43	CNN	87.57%	-	-	-
Saleem et al. (Saleem, Akhtar, Ahmed, & Qureshi, 2019)	Flavia dataset	25	CNN	99.48%	-	-	-
			KNN	98.93%	-	-	-
			Decision Tree	90.65%	-	-	-
			Naïve Bayes	94.5%	-	-	-
			Multi SVM	88.91%	-	-	-
Yacin et al. (Yalcin & Razavi, 2016)	TARBIL dataset	16	CNN	97.47%	-	-	-
			SVM	89.94%	-	-	-
Dyrmann et al. (Dyrmann, Karstoft, & Midtiby, 2016)	BBCH 12e16 dataset	22	CNN	86.2%	-	-	-
Our Methods	Our collected dataset	17	Cubic SVM	96.45%	96.04%	96.28%	96.16%
			Fine KNN	94.11%	94.09%	93.83%	93.92%

To evaluate the performance of the proposed method, feature extraction, training time, test time, and total test time have been computed. The calculated temporal complexity of the method is tabulated in Table 4.

Table 4
The average execution time of the proposed model per the used classifier

Classifiers	Feature extraction	Train time	Test time	Total test time
Cubic SVM	12.09 ms	6.41 ms	20.19 ms	32.28 ms
Fine KNN	12.09 ms	0.38 ms	15.91 ms	28 ms

The execution time of the proposed ultra-lightweight has been calculated for the proposed method in Table 2 and Table 4. It has been observed that a high accuracy has been obtained with SVM and KNN. The time taken for feature extraction and classification from an image (for 1280 × 720 × 3 sized images) has been computed as 32.28ms. 28ms has been reckoned using SVM and KNN respectively. Moreover, the time complexity (using asymptotic notation) of our presented ultra-lightweight model is calculated in Table 5.

Table 5
Time complexity analysis of our proposal

Step	Time complexity (Big O notation)
Layer separation	$O(n)$
Gray level conversion	$O(n)$
Histogram extraction	$O(4n)$
Median filtering	$O(4n)$
Classification using KNN	$O(nd)$
Classification using SVM	$O(nd^3)$
Total complexity using KNN	$O(10n + nd) \cong O(n + nd)$
Total complexity using SVM	$O(10n + nd^3) \cong O(n + nd^3)$

According to Table 5, this model has linear time complexity. Table 4 and 5 proves that our proposal meets the standards of an ultra-lightweight image classification model.

Moreover, these points can be highlighted for this research;

- A novel plant image dataset was acquired using a UAV.

- An ultra-lightweight feature creation method is proposed.

- Per Table 4 and Table 5, our proposal has low time complexity.

- Our ultra-lightweight model has a very simple structure, leading to a wide range of applications by researchers and developers.

- By using the presented ultra-lightweight model and an embedded system, a smart camera can be easily developed for plant species classification.

4. CONCLUSIONS

In this study, an image processing-based ultra-lightweight method is proposed. The proposed method has been developed in the MATLAB program and the results have been calculated. To test the method, 6170 images of 17 plant species have been collected. Features have been created by applying histogram extraction and median filtering functions to these images. For our collected dataset, 96.45% and 94.11% accuracies have been computed with SVM and KNN algorithms, respectively. The time complexity and execution times of our model have also been reckoned and these results show that our model is an ultra-lightweight model.

5. FUTURE WORKS

Classification of plant species and detection of disease in these plants are important for productivity in smart agriculture practices. Growth, disease symptoms, and yields are analyzed by monitoring plants continuously with cameras, thus leading to more efficient production. Real-time and inexpensive real-time implementation of smart agriculture applications is important for its usability. The proposed ultra-lightweight algorithms in future studies will be implemented in embedded systems. Consequently, operations such as automatic recognition of plants, identification of diseased leaves, and spraying will be carried out with cameras and embedded systems on the UAV.

Peer-review: Externally peer-reviewed.
Conflict of Interest: The authors have no conflict of interest to declare.
Grant Support: The authors declared that this study has received no financial support.
Author Contributions: Conception/Design of Study- O.Y., T.T.; Data Acquisition- O.Y.; Data Analysis/Interpretation- O.Y., T.T.; Drafting Manuscript- O.Y., T.T.; Critical Revision of Manuscript- O.Y., T.T.; Final Approval and Accountability- O.Y., T.T.; Material and Technical Support- O.Y., T.T.; Supervision- O.Y., T.T.

REFERENCES

Adak, M. F. (2020). Identification of Plant Species by Deep Learning and Providing as A Mobile Application. *Sakarya University Journal of Computer and Information Sciences*, *3*(3), 231–237. https://doi.org/10.35377/saucis.03.03.773465

Atila, Ü., Uçar, M., Akyol, K., & Uçar, E. (2021). Plant leaf disease classification using EfficientNet deep learning model. *Ecological Informatics*, *61*(October 2020), 101182. https://doi.org/10.1016/j.ecoinf.2020.101182

Babayigit, B., & Büyükpatpat, B. (2019). *Nesnelerin İnterneti Tabanlı Sulama ve Uzaktan İzleme Sisteminin Tasarımı ve Gerçekleştirimi*. 13–19.

Dyrmann, M., Karstoft, H., & Midtiby, H. S. (2016). Plant species classification using deep convolutional neural network. *Biosystems Engineering*, *151*(2005), 72–80. https://doi.org/10.1016/j.biosystemseng.2016.08.024

El, I., Es-saady, Y., El, M., Mammass, D., & Benazoun, A. (2017). Automatic Recognition of Vegetable Crops Diseases based on Neural Network Classifier. *International Journal of Computer Applications*, *158*(4), 48–51. https://doi.org/10.5120/ijca2017912796

Goyal, N., Kumar, N., & Gupta, K. (2021). Lower-dimensional intrinsic structural representation of leaf images and plant recognition. *Signal, Image and Video Processing*. https://doi.org/10.1007/s11760-021-01983-6

Hameed, K., Chai, D., & Rassau, A. (2018). A comprehensive review of fruit and vegetable classification techniques. *Image and Vision Computing*, *80*, 24–44. https://doi.org/10.1016/j.imavis.2018.09.016

Keivani, M., Mazloum, J., Sedaghatfar, E., & Tavakoli, M. B. (2020). Automated analysis of leaf shape, texture, and color features for plant classification.

Traitement Du Signal, *37*(1), 17–28. https://doi.org/10.18280/ts.370103

Murtaza, F., Saba, U., Haroon Yousaf, M., & Viriri, S. (2020). Plant species identification using discriminant bag of words (DBoW). *VISIGRAPP 2020 - Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, *5*, 499–505. https://doi.org/10.5220/0009161004990505

Pawara, P., Okafor, E., Schomaker, L., & Wiering, M. (2017). Data augmentation for plant classification. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10617 LNCS*(September), 615–626. https://doi.org/10.1007/978-3-319-70353-4_52

Pooja, V., Das, R., & Kanchana, V. (2018). Identification of plant leaf diseases using image processing techniques. *Proceedings - 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development, TIAR 2017*, *2018-Janua*(February), 130–133. https://doi.org/10.1109/TIAR.2017.8273700

Saleem, G., Akhtar, M., Ahmed, N., & Qureshi, W. S. (2019). Automated analysis of visual leaf shape features for plant classification. *Computers and Electronics in Agriculture*, *157*(January), 270–280. https://doi.org/10.1016/j.compag.2018.12.038

Selvam, L., & Kavitha, P. (2020). Classification of ladies finger plant leaf using deep learning. *Journal of Ambient Intelligence and Humanized Computing*, (0123456789). https://doi.org/10.1007/s12652-020-02671-y

Wang, J., Yang, J., Yu, L., Dong, H., & Wang, Y. (2021). *DBA_SSD : A Novel End-to-End Object Detection Using Deep Attention Module for 1 Helping Smart Device with Vegetable and Fruit Leaf Plant Disease Detection 2*.

Xie, C., Yang, C., & He, Y. (2017). Hyperspectral imaging for classification of healthy and gray mold diseased tomato leaves with different infection severities. *Computers and Electronics in Agriculture*, *135*, 154–162. https://doi.org/10.1016/j.compag.2016.12.015

Yalcin, H., & Razavi, S. (2016). Plant classification using convolutional neural networks. *2016 5th International Conference on Agro-Geoinformatics, Agro-Geoinformatics 2016*. https://doi.org/10.1109/Agro-Geoinformatics.2016.7577698

Yaman, O., Ertam, F., Tuncer, T., & Firat Kilincer, I. (2020). Automated UHF RFID-based book positioning and monitoring method in smart libraries. *IET Smart Cities*, *2*(4), 173–180. https://doi.org/10.1049/iet-smc.2020.0033

Yaman, O., & Tuncer, T. (2021). Ensemble NASNet Deep Feature Generator Based Underwater Acoustic Classification Model. *Veri Bilimi*, *4*(2), 33–39.

Zhu, X., Zhu, M., & Ren, H. (2018). Method of plant leaf recognition based on improved deep convolutional neural network. *Cognitive Systems Research*, *52*, 223–233. https://doi.org/10.1016/j.cogsys.2018.06.008

ACTA INFOLOGICA 2023;7(1):29-45
dergipark.org.tr/acin

İSTANBUL UNIVERSITY PRESS

DOI: 10.26650/acin.1165378

RESEARCH ARTICLE

Determining Online Travel Planning with AHP and TOPSIS Methods

Çevrimiçi Seyahat Planlamalarının AHP ve TOPSIS Yöntemleri ile Belirlenmesi

Abdullah Eren[1], **Heersh Azeez Khorsheed**[2]

ABSTRACT

Online shopping has become increasingly popular in recent years. Online shopping transactions, which are frequently carried out by consumers all over the world, are also very common in the tourism sector. Users avail themselves of a variety of alternative platforms such as websites, social media or recommendation systems in order to realize their travel plans. Travel transactions can be performed through many applications and platforms. Therefore, it is becoming increasingly important to make the right choice of platform in order to perform faster transactions and make the right decisions. Accordingly, it can sometimes be a difficult process for the user who intends to plan a journey choosing the most suitable online platform from among many alternatives. This study investigated which criteria are important in order to make online travel transactions. In addition, the study included research into which platforms the users can choose in accordance with the determined criteria. Thus, the correct order of the alternatives that people can choose is revealed. In the study, AHP and TOPSIS methods, which are multi-criteria decision-making methods, were preferred. Content quality, usefulness, satisfaction, interaction opportunity, accessibility and web design criteria were used as the main criteria. In addition, sub-criteria of the main criteria were also evaluated. Alternative options were determined such as websites, blogs, Instagram, Facebook, Twitter, Google Comments. The study concludes that the content quality feature is the most important criterion in online travel transactions. Of all the online platforms, websites took the first place among the determined alternatives.

Keywords: Travel planning, AHP, TOPSIS, social media

ÖZ

Online alışveriş günümüzde oldukça popüler hale gelmiştir. Tüm dünyada sıklıkla kullanılan bu işlemler turizm sektöründe de oldukça yaygındır. Kullanıcılar seyahat planlamalarını gerçekleştirmek adına web siteleri, sosyal medya veya öneri sistemleri gibi alternatif platformları kullanabilmektedirler. Bu bağlamda seyahat işlemleri birçok uygulama ve platform üzerinden yapılabilmektedir. Bu yüzden hızlı işlem yapıp doğru kararlar verebilme adına kullanılacak olan platform önem kazanmaktadır. Buna göre seyahat planlaması yapacak olan kullanıcının birçok alternatif içerisinden en uygun olanı seçmesi bazen zor bir süreç olabilmektedir. Bu çalışmada online olarak seyahat işlemlerini yapabilmek adına hangi kriterlerin önemli olduğu araştırılmıştır. Bununla beraber kullanıcıların belirlenmiş kriterler doğrultusunda hangi platformları tercih edebileceği de araştırılmıştır. Böylelikle kişilerin seçeceği alternatifler içerisinden doğru sıralamanın hangisi olduğu ortaya konulmuştur. Çalışmada çok kriterli karar verme yöntemlerinden AHP ve TOPSIS yöntemleri tercih edilmiştir. Çalışmada ana kriterler olarak içerik kalitesi, kullanışlılık, memnuniyet, etkileşim imkanı, erişilebilirlik ve web tasarımı kriterleri kullanılmıştır. Ayrıca ana kriterlerin alt kriterleri de değerlendirilmeye alınmıştır. Alternatif seçenekler ise web siteleri, bloglar, Instagram, Facebook, Twitter, Google Yorumlar olarak belirlenmiştir. Çalışma sonucunda online seyahat işlemlerinde içerik kalitesi özelliği en önemli kriter olmuştur. Belirlenmiş alternatifler içerisinden ilk sırayı web siteleri almıştır.

Anahtar Kelimeler: Seyahat planlaması, AHP, TOPSIS, sosyal medya

[1] (Assist. Prof.) Yuzuncu Yil University, Van Vocational School, Computer Technologies Department, Van, Turkiye

[2](Lect.) Soran University, Faculty of Science, Computer Department, Soran, Erbil, Iraq

ORCID: A.E. 0000-0003-0391-2825; H.A.K. 0000-0002-4468-6871

Corresponding author:
Abdullah EREN
Yuzuncu Yil University, Van Vocational School, Computer Technologies Department, Van, Turkiye
E-mail address: abdullaheren@yyu.edu.tr

Submitted: 22.08.2022
Revision Requested: 18.12.2022
Last Revision Received: 17.02.2023
Accepted: 21.02.2023
Published Online: 14.04.2023

Citation: Eren A. & Khorsheed HA. (2023). Determining online travel planning with AHP and TOPSIS methods. *Acta Infologica, 7*(1), 29-45. https://doi.org/10.26650/acin.1165378

1. INTRODUCTION

In parallel with developments in information and communication technologies, the trend of online shopping has become very popular all over the world. In this context, online tools are frequently used to meet the needs of people. The fact that online shopping has become so easy and convenient has brought innovations into the field of tourism as well as into many other sectors. (Tseng, 2017). Users can carry out their tourism activities through online websites, mobile applications or social networks. They can make plans according to their needs using their chosen online applications or websites. Therefore, research should be conducted into which applications among the many on offer are important for online travel planning. Consideration should be given to the quality elements of the digital platforms preferred by users .

Along with websites, social media platforms are used effectively in determining the travel plans of people. Social media tools play an important role in people's preferences by transferring their travel experiences and impressions to other users (Narangajavana et al., 2017). At the same time, in today's world where social media tools are quite common, tourism content producers also offer their products on social media channels to assist users in making travel plans (Mariani et al., 2017; Rathore et al., 2018). Social media and other web tools are important elements in product marketing and online shopping. With these tools, the company that markets any product and the real user meet in a common digital market. In order for a product to be marketed, the digital platform on which the product is located is expected to meet the expectations of the end user (Hänninen et al., 2018). Sufficient satisfaction of the platform used in product sales affects the marketing of the product and also the preference of the product. This is because the quality of the digital platform on which the product is presented affects the marketing of the product (Wang et al., 2019). In this sense, the quality of the digital platform becomes important. The aim of this study is to determine the importance levels of the factors affecting the quality of online tourism tools used for travel planning and to evaluate six online tourism tools. In this way, the platforms used by customers who carry out tourism activities with online tools will be ranked according to quality.

One of the main expectations of the tourism industry is customer loyalty. The quality of the service provided will affect customer choice of platform (Olorunniwo et al., 2006). If customers are satisfied with the platform they use when making a digital transaction, the expectation is that they will choose the same platform for their next transaction. Accordingly, the following questions were investigated in this study:

1. "Which criteria should be prioritized to assist a customer in choosing a particular digital platform for travel planning?" The aim of this question was to reveal the features that determine the quality of digital platforms used for travel planning. At the same time, the strengths and weaknesses of the digital platform were determined.
2. "Which digital tool is the best for travel planning?" This question was also addressed and evaluated in this study with the aim of creating a decision-making model that would help users choose the right digital platform.

In this study, clear criteria were determined in order to assess the quality of platforms used for travel planning. Multi-Criteria Decision Making-Methods (MCDM) were chosen for ranking these criteria. The Analytic Hierarchy Process (AHP) method, one of the multi-criteria decision-making methods, was used to solve the stated problem. With this method, the criteria are ranked among themselves. Then, the TOPSIS method was used to determine which digital platform is the best for travel planning (Filip, 2014). In this paper we first state the purpose and importance of the study to be conducted. This is followed by a literature review, after which multi-criteria decision-making methods are mentioned. Fourthly, AHP and TOPSIS methods are explained step by step. This is followed by a summary of the solutions developed using the AHP and TOPSIS methods in line with the data obtained from the users. Then the obtained results are presented, and alternative platforms are listed. In the last section, the results are evaluated, and suggestions are made.

2. LITERATURE REVIEW

As the internet has become an indispensable part of life, users have gained the opportunity to perform many transactions online thanks to easy access. According to TURKSTAT (2021) Household Information Technologies (IT) Usage Research data, the rate of internet use among the 16-74 age group in Turkey reached 82.6% as of 2021. Along with the high rate of

internet use, the rate of online shopping, which was 8.4% in 2011, increased to 44.3%. Such data clearly show that online shopping rates have been increasing over recent years. With the widespread use of online shopping, online transactions have also become very common in the tourism sector, as in many other areas as well. Accordingly, many tourism products have become more accessible online in recent years (Lin & Chen, 2013). Content providers in the tourism sector can best present their products to their users through online channels (Choi et al., 2018). Many online tools such as websites, social media platforms, blogs, forum sites, video sharing sites and digital encyclopedias are used effectively to deliver information to users (Gohil, 2015; Žanna, & Xuedong, 2016).

An examination of research conducted in recent years revealed that social media tools play a huge part in people's travel planning (Usui et al., 2018; Pop et al., 2021). According to GuestCentric's report for 2021, the use of social media is recognized as a growing trend for hotels to interact directly with guests and drive direct bookings. 48.75% of hoteliers who participated in the survey conducted in January 2021 see social media as the second most important sales and marketing priority. In this sense, social media has been found to be very important for hotel operators. Moreover, in recent years, recommendation systems have been used in travel planning. Recommendation systems aim to help individuals plan their travel by predicting their preferences and offering suggestions in line with their expectations. Thus, these systems aim to save time for users enabling them to make faster decisions.

The literature shows that multi-criteria decision-making methods are frequently used in research and aim to eliminate uncertainty (Bhole & Deshmukh, 2018). In online transactions, quality rankings of websites, social media tools or web tools can be made using multi-criteria decision-making methods. Kutbi and Alomar (2017) investigated the importance of social media tools in education in their study. Accordingly, video sharing tools were found to be important in education. Putri and Alawiah (2021) used the TOPSIS method to determine the right social media tool in product marketing. Bire et al. (2021) conducted a study to determine the order of importance of the factors affecting the destination choice of tourists. These researchers also used AHP and TOPSIS methods in their studies. As a result of the study, prominent tourism destinations were listed in line with certain criteria.

In the tourism sector, the importance of digital tools used in this field has increased with the increase in transactions, especially on the internet. Baki (2020) conducted a study that evaluates hotel websites according to the criteria of trust, information quality, customer relations, design and price with fuzzy AHP and fuzzy TOPSIS methods. Accordingly, information quality and trust criteria emerged as the most important criteria for hotel websites. Using the AHP-SAW method Sari (2021) tested a specific decision support system to determine the best tourism village . Accordingly, the alternatives most preferred by tourists were determined. Another study used the AHP method to examine the prominent factors in determining the travel preferences of tourists in rural areas (Hussain, & Wang, 2018). According to this study, social media tools were mostly preferred during the travel planning stage. Choedon and Lee (2018) conducted a study to classify and categorize service requirements in mobile tourism applications. Their study used the AHP method to determine the prominent criteria in the mobile travel application experiences of tourists. The study concluded that features such as geographic location map, language option, mobile ticketing, and mobile payment were found to be important in mobile tourism applications.

Various studies have been conducted in this area to improve tourism activities. While planning a trip, several challenges typically need to be addressed such as searching for a hotel, finding the most suitable ticket, and reaching the destination as soon as possible. It can sometimes take a long time to deal with these issues. Vieira (2018) conducted a study to reduce travel costs and optimize travel times. For this, travel data were evaluated using the ant colony optimization method. Thus, a model was developed for optimized travel planning. People also give importance to the opportunities offered by the destinations when planning their trips. In this direction, Gül and Topcu (2015) put forward a model that proposes activities for tourist candidates to spend their holidays more efficiently. In this study, a model was developed using AHP and TOPSIS methods that determines the activities that tourists may want to be engaged in according to their travel experiences . Chen and Wang (2021) conducted a study on travel planning during the Covid-19 period. Accordingly, they developed a model for travel destination recommendation using the fuzzy geometric mean (FGM) approach.

For many people, making travel plans in tourism has become an indispensable part of the whole travel experience. Sometimes, various problems can arise while planning a trip. These problems might include choosing a destination, choosing a hotel, or choosing a vehicle. Various studies have been carried out with the aim of eliminating these problems. Jiaoman et al. (2018) conducted a study to make destination selection more effective during travel procedures. They used fuzzy AHP method in their study on Japanese websites. The study concluded that travel planning design was found to be important. Different techniques have also been used to determine destinations through websites or applications. In their study, Alptekin and Büyüközkan developed an intelligent system proposal using case-based reasoning (CBR) and AHP methods to solve web-based destination determination problems. Combining two different techniques, this study yielded more effective results and the weaknesses of both techniques were eliminated. Do and Shih (2016) conducted a study to solve the destination determination problem during the travel planning phase. In this study, decision-making trial and evaluation laboratory (DEMATEL) and Analytic Network Process (ANP) models were combined. As a result of the research, the external search feature emerged as the most important criterion in determining the destination. As a result of the study tourists can make detailed searches on the web for their travel planning. In addition, they can benefit from the ideas of people who have had the same experience before in order to determine their destination. In this way, the comments on the websites are important in determining the destination. Ahani et al. (2019) discussed the comments on websites as a way of deciding on a hotel for a holiday. In this study, machine learning and TOPSIS methods were used with the aim of classifying tourists according to their satisfaction levels.

Since most travel planning is done in the digital environment, the features or quality of the digital platform used also shape the whole process of travel planning. Ip et al. (2012) conducted a study to determine the functionality of hotel websites. Using the fuzzy method, they listed the information that should be on the hotel websites in order to assist customers with their travel planning. The list includes a description of the hotel and its facilities , information on making a reservation, a description of the surrounding area , and user-generated information. The result of the study showed that the most important feature was the reservation information. A study conducted by Balouchi and Khanmohammadi (2015) investigated the priority features of social media platforms in tourism transactions. This study was conducted with Logarithmic fuzzy preference programming methodology (LFPP) and fuzzy AHP methods. The results of the study showed that, of the three stages when tourists might use social media platforms to assist them in travel arrangements, they mostly prefer to use the platforms before traveling (rather than during or after traveling) . In addition, the most important reason for the preference of these platforms was information sharing. Bire et al. (2021b) aimed to examine tourism supply chain management (TSCM) performance characteristics from the perspective of tourists. A questionnaire was used to obtain data which were evaluated with fuzzy AHP. In the research, the criteria of TSCM (reservation phase, pre-travel phase, travel phase, post-travel phase) were evaluated. According to the results of the research, the pre-travel performance criterion was found to be the most important step according to the tourists. Thus, when the studies in the literature are considered, the problems that require decision-making at the travel planning stage have been resolved. In many of these problems, multi-criteria decision-making methods were used.

3. MATERIALS AND METHODS

In order to carry out this research, data were obtained voluntarily from predetermined participants. The data obtained by the survey method were used in the main study by scoring. In this study, AHP and TOPSIS methods from multi-criteria decision-making methods were used. Ethics committee approval for this study was obtained from the Social and Human Sciences Publication Ethics Committee at Van Yüzüncü Yıl University.

- Meeting Date: 15/10/2021
- Number of Sessions: 2021/15
- Decision No: 2021/15-07
- Document Number: E-85157263-604.01.02-117700

3.1. Multi-Criteria Decision-Making Methods (MCDM)

The decision-making process is the act of selecting one option from a set of choices in order to accomplish a specific objective. While it might sometimes be quite simple to find alternatives in this process, it can also occasionally take on a complex structure (Eren & Kaya, 2019). Essentially, the decision-making stage serves as a problem-solving stage. The elements of intellect, design, and selection serve as the foundation for this problem-solving phase (Aurum & Wohlin, 2003). Multi-criteria decision-making methods (MCDM) methods are used to solve problems when it becomes difficult to make a decision and the options increase. MCDM is accepted as an operations research field that aims to develop and apply decision support tools and methods to deal with decision problems that are not easy to solve, including a specific purpose or conflicting objectives, with many criteria (Akgün and Erdal, 2019). Multi-criteria decision-making enables people to make choices in accordance with their own value judgments in the evaluation of criteria that do not have the same value in conflicting situations (Karaatli et al., 2015). There are many methods for multi-criteria decision-making. These methods have various advantages according to the content and solution of the problem (Ersöz & Kabak, 2010). The aim of this study was to determine the hierarchical order of the criteria and the importance of the alternatives using the AHP and TOPSIS methods, which belong to the multi-criteria decision-making methods.

3.1.2 Method of AHP

The AHP method, developed by Saaty (1980), has been accepted as a new approach to solving problems involving complexity and uncertainty. The AHP method is effectively used in making decisions that include many interrelated and categorized factors. The stage of decision making takes place by transforming user intuitions, abstract and unmeasurable factors and subjectively obtained judgments into a numerical and common basis and integrating them into this process. (Shapira, 2005). In this process, it is aimed for decision makers to compare a number of alternatives in a consistent degree and to reveal the relative priorities of the alternatives within certain ratios (Al-Harbi, 2001).

With AHP, the aim is to create a hierarchical structure among many quantitative and qualitative factors by ordering the specific and uncertain criteria ideally in a complex problem (Badri, 1999). With AHP, the aim is to measure the criteria under a hierarchical structure under user control with the help of analyzes such as problem modeling, calculation of criterion weights and sensitivity analysis (Franek and Kresta, 2014). In order to rank the criteria with the AHP method, the data obtained from the individuals are processed with comparative analysis. The AHP process steps are performed in 5 steps.

Step 1: Creating the Hierarchical Structure

Firstly, the criteria to be determined by the AHP method and the sub-criteria belonging to these criteria are determined. Then, the alternatives supported by the criteria are determined and made into a hierarchical structure. In this way, the meaning of the problem to be decided is provided.

Step 2: Scoring the Criteria

For this step, the scoring scale developed by Saaty (1980) is used (Table 1). Experts perform comparative scoring according to the importance of the criteria.

Table 1
Binary Comparison Scale

Scale Values	**Value Definitions**
1	Equal Importance
3	Moderate Importance
5	Strong Importance
7	Very Strong Importance
9	Extreme Importance
2-4-6-8	Intermediate Values

Step 3: Creating the pairwise comparison matrix

The criteria scored from 1 to 9 by the experts are compared among themselves in pairs. In these comparisons, values are entered into the columns by taking into account the relative importance of the i and j criteria (Table 2).

Table 2
Binary Comparison Matrix

$$A=\begin{bmatrix} 1 & a12 & \dots & a1n \\ \frac{1}{a12} & 1 & \dots & a2n \\ \vdots & \vdots & 1 & \vdots \\ \frac{1}{a1n} & \frac{1}{a2n} & \dots & 1 \end{bmatrix}$$

Step 4: Normalization of binary comparison matrices

In this step, the matrices are normalized by dividing each column value by the sum of the column values it is connected to. The weights of the criteria or sub-criteria are obtained by averaging the rows in the normalized matrix. Thus, the eigenvectors (W) of the criteria are obtained.

Step 5: Determining the Consistency Ratio

In order to calculate the consistency, first of all, the largest eigenvector value (λ max) should be obtained in the pairwise comparison matrix. In order for the obtained matrix to be fully consistent, the eigenvalue (λ max) must be equal to the number of elements (n) compared in the matrix (Saaty, 1990). In order to calculate this value, the values of the comparison matrix A are multiplied by the W column matrix and matrix D is obtained. The basic value E is obtained by dividing the row sums of this matrix by the previously obtained (W) eigenvector column elements. The arithmetic average of the E value gives the value (λ max) (Al-Harbi, 2001). The related equations are obtained as (1) and (2).

$$Ei = \frac{\mathrm{Di}}{\mathrm{Wi}} \quad (1) \tag{1}$$

$$\lambda \max = \frac{\sum_{i=1}^{n} Ei}{n} \tag{2}$$

After calculating the largest eigenvalue (λ max), the consistency ratio (CR) is determined and the consistency of the obtained matrix is tested. Thus, the degree of consistency between the comparisons made between the criteria is determined. Accordingly, the CR value is expected to be less than 0.10. The consistency ratio (CR) is obtained by dividing the CI given in equation (3) by the randomness index (RI) (Shapira, 2005).

$$CR = \frac{\mathrm{CI}}{\mathrm{RI}} \tag{3}$$

The CI value used in the equation is obtained as follows (4). The result is written in equation (3) instead of CI.

$$CI = \frac{\lambda \max - \mathrm{n}}{\mathrm{n}-1} \tag{4}$$

The RI value is obtained from the table below (Table 3).

Table 3
Random Index Values

n	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
RI	0.00	0.00	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49

3.1.3. TOPSIS Method

TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) is a method developed by Hwang and Yoon (1981) for ranking the determined alternatives. This method is preferred in determining the optimal alternatives in complex operations. Accordingly, in the TOPSIS method, the closest distance to the positive ideal solution and the farthest distance to the negative

ideal solution are preferred for the selection of alternatives (Opricovic and Tzeng, 2004). The implementation of the TOPSIS method takes place in 7 stages.

1. Creation of the Standard Decision Matrix

In the first stage, the initial matrix is created with the data obtained by the experts. In order to determine the advantages in this matrix (A), there are alternatives in the row and criteria in the column (1).

$$A_{ij} = \begin{bmatrix} a11 & a12 & \dots & a1n \\ a21 & a22 & \dots & a2n \\ \vdots & \vdots & \vdots & \vdots \\ am1 & am2 & \dots & amn \end{bmatrix} \quad (1)$$

2. Creation of Normalized Decision Matrix

At this stage, n_{ij} values are normalized according to the following formula (2).

$$n_{ij} = \frac{x_{ij}}{\sum_{i=1}^{m} x^2_{ij}} \quad i=1,..,m,\ j=1,..n. \quad (2)$$

3. Creating Weighted Normalized Decision Matrix

The calculation of the weighted normalized decision matrix v_{ij} is shown in the following equation (3).

$$v_{ij} = w_j n_{ij} \quad i=1,..,m,\ j=1,..n. \quad w_j = \text{where wj is the weight of the ith criterion and } \sum_{j=1}^{n} w_j = 1 \quad (3)$$

4. Determining the Ideal Positive and Ideal Negative Solutions

In the V matrix, the column values give the highest values A^+ positive ideal solution values and the smallest values A^- negative ideal solution values. Accordingly, *I* represents the benefit and *J* represents the cost.

$$A^+ = \{ v_1^+, \dots, v_n^+ \} = \{(\max_j vij | i \in I), (\max_j vij | i \in J)\}$$

$$A^- = \{ v_1^-, \dots, v_n^- \} = \{(\min_j vij | i \in I), (\min_j vij | i \in J)\}$$

5. Calculating the Separation Measures

At this stage, it is ensured that the distances to the ideal solutions are expressed. Accordingly, the distance to the positive ideal solution and the distance to the negative ideal solution are obtained by the following equations (4) and (5).

$$s_i^+ = \{\sum_{j=1}^{n} (vij - v_j^+)^2\}^{1/2} \quad (4)$$

$$s_i^- = \{\sum_{j=1}^{n} (vij - v_n^-)^2\}^{1/2} \quad (5)$$

6. Calculating the Relative Closeness to The Ideal Solution

C_i^* value was calculated in order to find the relative closeness to the ideal solution (6).

$$C_i^* = \frac{s_i^-}{s_i^- + s_i^+} \quad 0 \le C_i^* \le 1 \quad (6) \quad (6)$$

7. Ranking the Alternatives

In this step, alternatives with values between 0 and 1 are sorted according to the highest value. According to this solution, the alternative with the largest coefficient supports the best solution.

4. APPLICATION

In order to carry out online travel planning, users make travel plans with the help of many websites, applications, and social media tools. The level of importance of the features of these applications in determining travel planning was investigated in this study. The aim of the study was to determine the quality of the application used to make travel transactions. In this application, depending on the problem identified, literature support and expert opinion support were applied. Then it was

decided by which methods the problem would be solved. Accordingly, AHP and TOPSIS methods were preferred and criteria and sub-criteria were determined. In determining the criteria, support was received from three academicians working in computer science and management information systems, all of whom are experts in their fields. Then, the alternatives were determined. Accordingly, the features to be measured by the AHP method were made into criteria and their importance was ranked. In addition, this study also investigated which application in the digital environment is more important for users wishing to perform their travel transactions. For this, pre-determined applications with the TOPSIS method were turned into alternatives. Finally, these alternatives were listed in order of importance. The path to be followed in choosing a digital application for online travel transactions is shown in Figure 1.

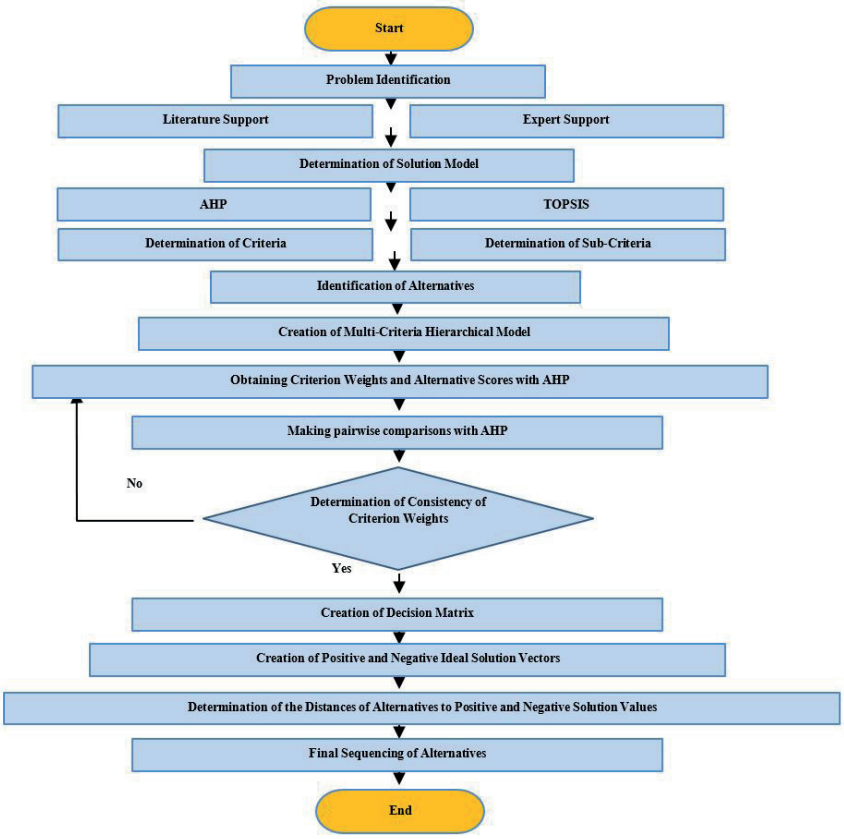


Figure 1. The Way to Choose a Digital Travel App

4.1. Solution with AHP Method

The first step in using the AHP method was that of determining the criteria that should be found in the web tools which perform travel transactions. Accordingly, the main criteria were determined as usability, satisfaction, interaction opportunity, accessibility, content quality and web design. In addition, sub-criteria supporting these criteria were created (Park et al., 2012; Cheng, 2014; Davis, 1989; Wang, 2008; Cheng, 2012; Wang et al., 2007). Once this had been done, alternatives were chosen for the determined criteria, which were Twitter, Instagram, Facebook, Blogs, Websites and Google Comments. The determined criteria, the sub-criteria belonging to these criteria, and the alternatives are shown in Figure 2.

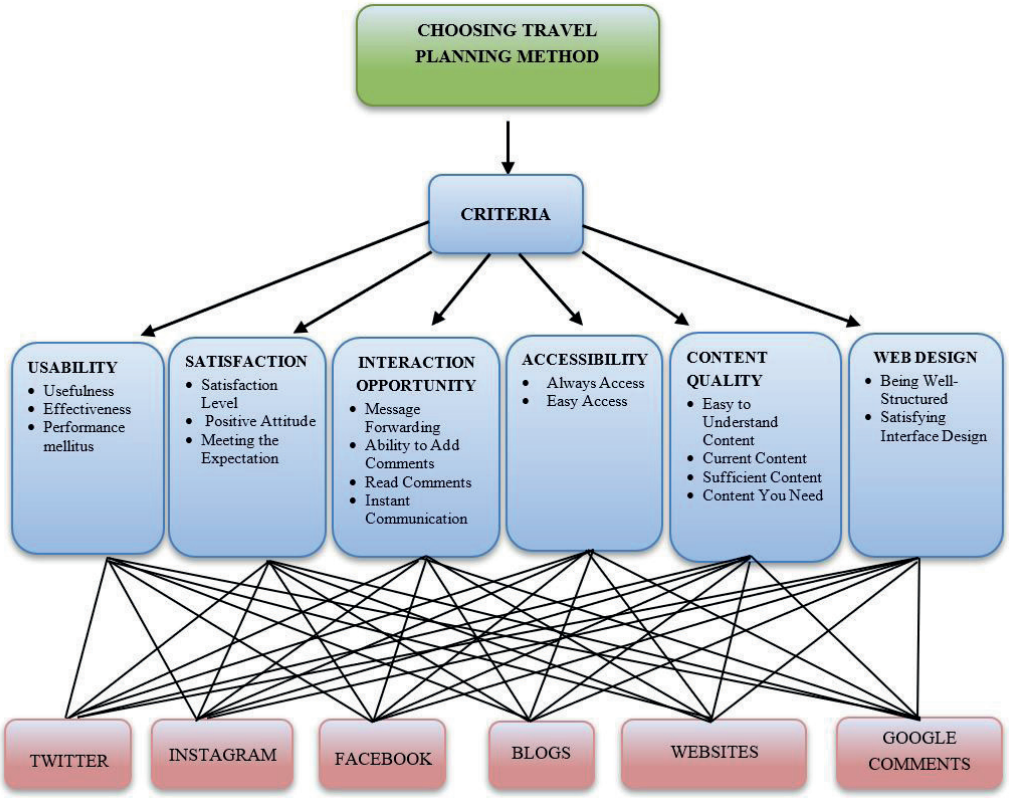


Figure 2. Criteria, Sub-Criteria and Alternatives

4.1.1. Weighting The Criteria

In order to conduct an analysis using AHP, 6 main criteria and 17 sub-criteria were evaluated. Following this, all criteria were converted into expert opinion forms. Sixteen participants, all of whom were academicians, were selected and asked to fill in the expert opinion forms. All the participants had a doctorate degree. Predetermined expert opinion forms were filled in during face-to-face interviews with the participants, all of whom had been selected due to their being frequent users of web tools and social media channels. In addition, all of them had done travel planning using online tools in the previous 5 years. The data obtained from the expert opinion forms were scored and placed in the comparison matrices. Pairwise comparison matrices were created by taking the geometric mean of individual matrix values to ensure the group decision (Saaty, 2008). In the next stages, the consistency of the final criterion weights obtained was tested.

The main criteria were determined as usability (USA), satisfaction (SAT), interaction opportunity (IO), accessibility (ACC), content quality (CQ) and web design (WD). After this, pairwise comparison matrices were created. Relative priorities were determined together with the weights obtained by scoring the criteria. In addition to these, (CR) consistency ratios were found to determine the consistency of the results. Table 4 shows how pairwise comparisons of the criteria were made and their superiority to each other proportioned. The general weights of the criteria were found in line with the determined ratios by normalizing the obtained matrix. It was also found to what extent the eigenvectors were consistent by using the obtained weights.

Table 4
Binary Comparison Matrix

Criteria	Interaction	Satisfaction	Usability	Accessibility	Content Quality	Web Design	Weights
Interaction	1.00	4.12	2.45	5.96	0.45	6.19	0.279
Satisfaction	0.24	1.00	0.37	2.71	0.19	3.66	0.093
Usability	0.41	2.69	1.00	4.43	0.49	4.23	0.175
Accessibility	0.17	0.02	0.23	1.00	0.16	2.71	0.051
Content Quality	2.21	5.14	2.06	6.24	1.00	5.38	0.363
Web Design	0.16	0.27	0.24	0.37	0.19	1.00	0.039
Consistency Rate	0.03						

After obtaining the main criteria, we worked on determining the relative weights of the sub-criteria. For this, the sub-criteria belonging to each criterion were compared with each other, just like the main criteria, and paired comparison matrices were created. Then, with the help of these normalized matrices, the eigenvectors of the sub-criteria were estimated, and their consistency calculated (Table 5). Accordingly, the global weights (W) of the sub-criteria were obtained by multiplying the sub-criteria with the weight of the main criterion to which they were attached. Thus, the final weights of the sub-criteria were revealed, and their importance levels were determined. In addition, all the obtained consistency ratios provided the desired criterion.

Table 5
Calculation of the Weights of the Main Criteria and Sub-Criteria

Criteria	Weights	Sub-Criteria	Weights	Global Weights (W)	Consistency Rate(CR)
Interaction Opportunity (IO)	0.279	IO1-Message Forwarding	0.326	0.091	0.033
		IO2-Ability to Add Comments	0.162	0.045	
		IO3-Ability to Read Comments	0.107	0.030	
		IO4-Instant Communication	0.405	0.113	
Satisfaction (SAT)	0.093	SAT1-Satisfaction Level	0.349	0.033	0.011
		SAT2-Positive Attitude	0.211	0.020	
		SAT3-Meeting the Expectation	0.440	0.041	
Usability (USA)	0.175	USA1-Usefulness	0.476	0.083	0.003
		USA2-Effectiveness	0.315	0.055	
		USA3-Performance	0.208	0.036	
Accessibility (ACC)	0.051	ACC1-Always Access	0.586	0.030	-----
		ACC2-Easy Access	0.414	0.021	
Content Quality (CQ)	0.363	CQ1-Easy to Understand Content	0.418	0.152	0.047
		CQ2-Current Content	0.254	0.092	
		CQ3-Sufficient Content	0.117	0.043	
		CQ4-Content You Need	0.211	0.077	
Web Design (WD)	0.039	WD1-Being Well Structured	0.390	0.015	-----
		WD2-Satisfying Interface Design	0.610	0.024	

5. RANKING WITH THE TOPSIS METHOD

The TOPSIS method was used to rank the alternative options to be used in travel planning. In order to do this, we first created a decision matrix. In the decision matrix, alternatives and options were scored according to all sub-criteria. According to this, the matrix to be created was scored by experts in a way to get values between 1-10. The final decision matrix was obtained by taking the geometric mean of the scores obtained from the experts (Table 6).

Table 6
Decision Matrix

	IO1	IO2	IO3	IO4	WD1	WD2	SAT1	SAT2	SAT3	ACC1	ACC2	CQ1	CQ2	CQ3	CQ4	USA1	USA2	USA3
A1-TWITTER	4.47	4.23	4.73	4.95	5.00	5.00	5.00	4.00	5.00	5.00	6.00	4.00	5.00	4.00	5.00	5.00	4.00	5.00
A2-INSTAGRAM	7.74	8.24	8.21	7.74	6.74	5.96	6.45	6.74	6.74	8.00	8.00	6.48	6.24	7.48	6.70	7.44	6.74	8.00
A3-FACEBOOK	7.74	8.45	8.24	8.21	6.00	6.74	7.00	6.74	6.96	8.24	8.74	6.24	7.24	6.48	7.48	6.00	6.00	6.00
A4-BLOGS	4.00	4.00	3.00	3.00	7.00	8.00	6.00	4.00	5.00	6.00	6.00	5.00	5.00	6.00	3.00	5.00	6.00	5.00
A5-WEBSITES	7.00	7.74	7.24	6.74	7.74	7.74	8.49	8.49	9.00	8.00	8.00	8.00	8.00	7.48	7.74	7.48	7.24	7.24
A6-GOOGLE COMMENTS	6.74	6.40	5.48	5.18	4.00	3.00	6.00	6.00	6.00	8.00	7.00	6.24	6.24	5.44	5.73	6.24	7.00	8,00

After the decision matrix was obtained, the matrix elements were normalized with the help of equation (2) used in the TOPSIS method (Table 7).

Table 7
Normalization Matrix

	IO1	IO2	IO3	IO4	WD1	WD2	SAT1	SAT2	SAT3	ACC1	ACC2	CQ1	CQ2	CQ3	CQ4	USA1	USA2	USA3
A1-TWITTER	0.283	0.255	0.300	0.324	0.329	0.324	0.310	0.263	0.310	0.279	0.333	0.267	0.320	0.261	0.331	0.325	0.261	0.306
A2-INSTAGRAM	0.489	0.498	0.520	0.507	0.443	0.386	0.400	0.443	0.417	0.447	0.444	0.432	0.399	0.488	0.444	0.484	0.440	0.490
A3-FACEBOOK	0.489	0.511	0.522	0.538	0.395	0.436	0.434	0.443	0.431	0.460	0.484	0.416	0.463	0.422	0.496	0.390	0.392	0.368
A4-BLOGS	0.253	0.242	0.190	0.197	0.460	0.518	0.372	0.263	0.310	0.335	0.333	0.334	0.320	0.391	0.199	0.325	0.392	0.306
A5-WEBSITES	0.443	0.467	0.459	0.441	0.509	0.501	0.527	0.559	0.557	0.447	0.444	0.534	0.512	0.488	0.513	0.487	0.472	0.443
A6-GOOGLE COMMENTS	0.426	0.387	0.347	0.339	0.263	0.194	0.372	0.395	0.371	0.447	0.388	0.416	0.399	0.354	0.380	0.406	0.457	0.490

After the normalization matrix was obtained, the global criteria weights (W) obtained by the AHP method were used in the weighted decision matrix. Accordingly, the weighted decision matrix was formed as a result of the multiplication of the normalized criterion values and (W) values (Table 8).

Table 8
Weighted Decision Matrix

	IO1	IO2	IO3	IO4	WD1	WD2	SAT1	SAT2	SAT3	ACC1	ACC2	CQ1	CQ2	CQ3	CQ4	USA1	USA2	USA3
W	0.091	0.045	0.030	0.113	0.015	0.024	0.033	0.020	0.041	0.030	0.021	0.152	0.092	0.043	0.077	0.083	0.055	0.036
A1-TWITTER	0.026	0.012	0.009	0.037	0.005	0.008	0.010	0.005	0.013	0.008	0.007	0.040	0.029	0.011	0.025	0.027	0.014	0.011
A2-INSTAGRAM	0.044	0.022	0.016	0.057	0.007	0.009	0.013	0.009	0.017	0.013	0.009	0.066	0.037	0.021	0.034	0.040	0.024	0.018
A3-FACEBOOK	0.044	0.023	0.016	0.061	0.006	0.010	0.014	0.009	0.018	0.014	0.010	0.063	0.043	0.018	0.038	0.033	0.022	0.013
A4-BLOGS	0.023	0.011	0.006	0.022	0.007	0.012	0.012	0.005	0.013	0.010	0.007	0.051	0.029	0.017	0.015	0.027	0.022	0.011
A5-WEBSITES	0.040	0.021	0.014	0.050	0.008	0.012	0.017	0.011	0.023	0.013	0.009	0.081	0.047	0.021	0.039	0.041	0.026	0.016
A6-GOOGLE COMMENTS	0.039	0.017	0.010	0.038	0.004	0.005	0.012	0.008	0.015	0.013	0.008	0.063	0.037	0.015	0.029	0.034	0.025	0.018

In the next step, A^+ (positive) ideal solution values and A^- (negative) ideal solution values are revealed with the help of a weighted decision matrix. An examination of the column values in the weighted decision matrix reveals that the maximum values are expressed as A^+ and the minimum values are determined as A^- (Wang et al.; 2020). Thus, the positive ideal solution and negative ideal solution sets of the alternatives were created (Table 9).

Table 9
Positive and Negative Ideal Solution Set

A+	0.044	0.023	0.016	0.061	0.008	0.012	0.017	0.011	0.023	0.014	0.010	0.081	0.047	0.021	0.039	0.041	0.026	0.018
A-	0.023	0.011	0.006	0.022	0.004	0.005	0.010	0.005	0.013	0.008	0.007	0.040	0.029	0.011	0.015	0.027	0.014	0.011

After obtaining the positive and negative ideal solution sets, the distances to these solution points were determined. The matrices for the distance to the positive ideal solution () and the distance to the negative ideal solution () are shown in Tables 10 and 11.

Table 10
Distance to Positive Ideal Solution (s_i^+)

	IO1	**IO2**	**IO3**	**IO4**	**WD1**	**WD2**	**SAT1**	**SAT2**	**SAT3**
A1-TWITTER	0.000352	0.000132	0.000044	0.000581	0.000007	0.000021	0.000050	0.000034	0.000103
A2-INSTAGRAM	0.000000	0.000000	0.000000	0.000012	0.000001	0.000010	0.000017	0.000005	0.000033
A3-FACEBOOK	0.000000	0.000000	0.000000	0.000000	0.000003	0.000004	0.000009	0.000005	0.000027
A4-BLOGS	0.000461	0.000147	0.000098	0.001485	0.000001	0.000000	0.000025	0.000034	0.000103
A5-WEBSITES	0.000018	0.000004	0.000004	0.000119	0.000000	0.000000	0.000000	0.000000	0.000000
A6-GOOGLE COMMENTS	0.000033	0.000031	0.000027	0.000502	0.000014	0.000058	0.000025	0.000010	0.000058
	ACC1	**ACC2**	**CQ1**	**CQ2**	**CQ3**	**CQ4**	**USA1**	**USA2**	**USA3**
A1-TWITTER	0.000029	0.000010	0.001635	0.000313	0.000093	0.000194	0.000182	0.000136	0.000045
A2-INSTAGRAM	0.000000	0.000001	0.000236	0.000108	0.000000	0.000028	0.000000	0.000003	0.000000
A3-FACEBOOK	0.000000	0.000000	0.000318	0.000020	0.000008	0.000002	0.000065	0.000020	0.000020
A4-BLOGLAR	0.000014	0.000010	0.000920	0.000313	0.000017	0.000581	0.000182	0.000020	0.000045
A5-WEBSITES	0.000000	0.000001	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000003
A6-GOOGLE COMMENTS	0.000000	0.000004	0.000318	0.000108	0.000032	0.000104	0.000046	0.000001	0.000000

Table 11
Distance To Negative Ideal Solution (s_i^-)

	IO1	**IO2**	**IO3**	**IO4**	**WD1**	**WD2**	**SAT1**	**SAT2**	**SAT3**
A1-TWITTER	0.000007	0.000000	0.000011	0.000208	0.000001	0.000009	0.000000	0.000000	0.000000
A2-INSTAGRAM	0.000461	0.000133	0.000097	0.001229	0.000007	0.000020	0.000009	0.000013	0.000019
A3-FACEBOOK	0.000461	0.000147	0.000098	0.001485	0.000004	0.000033	0.000016	0.000013	0.000025
A4-BLOGS	0.000000	0.000000	0.000000	0.000000	0.000009	0.000058	0.000004	0.000000	0.000000
A5-WEBSITES	0.000297	0.000104	0.000064	0.000764	0.000014	0.000052	0.000050	0.000034	0.000103
A6-GOOGLE COMMENTS	0.000247	0.000043	0.000022	0.000260	0.000000	0.000000	0.000004	0.000007	0.000006
	ACC1	**ACC2**	**CQ1**	**CQ2**	**CQ3**	**CQ4**	**USA1**	**USA2**	**USA3**
A1-TWITTER	0.000000	0.000000	0.000000	0.000000	0.000000	0.000103	0.000000	0.000000	0.000000
A2-INSTAGRAM	0.000025	0.000006	0.000629	0.000053	0.000093	0.000354	0.000176	0.000097	0.000045
A3-FACEBOOK	0.000029	0.000010	0.000511	0.000174	0.000047	0.000520	0.000029	0.000052	0.000005
A4-BLOGS	0.000003	0.000000	0.000102	0.000000	0.000031	0.000000	0.000000	0.000052	0.000000
A5-WEBSITES	0.000025	0.000006	0.001635	0.000313	0.000093	0.000581	0.000182	0.000136	0.000025
A6-GOOGLE COMMENTS	0.000025	0.000001	0.000511	0.000053	0.000016	0.000193	0.000045	0.000117	0.000045

The final () and () values of each alternative in the decision stage of the TOPSIS method were determined by equations (4) and (5) on TOPSIS. Finally, with the help of the values obtained for the alternatives, the value of was reached by using equation (6) (Table 12). Thus, with the help of the value, the relative closeness of the alternatives to the ideal solution was determined. In this way, the final ranking of the alternatives is also revealed.

Table 12
Closeness Of Alternatives to The Ideal Solution and Final Ranking

	s_i^+	s_i^-	C_i^*	Ranking
A1-TWITTER	0.062946	0.018450	0.226674	5
A2-INSTAGRAM	0.021318	0.058888	0.734204	2
A3-FACEBOOK	0.022358	0.060501	0.730166	3
A4-BLOGS	0.066750	0.016092	0.194254	6
A5-WEBSITES	0.012161	0.066914	0.846205	1
A6-GOOGLE COMMENTS	0.037049	0.039947	0.518822	4

6. DISCUSSION AND CONCLUSION

In today's world, where online shopping trends are at a very high level, internet users aim to carry out their shopping requirements in the best possible way. One of the areas where online transactions are widely used is travel planning. Many users use online platforms to carry out their travel planning or holiday transactions. Of course, knowing which platform is more convenient than others is also important. In addition, users want to benefit from the features offered by the online platforms they use for their travel planning. In this sense, the expectations of users from the platforms they use can also affect the quality of shopping. For some customers, accessibility is more important, while for others, factors such as speed or content quality are a priority. Thus, the aim of this study was to conduct research into which platform is effective in online travel planning and, accordingly, which criteria are prioritized on alternative platforms.

In this study, platforms that can perform online travel planning processes and the quality criteria of these platforms were evaluated. The AHP and TOPSIS methods, which are multi-criteria decision-making methods, were chosen for the study, and the data obtained from users were evaluated. Accordingly, the main criteria which formed the basis of this study were usability, satisfaction, interaction possibility, content quality, accessibility and web design criteria. Sub-criteria were listed using the AHP method. As a result of the evaluation, content quality was seen to be the most important criterion with 36.3%, and this criterion was followed by the possibility of interaction with 27.9%. The usefulness criterion had a priority of 17.5% and the satisfaction criteria had a priority value of 9.3%. While system accessibility had a priority of 5.1%, the web design criterion took last place with a rate of 3.9%. Our evaluation of the sub-criteria showed that the easy-to-understand content criterion, which is the content quality sub-criterion (CQ1), took first place with 41.7%. An examination of the sub-criteria of the interaction possibility criterion revealed that the instant communication feature took first place with 40.5% (IO4).

In the study, the best alternatives meeting the criteria were determined. Alternatives were listed using the TOPSIS method in order to determine the best platform for performing travel planning processes. Our results showed that websites have become the most preferred alternative in travel planning. Instagram and Facebook applications followed this alternative with very close values, respectively. The Google comments alternative took 4th place, with the Twitter application following it in 5th place. In this study, the blogs alternative took the last place.

An evaluation of the result of the study shows that the criterion of quality of content was found to be the most important criterion in the digital platforms preferred for online travel planning. In previous studies content quality is also called information quality (Delone & Mclean, 2003). Accordingly, in order for the information of any web application to be of high quality, it is expected that it should be presented in an easy, understandable, complete, appropriate, and up-to-date way, while also meeting the needs of customers (Delone & Mclean, 2003; Sirsat & Sirsat, 2016; Rai et al.; 2002). When the literature is examined, information quality plays an important role in determining the success of information systems (Lee et al.; 2002; Petter et al.; 2008). Thus, the fact that the content quality criterion is a priority feature in the current study supports the literature within the framework of information systems. The quality of the information obtained from any information system also positively affects the satisfaction obtained from the system and the success of the system (Mardiana et al.; 2015; Petter et al.; 2008; Zviran & Erlich, 2003). If there is sufficient and satisfactory content in information systems, users prefer to use that system (Alshibly, 2014; Delone & Mclean, 1992; Urbach & Müller, 2012). The current study also shows that content quality is the most important factor in increasing the quality of a digital platform used in tourism or travel transactions.

When studies conducted with multi-criteria decision-making methods are evaluated, it is clear that the quality of information has priority in the research. Accordingly, Lin (2010) reported that content quality is the most important criterion in the learning system in his study on web-based learning systems using fuzzy AHP method. Arora and Gupta (2017) examined the features of the system in their study on e-government systems with the AHP method. According to this study, content quality was found to be the most important criterion among other system features. Baki (2020) evaluated hotel websites in terms of quality with fuzzy AHP and fuzzy TOPSIS methods. The results showed that, when the criteria of trust, information quality, customer relations, design and price are taken into consideration, the criterion of trust takes the first place, while quality of information takes the second place. Tseng et al. (2021) developed a model for the selection of 3rd party travel reservation systems in their study. The study determined which criteria were important in the selection of the best tourism website with the AHP method. As a result of the research, while the security criterion among the five main criteria took the first place, the information quality criterion took the second place. The current study also shows that quality of the content is an important criterion for travel planning. In this respect, the result of the present study largely supports previous studies. According to the results of this research, it can be said that digital platforms with quality content will create more satisfaction on users. This is because the content that is sufficient and meets expectations affects platform users. Thus, the quality of the content created by the content providers gains importance in terms of guiding the customers. In this direction, the aim of content providers should be to provide better service to users by improving the content offered by the platforms in order for customers to plan travel transactions. In particular, if the content is up-to-date, comprehensible, meeting expectations and reliable customers will be affected positively. In this way, customer loyalty will be achieved as the expectations and needs of the customers are met. As a result of the study, the factors affecting the quality of digital platforms for travel planning were revealed. To this end, a solution model was developed for the problem of appropriate platform selection.

There are certain limitations within the scope of this study. The opinions of 16 experts in the field were evaluated. This number could be increased in future studies. In addition, different results may be obtained if this study were tested on generation Z or other age groups. This is because the use of social networks varies according to various age groups (Thuy and Duy, 2021). Different results may emerge in studies on the younger population, for example (Mamula-Nikolić et al.; 2022). In addition, the participants in the current study were determined from among academicians. This study could be repeated on different business groups or people with different incomes. However, more definitive results could be obtained if these and similar studies were also performed using new techniques such as fuzzy AHP, fuzzy TOPSIS, ANP, PIPRECIA and fuzzy MARCOS (Stević et al.; 2021; Sun, 2010; Chowdhury & Paul, 2020). In this study, the preferred platforms are limited to only six. In future studies, more social media platforms or digital content providers could be tested. In this study, six criteria were used to determine the quality of online tourism tools. In future studies, these criteria could be changed according to different dimensions or perspectives. Influencers or advertisements could also be considered as effective criteria in determining travel planning. As a result of the study, although the websites were chosen as the best platform, it should be taken into account that Instagram and Facebook platforms also receive very close values. Accordingly, managers in the sales and marketing sector are likely to turn to social media platforms in order to gain customers. In addition, in this study, the criteria of content quality and interaction possibility were shown to be the most important criteria in travel planning. Thus, improving the content on the platforms will create satisfaction on the users. Since information sharing is very popular in today's world, it will be important for users to have an interactive structure on digital platforms.

Peer-review: Externally peer-reviewed.
Conflict of Interest: The authors have no conflict of interest to declare.
Grant Support: The authors declared that this study has received no financial support.
Author Contributions: Conception/Design of Study- A.E., H.A.K.; Data Acquisition- A.E.; Data Analysis/Interpretation- A.E.; Drafting Manuscript- H.A.K.; Critical Revision of Manuscript- A.E., H.A.K.; Final Approval and Accountability- A.E.; Material and Technical Support- A.E., H.A.K.; Supervision- A.E., H.A.K.

REFERENCES

Ahani, A., Nilashi, M., Yadegaridehkordi, E., Sanzogni, L., Tarik, A. R., Knox, K., ... & Ibrahim, O. (2019). Revealing customers' satisfaction and preferences through online review analysis: The case of Canary Islands hotels. Journal of Retailing and Consumer Services, 51, 331-343.

Al-Harbi, K. M. A. S. (2001). Application of the AHP in project management. International journal of project management, 19(1), 19-27.

Akgün, İ., & Erdal, H. (2019). Solving an ammunition distribution network design problem using multi-objective mathematical modeling, combined AHP-TOPSIS, and GIS. Computers & Industrial Engineering, 129, 512-528.

Alptekin, G. I., & Büyüközkan, G. (2011). An integrated case-based reasoning and MCDM system for Web based tourism destination planning. Expert Systems with Applications, 38(3), 2125-2132.

Alshibly, H. H. (2014). Evaluating E-HRM success: A validation of the information systems success model. International Journal of Human Resource Studies, 4(3), 107-124.

Arora, M., & Gupta, A. (2017). E-Governance Parameters Assessment: An AHP Approach. Indian Journal of Computer Science, 2(1), 16-22.

Aurum, A., & Wohlin, C. (2003). The fundamental nature of requirements engineering activities as a decision-making process. Information and Software Technology, 45(14), 945-954.

Badri, M. A. (1999). Combining the analytic hierarchy process and goal programming for global facility location-allocation problem. International journal of production economics, 62(3), 237-248.

Baki, R. (2020). Evaluating hotel websites through the use of fuzzy AHP and fuzzy TOPSIS. International Journal of Contemporary Hospitality Management.

Balouchi, M., & Khanmohammadi, E. (2015). Using logarithmic fuzzy preference programming to prioritization social media utilization based on tourists' perspective. Foundations of Management, 7(1), 7-18.

Bhole, G. P., & Deshmukh, T. (2018). Multi-criteria decision making (MCDM) methods and its applications. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 6(5), 899-915.

Bire, R. B., Conterius, A. L., & Nasar, A. (2021, April). Tourist preference and destination competitiveness using the AHP–TOPSIS hybrid model. In International Conference on Applied Science and Technology on Social Science (ICAST-SS 2020) (pp. 43-48). Atlantis Press.

Bire, R. B., Nugraha, Y. E., & Welly, F. R. A. (2021). A fuzzy-analytic hierarchy process of tourism supply chain performance: customer perspectives. Enlightening Tourism. A Pathmaking Journal, 11(2), 531-557.

Chaudhari, K., & Thakkar, A. (2020). A comprehensive survey on travel recommender systems. Archives of Computational Methods in Engineering, 27(5), 1545-1571.

Chen, C. W. D., & Cheng, C. Y. J. (2009). Understanding consumer intention in online shopping: a respecification and validation of the DeLone and McLean model. Behaviour & Information Technology, 28(4), 335-345.

Chen, T., & Wang, Y. C. (2021). A calibrated piecewise-linear FGM approach for travel destination recommendation during the COVID-19 pandemic. Applied Soft Computing, 109, 107535.

Cheng, Y. M. (2012). Effects of quality antecedents on e-learning acceptance. Internet Research.

Cheng, Y. M. (2014). Why do users intend to continue using the digital library? An integrated perspective. Aslib Journal of Information Management.

Choedon, T., & Lee, Y. C. (2018). Classification and evaluation of service requirements in mobile tourism application using Kano model and AHP. The Journal of Information Systems, 27(1), 43-65.

Choi, M., Law, R., & Heo, C. Y. (2018). An investigation of the perceived value of shopping tourism. Journal of travel Research, 57(7), 962-980.

Chowdhury, P., & Paul, S. K. (2020). Applications of MCDM methods in research on corporate sustainability: A systematic literature review. Management of Environmental Quality: An International Journal.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS quarterly, 319-340.

DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. Information systems research, 3(1), 60-95.

DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: a ten-year update. Journal of management information systems, 19(4), 9-30.

De Wulf, K., Schillewaert, N., Muylle, S., & Rangarajan, D. (2006). The role of pleasure in web site success. Information & Management, 43(4), 434-446.

Do, T. H. N., & Shih, W. (2016). Destination decision-making process based on a hybrid MCDM model combining DEMATEL and ANP: The Case of Vietnam as a Destination. Modern Economy, 7(09), 966.

Eren, A., & Kaya, M. D. (2019). İş Zekâsı Sistemlerinde Karar Verme Başarısının İncelenmesi. Business & Management Studies: An International Journal, 7(5), 2148-2176.

Ersöz, F., & Kabak, M. (2010). Savunma sanayi uygulamalarinda çok kriterli karar verme yöntemlerinin literatür araştirmasi. Savunma Bilimleri Dergisi, 9(1), 97-125.

Filip, F. G. (2014). A decision-making perspective for designing and building information systems. International Journal of Computers Communications & Control, 7(2), 264-272.

Franek, J., & Kresta, A. (2014). Judgment scales and consistency measure in AHP. Procedia Economics and Finance, 12, 164-173.

Gohil, N. (2015). Role and impact of social media in tourism: a case study on the initiatives of Madhya Pradesh State Tourism. International Journal of Research in Economics and Social Sciences, 5(4), 8-15.

GuestCentric (2021). Oteller için Sosyal Medya, 2021, Retrieved September 1, 2021, from https://blog.guestcentric.com/social-media-for-hotels-is-it-worth-it/

Gul, S., & Topcu, I. (2015). A multi-attribute decision support model for the selection of touristic activities. International Journal of the Analytic Hierarchy Process, 7(3).

Hänninen, M., Smedlund, A., & Mitronen, L. (2018). Digitalization in retailing: multi-sided platforms as drivers of industry transformation. Baltic Journal

of Management, 13(2), 152-168.
Hussain, T., Li, B., & Wang, D. (2018). What factors influence the sustainable tour process in social media usage? Examining a rural mountain region in Pakistan. Sustainability, 10(7), 2220.
Hwang, C. L., & Yoon, K. (1981). Methods for multiple attribute decision making. In Multiple attribute decision making (pp. 58-191). Springer, Berlin, Heidelberg.
Ip, C., Law, R., & Lee, H. A. (2012). The evaluation of hotel website functionality by fuzzy analytic hierarchy process. Journal of Travel & Tourism Marketing, 29(3), 263-278.
Jiaoman, D., Lei, L., & Xiang, L. (2018, October). Travel planning problem considering site selection and itinerary making. In Proceedings of the 2018 Conference on Research in Adaptive and Convergent Systems (pp. 29-36).
Karaatli, M., Ömürbek, N., Budak, İ., & Okan, D. A. Ğ. (2015). Çok kriterli karar verme yöntemleri ile yaşanabilir illerin sıralanması. Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, (33), 215-228.
Kutbi, A., & Alomar, M. (2017). The Importance of Social Media in Education and Prioritizing Process. Journal of King Abdulaziz University, 28(2), 27-37.
Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: a methodology for information quality assessment. Information & management, 40(2), 133-146.
Lin, C. C. (2021). Investigating complimentary e-marketing strategy for small-and medium-sized enterprises at growth stage in taiwan. Information, 12(9), 380.
Lin, H. F. (2010). An application of fuzzy AHP for evaluating course website quality. Computers & Education, 54(4), 877-888.
Lin, Y. H., & Chen, C. F. (2013). Passengers' shopping motivations and commercial activities at airports–The moderating effects of time pressure and impulse buying tendency. Tourism Management, 36, 426-434.
Mamula-Nikolić, T., Perčić, K., & Nećak, M. (2022). Designing a sustainable brand by engaging brand-consumers' collaboration Generation Z co-creators and influencers in Serbia. Marketing, 53(1), 37-50.
Mardiana, S., Tjakraatmadja, J. H., & Aprianingsih, A. (2015). DeLone–McLean information system success model revisited: The separation of intention to use-use and the integration of technology acceptance models. International Journal of Economics and Financial Issues, 5(1), 172-182.
Mariani, M. M., Mura, M., & Di Felice, M. (2018). The determinants of Facebook social engagement for national tourism organizations' Facebook pages: A quantitative approach. Journal of Destination Marketing & Management, 8, 312-325.
Narangajavana, Y., Fiol, L. J. C., Tena, M. Á. M., Artola, R. M. R., & García, J. S. (2017). The influence of social media in creating expectations. An empirical study for a tourist destination. Annals of Tourism Research, 65, 60-70.
Nazarail, N. N., & Kasim, Z. (2021). The selection of social networking sites using fuzzy analytical hierarchy process. ESTEEM Academic Journal, 17, 1-11.
Olorunniwo, F., Hsu, M. K., & Udo, G. J. (2006). Service quality, customer satisfaction, and behavioral intentions in the service factory. Journal of services marketing, 20(1), 59-72.
Opricovic, S., & Tzeng, G. H. (2004). Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. European journal of operational research, 156(2), 445-455.
Park, S. Y., Nam, M. W., & Cha, S. B. (2012). University students' behavioral intention to use mobile learning: Evaluating the technology acceptance model. British journal of educational te Prabantoro, G., & Hariyanto, J. (2019, February). Social Media Preference as a Media of Business Marketing Communication of SMEs in Rawamangun East Jakarta. In 5th Annual International Conference on Management Research (AICMaR 2018) (pp. 165-169). Atlantis Press.chnology, 43(4), 592-605.
Petter, S., DeLone, W., & McLean, E. (2008). Measuring information systems success: models, dimensions, measures, and interrelationships. European journal of information systems, 17, 236-263.
Pop, R. A., Săplăcan, Z., Dabija, D. C., & Alt, M. A. (2021). The impact of social media influencers on travel decisions: The role of trust in consumer decision journey. Current Issues in Tourism, 1-21.
Putri, D. A., & Alawiah, E. T. (2021). Decision Support System To Choose The Best Social Media Platform For Product Marketing Using Topsis Method. IJISTECH (International Journal of Information System and Technology), 5(4), 394-399.
Rai, A., Lang, S. S., & Welker, R. B. (2002). Assessing the validity of IS success models: An empirical test and theoretical analysis. Information systems research, 13(1), 50-69.
Rathore, A. K., Joshi, U. C., & Ilavarasan, P. V. (2017). Social media usage for tourism: A case of Rajasthan tourism. Procedia computer science, 122, 751-758.
Saaty, T. L. (2008). Decision making with the analytic hierarchy process. Int. J. Services Sciences, 1(1), 83-98.
Saaty, T. L. (1990). How to make a decision: the analytic hierarchy process. European journal of operational research, 48(1), 9-26.
Saaty Thomas, L. (1980). The analytic Hierarchy process. New York: McGrow-Hill.

Sari, N. K. A. P. (2021). Implementation of the AHP-SAW method in the decision support system for selecting the best tourism village. Jurnal Teknik Informatika CIT Medicom, 13(1), 24-35.
Setiawan, B., Trisdyani, N. L. P., Adnyana, P. P., Adnyana, I. N., Wiweka, K., & Wulandani, H. R. (2018). The profile and behaviour of 'digital tourists' when making decisions concerning travelling case study: Generation Z in South Jakarta. Advances in research, 17(2), 1-13.
Shapira, A., & Goldenberg, M. (2005). AHP-based equipment selection model for construction projects. Journal of construction engineering and

management, 131(12), 1263-1273.

Sirsat, S. S., & Sirsat, M. S. (2016). A validation of the DeLone and McLean model on the educational information system of the Maharashtra State (India). International Journal of Education and Learning Systems, 1.

Stević, Ž., Karamaşa, Ç., Demir, E., & Korucuk, S. (2021). Assessing sustainable production under circular economy context using a novel rough-fuzzy MCDM model: A case of the forestry industry in the Eastern Black Sea region. Journal of Enterprise Information Management.

Sun, C. C. (2010). A performance evaluation model by integrating fuzzy AHP and fuzzy TOPSIS methods. Expert systems with applications, 37(12), 7745-7754.

Tang, C., & Ngerng, M. (2015). How do students select social networking sites? An analytic hierarchy process (AHP) model. International Journal of Cyber Society and Education, 8(2), 81-98.

Thuy, H. T. H., Hieu, V. M., & Duy, B. P. M. (2021). Intention to use facebook for travel planning: An investigation on generation Z users in Ho Chi Minh city. Webology, 18(Special Issue on Information Retrieval and Web Search), 92-106.

Tseng, A. (2017). Why do online tourists need sellers' ratings? Exploration of the factors affecting regretful tourist e-satisfaction. Tourism Management, 59, 413-424.

Tseng, T. H., Wang, Y. S., & Tsai, Y. C. (2021). Applying an AHP Technique for Developing A Website Model of Third-Party Booking System. Journal of Hospitality & Tourism Research, 45(8), 1440-1463.

TURKSTAT (2021). Hanehalkı Bilişim Teknolojileri (BT) Kullanım Araştırması, 2021, Retrieved September 5, 2022, from https://data.tuik.gov.tr/Bulten/Index?p=Hanehalki-Bilisim-Teknolojileri-(BT)-Kullanim-Arastirmasi-2021-37437

Urbach, N., & Müller, B. (2012). The updated DeLone and McLean model of information systems success. Information Systems Theory: Explaining and Predicting Our Digital Society, Vol. 1, 1-18.

Usui, R., Wei, X., & Funck, C. (2018). The power of social media in regional tourism development: a case study from Ōkunoshima Island in Hiroshima, Japan. Current Issues in Tourism, 21(18), 2052-2056.

Vieira, G. L. (2018). Trip-Planning Optimization: Minimizing Cost and Travel Time in Itineraries with Multiple Destinations (Doctoral dissertation, MSc. Thesis., School of Engineering, Federal University of Minas Gerais).

Wang, L., Ali, Y., Nazir, S., & Niazi, M. (2020). ISA evaluation framework for security of internet of health things system using AHP-TOPSIS methods. IEEE Access, 8, 152316-152332.

Wang, Y. Y., Wang, Y. S., Lin, H. H., & Tsai, T. H. (2019). Developing and validating a model for assessing paid mobile learning app success. Interactive Learning Environments, 27(4), 458-477.

Wang, Y. S. (2008). Assessing e-commerce systems success: a respecification and validation of the DeLone and McLean model of IS success. Information systems journal, 18(5), 529-557.

Wang, Y. S., Wang, H. Y., & Shee, D. Y. (2007). Measuring e-learning systems success in an organizational context: Scale development and validation. Computers in Human Behavior, 23(4), 1792-1808.

Žanna, M., & Xuedong, G. (2016). DMO tourism website's success evaluation model and framework. European journal of economics and management sciences, (4).

Zviran, M., & Erlich, Z. (2003). Measuring IS user satisfaction: review and implications. Communications of the Association for Information Systems, 12(1), 5.

ACTA INFOLOGICA 2023;7(1):47-58
dergipark.org.tr/acin

İSTANBUL UNIVERSITY PRESS

DOI: 10.26650/acin.1165100

RESEARCH ARTICLE

Website Performance Evaluation by Grey Relational Analysis: A Research on Companies in BIST Technology and Informatics Index

Gri İlişkisel Analiz ile Web Site Performans Değerlendirmesi: BIST Teknoloji ve Bilişim Endeksi Üzerine Bir Araştırma

Hülya Başeğmez[1], Kadir Özdemir[2]

[1] (Assist. Prof.), Beykent Unversity, Faculty of Economics and Administrative Sciences, Management Information Systems, Istanbul, Turkiye

[2](Res. Asst.), Bursa Technical University, Faculty of Humanities and Social Sciences, Department of Business Administration, Bursa, Turkiye

ORCID: H.B. 0000-0001-7768-1666; K.Ö. 0000-0002-2034-4797

Corresponding author:
Hülya BAŞEĞMEZ
Beykent Unversity, Faculty of Economics and Administrative Sciences, Management Information Systems, Istanbul, Turkiye
E-mail address: hulyabasegmez@gmail.com

Submitted: 22.08.2022
Revision Requested: 24.10.2022
Last Revision Received: 16.12.2022
Accepted: 04.02.2023
Published Online: 09.05.2023

Citation: Basegmez, H. & Ozdemir, K. (2023). Website performance evaluation by grey relational analysis: a research on companies in BIST technology and informatics index. *Acta Infologica, 7*(1), 47-58.
https://doi.org/10.26650/acin.1165100

ABSTRACT

Since websites are the visible face of companies, they are a significant determinant of how customers perceive interactions. Websites give the first impression of ensuring customer satisfaction and meeting expectations. Therefore, a quality website will provide competitive advantages. In this study, the website performance of 27 businesses in the BIST Technology and Informatics Index was evaluated, and 24 websites were analysed because the websites of three companies gave analysis errors. Website performance was measured by Page Size (Mb), Onload Time (s), First contentful paint (ms), Performance Score (%), Largest Contentful Paint (ms), Total Blocking Time (ms) and Speed Index (ms) criteria. The performance evaluation criteria of the websites were determined based on the literature and expert opinion. The data of the websites were obtained from GTmetrix automated test tool and analysed by using the Grey Relational Analysis (GRA) method. As a result of the analysis, companies' websites were ranked according to their performance score. Accordingly, Alcatel-Lucent, Plastik Kart, Escort, Smartiks, and Papilon have the highest website performance scores and are in the top five. Aselsan, Index, Datagate, Despec and Mia Teknoloji have the lowest performance scores.

Keywords: Website performance evaluation, Grey Relational Analysis (GRA), GTmetrix

ÖZ

Web siteler, işletmelerin görünen yüzü olduğundan işletmelerin müşteriler tarafından nasıl algılanacağının önemli belirleyicisidir. Müşteri memnuniyetinin sağlanması ve beklentilerin karşılanması noktasında web siteler ilk izlenimi vermektedir. Dolayısıyla kaliteli bir web sitesine sahip olmak işletmelere rakipleri karşısında avantajlar sağlayacaktır. Bu çalışmada BİST Teknoloji ve Bilişim Endeksi'nde yer alan 27 işletmenin web site performansı değerlendirilmiş olup üç işletmenin web sitesi analiz hatası verdiği için 24 web sitesinin analizi yapılmıştır. Web sitelerin performansı sayfa boyutu, yüklenme süresi, FCP (First Contentful Paint), performans skoru, LCP (Largest Contentful Paint), toplam engelleme süresi ve hız endeksi kriterleri kullanılarak ölçülmüştür. Kriterler uzman görüşüne ve literatüre dayanarak belirlenmiştir. Web sitelere ait veriler GTmetrix otomatik test aracı kullanılarak elde edilmiş olup Gri İlişkisel Analiz (GRA) yöntemi kullanılarak analiz edilmiştir. Analizler sonucunda şirketlere ait web siteler performans skorlarına göre sıralanmıştır. Buna göre Alcatel-Lucent, Plastik Kart, Escort, Smartiks, ve Papilon şirketleri en yüksek web sitesi performans skoruna sahiptir ve ilk beşte sırada yer almaktadır. Ayrıca Aselsan, Index, Datagate, Despec ve Mia Teknoloji şirketleri en düşük web sitesi performans skoruna sahiptir.

Anahtar Kelimeler: Web site performans değerlendirmesi, gri ilişkisel analiz (GRA), GTmetrix

1. INTRODUCTION

The rapid increase in internet usage worldwide increases the importance of websites for companies and customers. Since websites are the visible face of companies, they are an essential determinant of how customers perceive interactions. Websites give the first impression of ensuring customer satisfaction and meeting expectations. Companies and customers initially interact through websites and customers obtain basic data and information about companies from the websites. Therefore, a quality website will provide companies with advantages over their competitors (Ozbek, 2020). There are many factors that determine the quality of a website. Some quality factors are related to design dimensions (colors, font, font size, photos, etc.), while others are more technical (page speed, page size, HTTPS request, etc.). In this framework, the performance of websites can be evaluated in the context of different quality factors.

Website quality and performance have a significant impact on customers' perceptions and attitudes about the companies. The websites that have high quality and high performance create a more favorable impression on customers (Rababah ve Masoud, 2010; Ozbek, 2020). This phenomenon is the same for investors and other stakeholders. Like customers, investors in the stock market get their initial impressions and information about companies from websites. Companies in the IT sector in particular provide consultancy services about technology and informatics to other companies and individual users. In this case, it is expected that the website performance of IT companies will be high. Therefore, in this study, the website performance of 27 businesses in the BIST Technology and Informatics Index was evaluated, and 24 websites were analysed because the websites of three companies gave analysis errors. The performance evaluation criteria of the websites were determined based on the literature and expert opinion. In this regard, the website performance of 24 companies was measured by using seven criteria which are Page Size (Mb), Onload Time (s), First contentful paint (ms), Performance Score (%), Largest Contentful Paint (ms), Total Blocking Time (ms) and Speed Index (ms). Data were obtained using the GTmetrix automated testing tool. Also, data were analysed using the Grey Relational Analysis (GRA) method, and 24 companies' websites were ranked according to their performance score.

It is important to determine the weights of the criteria relative to each other when evaluating the performance in the grey relational analysis method. The criteria weights can be accepted as equal or different weight values can be assigned to the criteria by using methods such as the modified digital logic method (Dehghan-Manshadi et al., 2007), Analytical network process (ANP) (Tsai et al., 2011), Analytical hierarchy process (AHP) (Sun et al., 2013; Wang et al., 2012; Sandeep et al., 2011), and Entropy method (Shemshadi et al., 2011; Ayçin, 2018; Vatansever and Akgül, 2018). However, in a study by Kung et al. (2006), the criteria weights were taken equally and it was determined that the profitability ratios affected the ranking obtained as a result of the analyses (Yıldırım et al., 2021).

Many studies use the grey relational analysis (GRA) methods in diverse research areas when the literature is examined. For instance, evaluation of online travel agencies (Gavcar and Organ, 2020) and evaluation of the companies' financial performance (Kula et al., 2016; Ecer and Gunay, 2014; Gunay et al., 2018; Suvvari et al., 2019; Dinler, 2021; Çekici and Babacan, 2022), evaluation of smartphone technical features' significance level (Sahin and Aydemir, 2019), supplier selection (Rajesh and Ravi 2015; Wang et al., 2017; Soylemez et al., 2021), evaluation of energy consumption and CO_2 emission of transportation sector (Yuan et al., 2017), evaluation of bank performance (Guru and Mahalik, 2019), evaluation of social media usage performance of companies (Mercangoz et al., 2018), and evaluation of website performance (Wuwei, 2009; Vatansever and Akgul, 2018; Zhang, 2018).

Studies examining website performance with Grey Relational Analysis are insufficient in the literature. In this study, the website performance of the companies in the BIST technology and informatics index was evaluated using grey relational analysis. In this context, it is thought that this study will contribute to the literature on grey relational analysis and website performance evaluation. In this study, the main reason for choosing the GRA method, which is one of the popular multi-criteria decision-making methods, is its advantage over other methods.

This study is organized as: Section 2 discusses the related works in literature. In Section 3, information about used data and method are given. In section 4, the findings arranged according to the results of the analysis are included. Section 5 accommodates the conclusion.

2. LITERATURE REVIEW

In this section, different decision-making methods used in the performance measurement of websites and some studies published in the literature on grey relational analysis method, especially in the last five years, are presented in chronological order without any sector restrictions.

Kaur et al. (2016), evaluated the performance of 12 university websites by using diverse automated testing tools that are Pingdom, GTmetrix, Website Grader and Site Speed Checker. They analysed performance score, speed, response time, page size, number of requests and load time of websites. In the study, they also compared automated testing tools evaluation results and they revealed measured components of each website from each testing tool.

Beneida and Namoun (2018), evaluated the performance of top-rank Arabic educational websites in their study with seven components, which are efficiency, effectiveness, learnability, memorability, errors, content and satisfaction. Additionally, they evaluate technical performance of the websites using GTmetrix and Web Page Analyser. In this case, page speed, Yslow grade, fully loaded time, total page size, total number of requests, HTML size, image size, script size, CSS size were examined in evaluating the technical performance of Arabic educational websites. They saw that according to the web performance evaluation results of GTmetrix and Web Page Analyser, all educational websites have weak performance scores.

Bilal et al. (2019), evaluated the performance of the top 60 government websites of China and Pakistan according to accessibility and usability issues. They specified seven criteria of websites and analysed these criteria by using automated testing tools. In this context, loading time and page size analysed by using GTmetrix, mobile friendliness were analysed by using Google-Mobile Friendly Tool, broken links were analysed by using Dead Link Checker, HTML5-ARIA and content errors were analysed by using WAVE and accessibility performance of the websites were analysed by using the AChecker testing tool.

Stringam and Gerdes (2019), evaluated and compared load performance of 259 international hotel websites using both mobile and desktop devices. They analysed load time and speed index of hotel websites with the GTmetrix automated testing tool. In the study, they interpreted results according to continent (Africa, Asia, Australia, Europe, North America and South America) and size of hotels.

Salvio and Palaoag (2019), evaluated the performance of 5 selected Philippine e-government websites by using Website Grader, GTmetrix and Pingdom automated testing tools. They analysed performance score, page size, page requests and page speed of the websites. In the context of each analysed parameters, they also made various recommendations to increase performance of each website.

Ozdemir and Turna (2020), evaluated the performance of 10 commercial banks' websites in Turkey by using AHP, TOPSIS and VIKOR MCDM methods. They determined six performance criteria: load time, page speed, markup, page views, visitors and speed index. Accordingly, they used five automated testing tools, which are Pingdom, GTmetrix, Validator, Webpagetest and Websiteinformer, to evaluate performance of websites. They obtained data at the same time of the day by using the same devices over the 60-day period.

Csontos and Heckl (2021), evaluated the performance of 25 Hungarian public sector websites based on accessibility, usability and security components. For testing the accessibility component, they used the WAVE automated testing tool, however for testing usability they used GTmetrix and for security issue they used Securi online checker. In this regard, they analysed page speed, page load time, YSlow value, total size of websites, number of total requests, GDPR compliance and multilingual support of websites.

Hossain et al. (2021), evaluated the performance of 10 e-commerce websites in Bangladesh by considering nine parameters: first byte, load time, start render, speed index, first contentful paint (FCP), largest contentful paint (LCP), total blocking time, cumulative layout shift (CLS) and interactive time. They collected data from WebpageTest, PageSpeed Insights and GTmetrix testing tools.

Maruf and Ozdemir (2021), evaluated the performance of 15 commercial banks' websites in Turkey by using SWARA (Step-wise Weight Assessment Ratio Analysis) and ARAS (Additive Ratio Assessment) methods. They analysed performance score, page speed, fully loaded time, page size, bounce rate, number of visitors, average number of page views, average time spent on websites, world ranking and Turkey ranking of websites. In this regard, to evaluate websites performance they used GTmetrix and Similarweb testing tools.

Al-Sakran and Alsudairi (2021), evaluated the usability and accessibility components of 22 e-government websites from various sectors, such as education, social development, health, environment, finance, foreign affairs and justice etc. in Saudi Arabia. In the first part of the study, they analysed websites with a manual evaluation by experts, and in the second part they used automated testing tools for desktop (GTmetrix, WAVE) and mobile (Dareboost). With GTmetrix they analysed page speed, Yslow grade, fully loaded time, page size and number of requests of e-government websites. According to the findings, e-government websites have usability and accessibility problems that negatively affect website performance.

Belinda et al. (2021), evaluated four websites' (FUTA, UMaT, Yahoo and Google) performance with internal and external usability factors. The internal usability factors are performance, page size, speed, load time, number of requests and the external usability factors are ease of navigate, easy to find information, clear organization of information, pleasant interface, useful image and right presentation of content. They used GTmetrix, Website Grader and Pingdom automated testing tools for evaluating internal usability of websites. However, in the second part of the study they conducted a survey on users to assess the external usability of websites.

Dawis and Setiawan (2022), evaluated the performance of university websites by using the automated testing tool GTmetrix. They analysed page speed, Yslow grade, page load, total page size and number of requests of university website. As a result of the analysis, performance score of the university website was determined as E grade which means the website performance is not satisfactory and needs to be improved.

3. METHODOLOGY

This study aims to test and measure the website quality of companies in BIST technology and informatics index with online web diagnostic tools.

3.1. DATA COLLECTION

In this study, the website performance of 24 companies in the BIST technology and informatics index were examined. For this purpose, first of all, 26 criteria were determined to examine the performance of these websites. Then, the number of criteria was reduced to 7 by taking expert opinion. The expert team consists of four independent individuals, one academician, two web designers and one computer engineer. In addition to the expert opinion, determined criteria were used in the literature for website performance measurement. For instance, page size (Nacar and Ozdemir, 2021; Bilal et al., 2019), onload time, speed index (Stringam and Gerdes, 2019; Hossain et al., 2021), first contentful paint, largest contentful paint, total blocking time (Hossain et al., 2021) and performance score (Al-Sakran and Alsudairi, 2021; Belinda et al., 2021). The criteria utilized in this study and their descriptions are shown in Table 1.

Table 1
Criteria for Website Performance Measurement

Criteria	Objective	Definition
Page Size (Mb)	Min	Page size is a significant component of website performance as it is directly related to speed and quality. Page size increases depending on the size and quality of images, videos, and excess content on the websites (Nacar and Ozdemir, 2021: 210). The recommended page size for fast-loading websites is<= 12 KB, average is <=2 MB and slow-loading is >= 2 MB (Bilal et al., 2019: 318).
Onload Time (s)	Min	Onload time is defined as completing the processing of the web page and downloading the whole resources (images, videos, texts, font size, etc.) on the page (Gtmetrix, 2022a). In this case, onload time is directly related to the customer waiting time and satisfaction (Zhang et al., 1999).
First contentful paint (ms)	Min	First contentful paint (FCP) measures how quickly visitors can view the actual content of websites (images, videos, texts, etc) and it depends on page load speed (Gtmetrix, 2022b).

Performance Score (%)	Max	The performance score measures how well the website performs from the visitor's perspective. The Performance score is determined by some key performance metrics such as loading performance, interactivity, and visual stability. Since the performance score provides measurement from a visitor perspective, it can vary due to geographic, hardware, and network differences (Gtmetrix, 2022c). The grading scale is between 0-100 and if the score is 90-100, it is indicated by A grade, 80-90 is indicated by B Grade, 70-80 is indicated by C grade, and so on (Al-Sakran and Alsudairi (2021).
Largest Contentful Paint (ms)	Min	Largest Contentful Paint measures how long it takes visitors to see the largest content item (images, videos, heading text, etc.) on the web pages. In this regard, Largest Contentful Paint is a substantial indicator of visitors' perceived speed of web pages (Gtmetrix, 2022d).
Total Blocking Time (ms)	Min	Total blocking time measures the total amount of time the web page is blocked, which is between First Contentful Paint and Time to Interactive (Gtmetrix, 2022e).
Speed Index (ms)	Max	Speed index measures how quickly a web page is visually complete up to the fold. The speed index tells the visibility of the content of the web page to the visitors and analyses the viewport of the web browser (Stringam and Gerdes, 2019: 21).

GTmetrix automatic evaluation tool was used to analyse the performance of 24 websites to be used in the study. GTmetrix is an automated software tool that measures websites performance. GTmetrix use Google Pagespeed and Yahoo YSlow as a machine for website analysis. This tool measures website performance based on available parameters such as website speed and page size (Dawis and Setiawan, 2022; Kaur et al., 2016). The data was collected from 02/01/2022-02/28/2022. Table 2 shows descriptive statistics of the dataset.

Table 2
Descriptive Statistics of the Dataset

Criteria	Mean	Std. Dev.	Max	Min
Page Size (Mb)	6.83	7.71	25.77	0.28
Onload Time (s)	87.98	81.58	283.33	17.81
First Contentful Paint (s)	2.51	0.87	4.17	0.03
Performance Score (%)	53.71	15.98	86	9
Largest Contentful Paint (s)	5.50	3.61	16.22	0.59
Total Blocking Time (ms)	226.25	563.82	2800	0
Speed Index (s)	5.45	3.62	17.67	1.07

Internet speed of visitors is a significant variable in collecting data for website performance measurement with GTmetrix automatic testing tools. Because the data can differ according to the internet quality. For this reason, we also recorded the speed of the internet network via Speedtest.net website. In this regard, during the data collection period, our average ping was 16.3 ms, download speed was 28.84 Mbps and upload speed was 20.2 Mbps.

3.2. GREY RELATIONAL ANALYSIS METHODS

Grey system theory offers a mathematical method for dealing with poor, incomplete, and uncertain information. It was first developed by Deng Julong, to study the uncertainties in system models, to help in prediction and decision making. Grey system theory uses a certain idea of information. It categorizes situations with no information as black and those with perfect information as white. If a system has partially known information, it is called a grey system. But in actual problems, neither of these idealistic scenarios ever happens (Ju-long, Deng, 1982). The grey system theory has five major parts: grey prediction, grey decision, grey relational analysis, grey control, and grey programming (Ju-Long, Deng, 1982; Ju-Long, Deng, 1989).

In the grey relational analysis, also known as the grey relational generation, experimental findings are first normalized in the 0–1 range. To express the link between the desired and actual experimental data, the grey relational coefficient is generated from the normalized experimental data. The grey relational grade is then calculated by averaging each process response's respective grey relational coefficient. The grey relational grade serves as the foundation for the overall assessment of the multiple process answers. As a consequence, the optimization of a single grey relational grade can be replaced by the optimization of the complex multiple process answers. For the multi-response process, the grey relational grade may be thought of as the comprehensive assessment of experimental data (Lin C. L., 2004; Acır et all. 2017; Mia et al. 2018; Ramesh, Baranithara and Sakthivel, 2019; Uzun, 2019).

Table 3
Comparative performance of some well-known MCDM techniques (Wang et al., 2013).

MCDM method	Computational time	Simplicity	Mathematical calculations involved	Stability
GRA	Moderate	Moderately critical	Moderate	Medium
TOPSIS	Moderate	Moderately critical	Moderate	Medium
AHP	Very high	Very critical	Maximum	Poor
ELECTRE	High	Moderately critical	Moderate	Medium
PROMETHEE	High	Moderately critical	Moderate	Medium

The main reason for choosing the GRA method, which is one of the popular multi-criteria decision-making methods in this study is the advantages compared with others. In addition, since the GRA method is independent of a probability distribution, it is known that it provides better results in studies with small sample sizes than other statistical analysis techniques (Kung et al., 2006). Table 3 shows the performance comparison of some popular MCDM methods discussed by Wang et al (2013). According to Table 3, it can be said that the grey relational analysis method clearly outperforms other methods in terms of being more effective in solving complex decision-making problems, having a wide application area, shorter computation time, and flexibility.

The calculation steps of the grey relational analysis are presented as follows:

Step 1: Calculating the decision matrix.

To n alternatives and m criteria, a decision matrix is formulated as $F = \left(f_{ij}\right)_{n\times m}$. Here, f_{ij} is the value of *j*th criterion function for alternative A_i $(i = 1,2,\dots,n\,; j = 1,2,\dots,m)$ and obtained decision matrix is as follows (Wang et al. 2013; Uzun, 2019):

$$F = \left(f_{ij}\right)_{n\times m} = \begin{bmatrix} x_1(1) & \dots & x_1(n) \\ \vdots & \ddots & \vdots \\ x_m(1) & \dots & x_m(n) \end{bmatrix} \quad (1)$$

Step 2: Determination of the series to be compared.

To compare the factors in the decision problem, a reference series is determined and the comparison matrix is obtained by adding this series to the decision matrix.

$x_i^0(k)$ stands for the original data sequence, $x_i(k)$ stands for normalized data, that is, data sequence after pre-processing. $\max x_i^0\,(k)$ denotes the maximum of $x_i(k)$ and $\min x_i^0(k)$ represents the minimum of $x_i(k)$ values (Uzun, 2019).

Step 3: Calculating normalized decision matrix.

In order to acquire normalization values Eq. (2) is calculated as follows:

$$\boldsymbol{x_i(k)} = \frac{\mathbf{\max} x_i^0(k) - x_i^0(k)}{\mathbf{\max} x_i^0(k) - \mathbf{\min} x_i^0(k)} \quad (2)$$

where $i = 1,\dots,m; k = 1,\dots,n.$ m m is the number of experimental data and n is the number of parameters. After the normalization values are determined, the decision matrix is converted to a normalized decision matrix (Uzun, 2019).

Step 4: Determination of grey relational coefficient.

The grey relational grade in grey relational analysis is the measure of the relevance between two systems or two sequences. A local grey relation measurement is what is used when there is just one sequence, $x_0(k)$, that may be used as the reference sequence and all other sequences perform as comparison sequences. After data pre-processing is completed the grey relation coefficient $\xi_i(k)$ for the *k*th performance characteristic in the *i*th experiment may be expressed (Lin, 2004; Acır et al. 2017; Zerti et al., 2018; Mia et all. 2018; Ramesh et al. 2019; Uzun, 2019) as:

$$\boldsymbol{\xi_i(k)} = \frac{\Delta_{\mathbf{min}} + \xi\,\Delta_{\mathbf{max}}}{\Delta_{\mathbf{0i}}(\boldsymbol{k}) + \xi\Delta_{\mathbf{max}}} \quad (3)$$

Where, the reference sequence's and the comparability sequence's deviation sequence is denoted by Δ_{0i}.

$$\Delta_{0i}(k) = |\,x_0(k) - x_j(k)| \quad (4)$$

$$\Delta_{\min} = \min_j \min_k |x_0(k) - x_j(k)| \quad (5)$$

$$\Delta_{\max} = \max_j \max_k |x_0(k) - x_j(k)| \quad (6)$$

Here, the reference sequence is denoted by $x_0(k)$, while the comparability sequence is denoted by $x_j(k)$. In addition, the identifying or distinguishing coefficient is $\xi \in [0,1]$. The value of ξ may be adjusted according to actual system requirements. ξ value is generally used as 0.5 and ξ =0.5 was also used in this study. The value of ξ is the smaller and the distinguished ability is the larger.

Step 5. Determination of grey relational grade.

The average value of the grey relational coefficients is taken as grey relational grade (Lin, 2004; Tzeng et al., 2009; Acır et al., 2017). As a result, if criteria weights are determined equally, grey relational grade is defined as follows:

$$\gamma_i = \frac{1}{n} \sum_{k=1}^{n} \xi_i(k). \quad (7)$$

Step 6: Ranking the alternatives.

The values of the grey relationship grade are used to rank the existing alternatives in terms of similarity in the reference series. A higher value means higher similarity (Wang et al., 2013).

4. FINDINGS

In this part of the study, the Grey Relational Analysis method was used evaluate the performance of 24 websites in the BIST Technology and Informatics index. The variables used in the analysis are given in Table 1. In addition, the names and websites of the companies included in the analysis are shown in Table 4.

Table 4
Companies and Websites

Website Code	Companies	Websites	Website Code	Companies	Websites
WS1	Alcatel - Lucent	www.alcatel-lucent.com.tr	WS13	Kron	www.kron.com.tr
WS2	ARD Bilişim	www.ardbilisim.com.tr	WS14	Link	www.link.com.tr
WS3	Arena	www.arena.com.tr	WS15	Logo	www.logo.com.tr
WS4	Armada	www.armada.com.tr	WS16	Manas	www.manas.com.tr
WS5	Aselsan	www.aselsan.com.tr	WS17	Matriks Data	www.matriksdata.com
WS6	ATP	www.atp.com.tr	WS18	Mia Teknoloji	www.miateknoloji.com
WS7	Datagate	www.datagate.com.tr	WS19	Mobiltel	www.mobiltel.com.tr
WS8	Despec	www.despec.com.tr	WS20	Papilon	invest.papilon.com.tr
WS9	Escort	www.escort.com.tr	WS21	Penta	www.penta.com.tr
WS10	Index	www.index.com.tr	WS22	Plastik Kart	www.plastkart.com.tr
WS11	Kafein	www.kafein.com.tr	WS23	Smartiks	www.smartiks.com.tr
WS12	Karel	www.karel.com.tr	WS24	VBT	www.vbt.com.tr

Initially, the data set was arranged by obtaining the minimum and maximum values of each variable. Then, in order to express the whole sample with the same unit, the data were normalized to the range of 0-1 using Equation (2). Normalized data are given in Table 5. While the data were normalized, calculations were made in the direction that the performance score (%) and speed index (ms) criteria were maximum and the others were minimum (Gtmetrix, 2022e).

Table 5
Normalized Values

	min	min	min	max	min	min	max
Website Code	Page Size (Mb)	Onload Time (s)	First contentful paint (ms)	Performance Score (%)	Largest Contentful Paint (ms)	Total Blocking Time (ms)	Speed Index (ms)
WS1	0.97	0.96	0.80	1.00	0.95	0.94	0.03
WS2	0.86	0.82	0.42	0.61	0.70	0.99	0.21
WS3	0.95	0.91	0.39	0.51	0.85	0.87	0.14
WS4	0.69	0.67	0.39	0.45	0.53	1.00	0.31
WS5	0.70	0.02	0.42	0.51	0.53	0.96	0.26

WS6	0.77	0.69	0.13	0.44	0.00	0.97	1.00
WS7	0.07	0.01	0.59	0.61	0.62	0.96	0.13
WS8	0.05	0.00	0.65	0.61	0.54	0.96	0.21
WS9	0.98	0.95	0.57	0.87	0.91	1.00	0.15
WS10	0.42	0.53	0.62	0.30	0.15	0.95	0.39
WS11	0.83	0.93	0.40	0.43	0.71	0.99	0.20
WS12	0.91	0.85	0.12	0.57	0.75	0.99	0.21
WS13	0.87	0.87	0.52	0.62	0.81	0.97	0.31
WS14	0.95	0.90	0.46	0.70	0.85	1.00	0.15
WS15	0.84	0.87	0.46	0.43	0.74	0.90	0.32
WS16	0.88	0.81	0.23	0.58	0.60	1.00	0.26
WS17	0.97	0.89	0.46	0.62	0.74	1.00	0.17
WS18	0.64	0.83	0.00	0.00	0.64	0.00	0.77
WS19	0.96	0.90	0.28	0.48	0.76	0.92	0.32
WS20	0.96	0.96	0.38	0.73	0.85	1.00	0.15
WS21	0.82	0.80	0.36	0.57	0.78	0.94	0.24
WS22	1.00	0.97	0.71	0.88	0.88	1.00	0.08
WS23	0.00	1.00	1.00	0.86	1.00	0.81	0.00
WS24	0.74	0.54	0.14	0.55	0.59	0.95	0.34
Max	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00

After that, the distance matrix was determined with the help of Equation (3-6) to find the grey correlation coefficient. The distance matrix is shown in Table 6.

Table 6
Distance Matrix

Website Code	Page Size (Mb)	Onload Time (s)	First contentful paint (ms)	Performance Score (%)	Largest Contentful Paint (ms)	Total Blocking Time (ms)	Speed Index (ms)
WS1	0.03	0.04	0.20	0.00	0.05	0.06	0.97
WS2	0.14	0.18	0.58	0.39	0.30	0.01	0.79
WS3	0.05	0.09	0.61	0.49	0.15	0.13	0.86
WS4	0.31	0.33	0.61	0.55	0.47	0.00	0.69
WS5	0.30	0.98	0.58	0.49	0.47	0.04	0.74
WS6	0.23	0.31	0.87	0.56	1.00	0.03	0.00
WS7	0.93	0.99	0.41	0.39	0.38	0.04	0.87
WS8	0.95	1.00	0.35	0.39	0.46	0.05	0.79
WS9	0.02	0.05	0.43	0.13	0.09	0.00	0.85
WS10	0.58	0.47	0.38	0.70	0.85	0.06	0.61
WS11	0.17	0.07	0.60	0.57	0.29	0.01	0.80
WS12	0.09	0.15	0.88	0.43	0.25	0.01	0.79
WS13	0.13	0.13	0.48	0.38	0.19	0.03	0.69
WS14	0.05	0.10	0.54	0.30	0.15	0.00	0.85
WS15	0.16	0.13	0.54	0.57	0.26	0.10	0.68
WS16	0.12	0.19	0.77	0.42	0.40	0.00	0.74
WS17	0.03	0.11	0.54	0.38	0.26	0.00	0.83
WS18	0.36	0.17	1.00	1.00	0.36	1.00	0.23
WS19	0.04	0.10	0.72	0.52	0.24	0.08	0.68
WS20	0.04	0.04	0.62	0.27	0.15	0.00	0.85
WS21	0.18	0.20	0.64	0.43	0.22	0.06	0.76
WS22	0.00	0.03	0.29	0.12	0.12	0.00	0.92
WS23	1.00	0.00	0.00	0.14	0.00	0.19	1.00
WS24	0.26	0.46	0.86	0.45	0.41	0.05	0.66
Max	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Then, using the obtained distance matrix and Equation (3), the grey correlation coefficients were determined.

Table 7
Grey Relational Coefficient

Website Code	Page Size (Mb)	Onload Time (s)	First contentful paint (ms)	Performance Score (%)	Largest Contentful Paint (ms)	Total Blocking Time (ms)	Speed Index (ms)
WS1	0.9499	0.9317	0.7143	1.0000	0.9018	0.8917	0.3392
WS2	0.7813	0.7375	0.4646	0.5620	0.6249	0.9838	0.3872
WS3	0.9137	0.8429	0.4490	0.5033	0.7722	0.7973	0.3681
WS4	0.6201	0.6022	0.4493	0.4783	0.5131	1.0000	0.4198
WS5	0.6278	0.3370	0.4622	0.5033	0.5150	0.9340	0.4039
WS6	0.6892	0.6153	0.3639	0.4724	0.3333	0.9365	1.0000
WS7	0.3495	0.3350	0.5506	0.5620	0.5662	0.9265	0.3647
WS8	0.3454	0.3333	0.5851	0.5620	0.5183	0.9174	0.3868
WS9	0.9580	0.9113	0.5387	0.7938	0.8479	1.0000	0.3696
WS10	0.4640	0.5129	0.5678	0.4162	0.3712	0.9009	0.4493
WS11	0.7429	0.8723	0.4532	0.4667	0.6341	0.9715	0.3853
WS12	0.8517	0.7709	0.3620	0.5385	0.6653	0.9838	0.3883
WS13	0.7878	0.7895	0.5090	0.5704	0.7241	0.9479	0.4207
WS14	0.9066	0.8269	0.4811	0.6260	0.7727	1.0000	0.3692
WS15	0.7577	0.7961	0.4798	0.4667	0.6622	0.8289	0.4248
WS16	0.8045	0.7203	0.3943	0.5461	0.5541	0.9993	0.4023
WS17	0.9426	0.8253	0.4829	0.5704	0.6618	0.9986	0.3746
WS18	0.5813	0.7462	0.3333	0.3333	0.5791	0.3333	0.6852
WS19	0.9177	0.8302	0.4110	0.4904	0.6728	0.8690	0.4228
WS20	0.9257	0.9236	0.4483	0.6471	0.7663	1.0000	0.3697
WS21	0.7321	0.7161	0.4391	0.5385	0.6990	0.8940	0.3978
WS22	1.0000	0.9476	0.6317	0.8105	0.8061	0.9908	0.3517
WS23	0.3333	1.0000	1.0000	0.7778	1.0000	0.7216	0.3333
WS24	0.6540	0.5184	0.3684	0.5238	0.5465	0.9132	0.4328

Finally, grey relational degrees are determined by using Equation (7).

Table 8
Grey Relational Grade

Website Code	Company Name	Performance Score	Ranking	Website Code	Company Name	Performance Score	Ranking
WS1	Alcatel - Lucent	0.818	1	WS13	Kron	0.678	8
WS2	ARD Bilişim	0.649	12	WS14	Link	0.712	6
WS3	Arena	0.664	9	WS15	Logo	0.631	16
WS4	Armada	0.583	18	WS16	Manas	0.632	14
WS5	Aselsan	0.540	20	WS17	Matriks Data	0.694	7
WS6	ATP	0.630	17	WS18	Mia Teknoloji	0.513	24
WS7	Datagate	0.522	22	WS19	Mobiltel	0.659	10
WS8	Despec	0.521	23	WS20	Papilon	0.726	5
WS9	Escort	0.774	3	WS21	Penta	0.631	15
WS10	Index	0.526	21	WS22	Plastik Kart	0.791	2
WS11	Kafein	0.647	13	WS23	Smartiks	0.738	4
WS12	Karel	0.652	11	WS24	VBT	0.565	19

Table 8 shows the final results of the analyses. According to the results of the analysis using the data set obtained by taking the average data, the website with the highest performance was the Alcatel-Lucent company. Also, the website of the Plastik Kart company is ranked second, and the website of the Escort company ranked third. The company with the lowest website performance is Mia Teknoloji.

4. DISCUSSION AND CONCLUSION

Today, every company desires to provide a great user experience to the visitors of their websites. With the advancement of technology, user expectations are increasing, and so is better performance of websites which means more satisfied users and companies. However, if the website is slow, various problems will arise, such as the page opening late, and due to such issues,

user satisfaction will decrease, and the website will lose visitors. Because websites are the visible face of businesses, they are an essential determinant of how customers perceive companies. Websites give also the first impression of ensuring customer satisfaction and meeting expectations. Therefore, having a quality website will provide competitive advantages.

The purpose of this study is to evaluate the website performance of 27 companies in the BIST Technology and Informatics Index. Since the websites of 3 companies gave an analysis error, the data could not be obtained, so the websites of 24 companies were evaluated. The performance evaluation criteria of the websites were determined based on the literature and expert opinion. In this context, the website performance of 24 companies was measured by using seven criteria which are Page Size (Mb), Onload Time (s), First contentful paint (ms), Performance Score (%), Largest Contentful Paint (ms), Total Blocking Time (ms) and Speed Index (ms). Data were obtained using the Gtmetrix automated testing tool. Relevant data were analysed using the Grey Relational Analysis (GRA) method, and as a result of the analysis, 24 companies' websites were ranked according to their performance score. In addition, no weight assignment was made for the criteria in the analysis, that is, the criteria weights were accepted as equal.

According to the performance scores calculated as a result of the Grey Relational Analysis method, Alcatel-Lucent (0.818), Plastik Kart (0.791), Escort (0.774), Smartiks (0.738) and Papilon (0.726) have the highest website performance scores and are in the top five. Aselsan (0.540), Index (0.526), Datagate (0.522), Despec (0.521) and Mia Teknoloji (0.512) have the lowest performance scores. These companies' websites are in the last five in the ranking.

When the website performance evaluation literature was examined, studies examining website performance with Grey Relational Analysis were found to be insufficient. In this context, it is thought that the study will contribute to the literature. Also, the website performances of the companies in the BIST Technology and Informatics Index were measured for the first time, and the company's websites were ranked.

As with any study, this study also has some limitations. The most important limitation of the study is the measurement of website performance with seven criteria. In future studies, website performance can be measured using different criteria. The analyses results also differ when other criteria are used. Besides that, obtaining the data of 24 companies using the GTmetrix automatic test tool is a significant limitation. In this case, different tools such as Pingdom, Website Grader, Site Speed Checker, Validator, Webpagetest, Websiteinformer, etc., can be used in future studies. Another limitation of the study is that website performance is only considered using the "speed" measure. A website's performance can also be evaluated using different measures such as content, usability, and user interface. These metrics are beyond the scope of the study. Consequently, the results are valid in the context of criteria, testing tool, and method. Thus, the results might also differ when other website performance criteria, testing tools, measured metrics, and method differ.

Peer-review: Externally peer-reviewed.
Author Contributions: Conception/Design of Study- H.B., K.Ö.; Data Acquisition- K.Ö.; Data Analysis/Interpretation- H.B.; Drafting Manuscript- H.B., K.Ö.; Critical Revision of Manuscript- H.B., K.Ö.; Final Approval and Accountability- H.B., K.Ö.
Conflict of Interest: The authors have no conflict of interest to declare.
Grant Support: The authors declared that this study has received no financial support.

REFERENCES

Acır, A., Canlı, M. E., Ata, İ., & Çakıroğlu, R. (2017). Parametric optimization of energy and exergy analyses of a novel solar air heater with grey relational analysis. *Applied Thermal Engineering*, *122*, 330-338.

Al-Sakran, H. O., & Alsudairi, M. A. (2021). Usability and accessibility assessment of Saudi Arabia mobile e-government websites. *IEEE Access*, *9*, 48254-48275.

Benaida, M., & Namoun, A. (2018). Technical and perceived usability issues in Arabic educational websites. *International Journal of Advanced Computer Science and Applications*, *9*(5), 391-400.

Belinda, B. I., Kayode, A. B., Solomon, N., & Favour-Bethy, T. A. (2021). Analysis of internal and external website usability factors. *Communications on Applied Electronics (CAE)*, 7 (35), 9-18.

Bilal, M., Yu, Z., Song, S., & Wang, C. (2019). Evaluate accessibility and usability issues of particular China and Pakistan government websites. In *2019*

2nd International Conference on Artificial Intelligence and Big Data (ICAIBD) 316-322. IEEE.

Csontos, B., & Heckl, I. (2021). Accessibility, usability, and security evaluation of Hungarian government websites. *Universal Access in the Information Society*, *20*(1), 139-156.

Çekici, E. M. & Babacan, Ş. F. (2022). Gri İlişkisel Analiz ile Kurumsal Yönetim ve Finansal Performans Karşılaştırması: Borsa İstanbul'da Bir Uygulama. Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi , 44 (1) , 221-238 . DOI: 10.14780/muiibd.1135582

Dawis, A. M., & Setiawan, I. (2022). Evaluation of the website Aisyiyah Surakarta of University performance based on search engine optimization using automated software testing Gtmetrix. *International Journal of Computer and Information System (IJCIS)*, *3*(1), 17-20.

Dehghan-Manshadi, B., Mahmudi, H., Abedian, A., & Mahmudi, R. (2007). A novel method for materials selection in mechanical design: combination of non-linear normalization and a modified digital logic method. *Materials & design*, *28*(1), 8-15.

Dinler, S. Turkiye'de faaliyet gosteren hayat dısı sigorta şirketlerinin gri ilişkisel analizi ile performansının değerlendirilmesi. *Oltu Beseri ve Sosyal Bilimler Fakultesi Dergisi*, 2(2), 200-212.

Ecer, F., & Gunay, F. (2015). Borsa İstanbul'da işlem goren turizm şirketlerinin finansal performanslarının gri ilişkisel analiz yöntemiyle ölçülmesi. Anatolia: *Turizm Arastırmaları Dergisi,* 25(1), 35-48.

Gavcar, C. T., & Organ, A. (2020). AHP-Gri İliskisel Analiz ve AHP-WASPAS yöntemleri ile online satıs yapan seyahat acentalarinin değerlendirilmesi. *Business & Management Studies*: *An International Journal*, 8(1), 731-753.

GTmetrix. (2022a). GTmetrix performance blog. Retrieved From https://gtmetrix.com/blog/browser-timings/#onload-time. Date of access: 13.04.2022.

GTmetrix. (2022b). Lighthouse: First Contentful Paint. Retrieved From https://gtmetrix.com/first-contentful-paint.html. Date of access: 13.04.2022.

GTmetrix. (2022c). Everything you need to know about the new GTmetrix Report (powered by Lighthouse). Retrieved From https://gtmetrix.com/blog/everything-you-need-to-know-about-the-new-gtmetrix-report-powered-by-lighthouse/#performance-score. Date of access: 15.04.2022.

GTmetrix. (2022d). Lighthouse: Largest Contentful Paint. Retrieved From https://gtmetrix.com/largest-contentful-paint.html. Date of access: 15.04.2022.

GTmetrix. (2022e). Lighthouse: Total Blocking Time. Retrieved From https://gtmetrix.com/total-blocking-time.html. Date of access: 15.04.2022.

Guru, S., & Mahalik, D. K. (2019). A comparative study on performance measurement of Indian public sector banks using AHP-TOPSIS and AHP-grey relational analysis. *Opsearch*, 56(4), 1213-1239.

Gunay, F., Karadeniz, E., & Dalak, S. (2018). Turkiye'de en yuksek net satıs gelirine sahip 20 şirketin finansal performanslarının gri ilişkisel analiz yöntemiyle incelenmesi. *Omer Halisdemir Universitesi İktisadi ve İdari Bilimler Fakultesi Dergisi*, 11(2), 51-73.

Hossain, M., Hassan, R., Amjad, M., & Rahman, M. (2021). Web performance analysis: An empirical analysis of e-commerce sites in Bangladesh. *International Journal of Information Engineering & Electronic Business*, 13(4), 47-54.

Ju-Long, Deng (1982). Control problems of grey systems. *Systems & control letters*, 1(5), 288-294.

Ju-Long, Deng (1989). Introduction to grey system theory. *The Journal of grey system*, *1*(1), 1-24.

Kaur, S., Kaur, K., & Kaur, P. (2016). An empirical performance evaluation of universities website. *International Journal of Computer Applications*, *146*(15), 10-16.

Kula, V., Kandemir, T., & Baykut, E. (2016). Borsa İstanbul'da işlem goren sigorta ve BES şirketlerinin finansal performansının gri ilişkisel analiz yöntemi ile incelenmesi. *Journal of Economics and Administrative Sciences,* 17(1), 37-53.

Kung, C. Y., Yan, T. M., & Chuang, S. C. (2006). GRA to Assess the Operating Performance of Non-Life Insurance Companies in Taiwan. *Journal of Grey System*, *18*(2).

Lin, C. L. (2004). Use of the Taguchi method and grey relational analysis to optimize turning operations with multiple performance characteristics. *Materials and manufacturing processes*, *19*(2), 209-220.

Maruf, M. and Ozdemir, K. (2021). Turkiye'deki ticari bankalara ait web sitelerin performanslarının SWARA ve ARAS yöntemi ile sıralanması. OPUS–Uluslararası Toplum Arastırmaları Dergisi, 18(Yonetim ve Organizasyon Ozel Sayısı), 1514-1537.

Mercangoz, B. A., Calıskan, B. O. O., & Yıldırım, B. F. (2018). Fortune 500 Turkiye'de yer alan lojistik isletmelerinin sosyal medya kullanımının analizi ve değerlendirilmesi. *İsletme Arastırmaları Dergisi*, 10(4), 1321-1341.

Mia, M., Rifat, A., Tanvir, M. F., Gupta, M. K., Hossain, M. J., & Goswami, A. (2018). Multi-objective optimization of chip-tool interaction parameters using Grey-Taguchi method in MQL-assisted turning. *Measurement*, *129*, 156-166.

Nacar, R., & Ozdemir, K. (2021). E-ticaret pazaryeri web sitelerinin dinamikleri: covid-19 pandemi sürecindeki değişim. *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 16(1), 205-226.

Ozbek, A. (2020). Web Sitesi Kalitesinin MULTIMOORA Yöntemi İle Değerlendirilmesi. *International Journal of Engineering Research and Development*, 12(2), 479-491.

Ozdemir, M. and Turna, G. B. (2020). Website performances of commercial banks in Turkey. *Dumlupınar Universitesi Sosyal Bilimler Dergisi*, *64*, 38–52.

Rababah, O., & Masoud, F. (2010). Key factors for developing a successful e-commerce website. *Communications of the IBIMA*, 2010, 1–9.

Rajesh, R., & Ravi, V. (2015). Supplier selection in resilient supply chains: a grey relational analysis approach. *Journal of Cleaner Production*, 86, 343-359.

Ramesh, K., Baranitharan, P., & Sakthivel, R. (2019). Investigation of the stability on boring tool attached with double impact dampers using Taguchi based Grey analysis and cutting tool temperature investigation through FLUKE-Thermal imager. *Measurement*, *131*, 143-155.

Sahin, Y., & Aydemir, E. (2019). Akıllı telefon teknik ozellik onem derecelerinin AHP ağırlıklı gri ilişkisel analizi yöntemi ile belirlenmesi. *Eskisehir Osmangazi Universitesi İktisadi ve İdari Bilimler Dergisi,* 14(1), 225-238.

Salvio, K. B. V., & Palaoag, T. D. (2019, April). Evaluation of the selected Philippine e-government websites' performance with prescriptive analysis. In

Proceedings of the 2019 5th International Conference on Computing and Artificial Intelligence, 129-137.

Sandeep, M., Kumanan, S., & Vinodh, S. (2011). Supplier selection using combined AHP and GRA for a pump manufacturing industry. *International Journal of Logistics Systems and Management*, *10*(1), 40-52.

Shemshadi, A., Shirazi, H., Toreihi, M., & Tarokh, M. J. (2011). A fuzzy VIKOR method for supplier selection based on entropy measure for objective weighting. *Expert systems with applications*, *38*(10), 12160-12167.

Soylemez, E. Y., Kayabası, A., & Doğan, S. (2021). Seramik sektoru tedarikci secim surecinde gri ilişkisel analiz (GİA) ile entegre edilmis analitik hiyerarsi prosesi (AHP) modeli. *İzmir İktisat Dergisi*, 36(3), 497-516.

Stringam, B., & Gerdes, J. (2019). Service gap in hotel website load performance. *International Hospitality Review.* 33(1), 16-29.

Sun, S., Zhou, S., & Zhao, X. (2013). Analysis of Construction contractor selection and evaluation based on AHP and GRA. In *ICCREM 2013: Construction and Operation in the Context of Sustainability* (pp. 36-44).

Suvvari, A., Durai, S., R., & Goyari, P. (2019). Financial performance assessment using Grey relational analysis (GRA): An application to life insurance companies in India. *Grey Systems: Theory and Application.* 9(4), 502-516.

Tsai, W. H., Chou, W. C., & Leu, J. D. (2011). An effectiveness evaluation model for the web-based marketing of the airline industry. *Expert Systems with Applications*, *38*(12), 15499-15516.

Tzeng, C. J., Lin, Y. H., Yang, Y. K., & Jeng, M. C. (2009). Optimization of turning operations with multiple performance characteristics using the Taguchi method and Grey relational analysis. *Journal of materials processing technology*, *209*(6), 2753-2759.

Uzun, G. (2019). Analysis of grey relational method of the effects on machinability performance on austempered vermicular graphite cast irons. *Measurement*, *142*, 122-130.

Vatansever, K., & Akgűl, Y. (2018). Performance evaluation of websites using entropy and grey relational analysis methods: The case of airline companies. *Decision Science Letters*, 7(2), 119-130.

Wang, B. T., Sheu, T. W., Liang, J. C., Tzeng, J. W., & Nagai, M. (2012, December). Optimal method for choosing an English coursebook by using GRA and AHP. In 2012 *IEEE/SICE International Symposium on System Integration (SII)* (pp. 558-562). IEEE.

Wang, P., Meng, P., Zhai, J. Y., & Zhu, Z. Q. (2013). A hybrid method using experiment design and grey relational analysis for multiple criteria decision making problems. *Knowledge-Based Systems*, *53*, 100-107.

Wang, T. K., Zhang, Q., Chong, H. Y., & Wang, X. (2017). Integrated supplier selection framework in a resilient construction supply chain: An approach via analytic hierarchy process (AHP) and grey relational analysis (GRA). *Sustainability,* 9(2), 289.

Wuwei, L. (2009). Research on evaluation of e-commerce website based on grey relational analysis. *International Forum on Information Technology and Applications* (Vol. 2, pp. 473-476). IEEE.

Yıldırım, M., Kıvanç, B. A. L., & Doğan, M. (2021). Gri İlişkisel Analiz Yöntemi ile Finansal Performans Analizi: Bist'te İşlem Gören Demir Çelik Şirketleri Üzerinde Bir Uygulama. *Muhasebe Bilim Dünyası Dergisi*, *23*(1), 122-143.

Yuan, C., Wu, D., & Liu, H. (2017). Using grey relational analysis to evaluate energy consumption, CO2 emissions and growth patterns in China's provincial transportation sectors. *International journal of environmental research and public health*, 14(12), 1536.

Zerti, O., Yallese, M., Zerti, A., Belhadi, S., & Girardin, F. (2018). Simultaneous improvement of surface quality and productivity using grey relational analysis based Taguchi design for turning couple (AISI D3 steel/mixed ceramic tool (Al2O3+ TiC)). *International Journal of Industrial Engineering Computations*, *9*(2), 173-194.

Zhang, H. (2018). Evalution model of website usability. *In IOP Conference Series: Materials Science and Engineering*, 394 (3), 03253, IOP Publishing.

Zhang, P., Dran, G. M., Small, R. V., & Barcellos, S. (1999). Websites that satisfy users: A theoretical framework for web user interface design and evaluation. *Proceedings of the Hawaii International Conference on System Sciences.*

ACTA INFOLOGICA 2023;7(1):59-70
dergipark.org.tr/acin

DOI: 10.26650/acin.1259067

RESEARCH ARTICLE

Short-Term Sales Forecasting Using LSTM and Prophet Based Models in E-Commerce

E-Ticarette LSTM ve Prophet Esaslı Modeller Kullanarak Kısa Dönemli Satış Tahmini

Alp Ecevit[1], İrem Öztürk[1], Mustafa Dağ[1], Tuncay Özcan[2]

[1] Private corporation

[2](Assoc. Prof.) Istanbul Technical University, Management Engineering Department, Istanbul Technical University, Istanbul, Turkiye

ORCID: A.E. 0000-0003-1685-7642;
İ.Ö. 0000-0002-0264-1798;
M.D. 0000-0003-0291-604X;
T.Ö. 0000-0002-9520-2494

Corresponding author:
Tuncay ÖZCAN
Istanbul Technical University, Management Engineering Department, Istanbul Technical University, Istanbul, Turkiye
E-mail address: tozcan@itu.edu.tr

Submitted: 02.03.2023
Accepted: 14.03.2023
Published Online: 14.04.2023

Citation: Ecevit, A., Ozturk, I., Dag, M., Ozcan, T. (2023). Short-term sales forecasting using LSTM and prophet based models in e-commerce. *Acta Infologica, 7*(1), 59-70. https://doi.org/10.26650/acin.1259067

ABSTRACT

The accuracy of sales forecasting is crucial for e-commerce businesses to optimize inventory management, pricing decisions, marketing strategies and staff scheduling. At this point, different approaches such as statistical models, fuzzy systems, machine learning and deep learning algorithms are widely used for sales forecasting. This study investigates the performance of the deep learning based the Long-Short Term Memory (LSTM) model and the Facebook Prophet model on short-term sales forecasting. The performance of the proposed models is compared with the seasonal autoregressive integrated moving average (SARIMA) using real-life data from an e-commerce site. For the comparative analysis of the proposed forecasting models, weighted average absolute percent error (wMAPE), root mean square error (RMSE) and R-squared are selected as performance measures. The numerical results show that the LSTM model outperforms the Prophet and SARIMA models in terms of forecast accuracy for hourly sales forecasting.

Keywords: Sales forecasting, e-commerce, LSTM, prophet

ÖZ

Satış tahmininin doğruluğu, e-ticaret işletmelerinin envanter yönetimini, fiyatlandırma kararlarını, pazarlama stratejilerini ve personel planlamasını en iyilemesi için çok önemlidir. Bu noktada, satış tahmini için istatistiksel modeller, bulanık sistemler, makine öğrenmesi ve derin öğrenme algoritmaları gibi farklı yaklaşımlar yaygın olarak kullanılmaktadır. Bu çalışma, derin öğrenme tabanlı Uzun-Kısa Süreli Bellek (LSTM) modeli ve Facebook Prophet modelinin kısa vadeli satış tahmini üzerindeki performansını incelemektedir. Önerilen modellerin performansı, bir e-ticaret sitesinden alınan gerçek hayat verileri kullanılarak mevsimsel otoregresif bütünleşik hareketli ortalama (SARIMA) ile karşılaştırılmıştır. Önerilen tahmin modellerinin karşılaştırmalı analizi için, performans ölçütleri olarak ağırlıklı ortalama mutlak yüzde hata (wMAPE), hata kareleri ortalamasının karekökü (RMSE) ve R-kare seçilmiştir. Sayısal sonuçlar, LSTM modelinin saatlik satış tahmini için tahmin doğruluğu açısından Prophet ve SARIMA modellerinden daha iyi performans gösterdiğini göstermiştir.

Anahtar Kelimeler: Satış tahmini, e-ticaret, LSTM, prophet

1. INTRODUCTION

Sales forecasting is a crucial task for e-commerce businesses as it allows them to anticipate future demand and make informed decisions on inventory management, pricing strategies, marketing campaigns and staff scheduling. Accurately predicting future sales allows retailers to optimize their stock and inventory levels, avoiding the costly consequences of both stockouts and excess inventory. Stockouts can lead to customer disappointment and loss of sales, whereas excess inventory can result in increased storage and holding costs, negatively impacting overall profits (Jing & Lewis, 2011). Therefore, retailers use various forecasting models to improve the accuracy of their sales forecasts. At this point, especially, machine learning and deep learning-based models are very popular.

This study presents the performance analysis of the deep learning based the Long-Short Term Memory (LSTM) model and the Prophet model on hourly sales forecasting. The performance of the proposed models is evaluated by comparison with the seasonal autoregressive integrated moving average (SARIMA). For this analysis, e-commerce data of a grocery retailer is used. In addition, weighted average absolute percent error (wMAPE), root mean square error (RMSE) and R-square of these models are calculated to measure prediction performance. In other words, the aim of this study is to perform a comparative analysis of three popular time series forecasting models such as LSTM, Prophet and SARIMA in order to identify the best model for short-term sales forecasting. Recent studies have shown that deep learning frameworks and Prophet have been effective in sales forecasting for e-commerce applications (Ensafi, Amin, Zhang, & Shah, 2022), whereas SARIMA remained a popular choice for traditional time series forecasting (Zhang, 2003). Each model has its own unique strengths and limitations, making it essential to thoroughly compare them in order to determine the most accurate and efficient model for sales forecasting. The findings of this study have practical implications for e-commerce businesses and contributes to the existing literature on sales forecasting in the field.

The rest of this paper is organized as follows: In the next section, a literature review of the existing methods for sales forecasting in e-commerce and retailing is presented. In Section 3, the methods used for sales forecasting and the performance metrics are explained. In Section 4, the dataset and the data preprocessing and application steps performed are described. In Section 5, the performance results obtained are presented and discussed. Finally, the conclusion of the paper is presented along with the suggested future work.

2. RELATED WORKS

The accuracy of sales forecasting significantly influences the marketing, pricing, inventory and scheduling decisions, especially for retailers and e-commerce companies. Loureiro, Miguéis, & da Silva (2018) acknowledged this statement by emphasized the crucial place of sales forecasting for businesses. Therefore, sales forecasting has been an interesting topic for practitioners and academics alike. There are numerous studies related to sales forecasting in the literature. Some of these studies are discussed in this section.

The importance of statistical models such as Autoregressive integrated moving averages (ARIMA) cannot be denied. However, these kinds of statistical forecasting frameworks have certain limitations. ARIMA models do perform well under the condition of data's linearity (Zhang, 2003; Loureiro et al., 2018; Ji, Wang, Zhao, & Guo, 2019; Bandara et al., 2019; Punia, Nikolopoulos, Singh, Madaan, & Litsiou, 2020). As the limitations of statistical models have been acknowledged and machine learning/ deep learning approaches started to become more efficient, Zhang (2003) proposed a hybrid approach that combined ARIMA and Artificial Neural Networks (ANN) intended to capture both the linear and nonlinear relationship between data points. In this study, various datasets have been implemented to test the hybrid model and this architecture has been evaluated with mean squared error (MSE) and mean absolute deviation (MAD). The hybrid model outperformed both ARIMA and ANN models based on the performance metrics mentioned above in all datasets. Sun, Choi, Au, & Yu (2008), showed the usage of neural networks in sales forecasting tasks. This study took advantage of a neural network framework called extreme machine learning (EML) and used it in the fashion retail industry by examining the relationship between sales amount and product-related features. Chang, Liu, & Fan (2009) used a hybrid structure that combined a k-means clustering algorithm with a neural network to predict the sales of circuit boards. In this study, the performance of the model was evaluated by

comparing it with different kinds of models such as back propagation neural networks, radial basis function neural networks. Different from previous studies, Yu, Choi, & Hui (2011) focused on the time-consuming aspect of Artificial Neural Networks and developed an extreme learning machine (ELM) model. Although it has been indicated in this study, that the extreme learning machines are not stable when compared to ANNs, the developed model has been used in the retail fashion industry which sacrificed performance over agility. Choi, Hui, Liu, Ng, & Yu (2014) proposed a hybrid approach of combining the grey model and the extreme learning machine using synthetic fashion retail data in order to predict the sales of the fast fashion industry. In this study, the proposed model was highly appropriate for industries where products are highly seasonal and are constantly changing i.e., there is a low amount of data available for every product point. Arunraj & Ahrens (2015) developed a hybrid autoregressive model to predict the sales of a banana retailer using features such as day of the year, the month of the year, holiday, weather effect, etc. In this study, focused the seasonal autoregressive integrated moving average with external variables (SARIMAX) and hybrid SARIMA and Quantile Regression (SARIMA-QR) that enabled high-low quantile predictions. The results of this study indicated that SARIMA-QR has been more successful in aspects such as business interpretability, accurate decision-making and insightfulness. Zhao & Wang (2017) proposed a CNN-based model that used a one-dimensional convolution layer to predict sales of an e-commerce site using variables such as user actions and content of the product. Loureiro et al. (2018) presented an Artificial Neural Network model to predict the sales of retail products using product-related data. In this study, the proposed model was compared with different models such as Random Forests, Decision Trees and Support Vector Regression.

Another state-of-the-art approach in time series forecasting tasks is to use Recurrent Neural Networks (RNN) and particularly Long Short-Term Memory (LSTM) models. Yu, Wang, Strandhagen, & Wang (2018) created a LSTM model for the product-based sales forecasting. The model achieved strong results in only one-fourth of the products. These results showed that the necessity of a larger amount of data points and a multivariate structure of both the model and the dataset. One of the more recent algorithms that have been used in time series forecasting tasks was Meta's Prophet (Taylor & Letham, 2017). Weytjens, Lohmann, & Kleinsteuber (2021) compared Prophet, ARIMA and LSTM models in predicting cash flows using transaction-related data. Rather than using mean squared error as a performance metric, this study adopted a domain-related measure called the Interest Opportunity Cost (IOC) which is used to evaluate the above models. In this study, numerical results showed that LSTMs both outperformed the statistical ARIMA model and Meta's Prophet. Ji et al. (2019) developed a hybrid model called C-A-XGBoost for an e-commerce company, which combined clustering, ARIMA and XGBoost algorithms that used features such as user actions, sales, weather and holidays. In order to evaluate the performance of the model, metrics such as mean squared error, root mean square error and mean absolute error were used, and numerical results showed that the proposed C-A-XGBoost model performed better than statistical forecasting algorithms. Bandara et al. (2019) created a model that combined LSTM with K-means using product-related, sales-related and time-related features for sales forecasting. In this study, the numerical results showed that the proposed LSTM-based time series forecasting model outperformed linear forecasting models. The usage of statistical methods for time series forecasting has been limited during the last few years and literature started to emphasize the argument that machine learning and deep learning models outperform models like ARIMA, SARIMA and linear regression. Punia et al. (2020) indicated that machine learning models are insufficient in handling the trend and seasonality in the data, and developed a hybrid model that combined LSTM and Random Forests. The performance of the developed model was compared with LSTM, ARIMAX and multiple regression. The proposed hybrid model performed slightly better than all of the other models. Another example of a forecasting study has been conducted by Chandriah & Naraganahalli (2021), created a LSTM-based model using the Modified-Adam algorithm to predict the demand for spare automobile parts. In this study, sales and stock-related features were used in order to train the model, and the model was evaluated using performance metrics of mean error and mean squared error. Ensafi et al. (2022) performed a comparative analysis of a Seasonal Autoregressive Integrated Moving Average (SARIMA), Prophet model, LSTM model and CNN model for predicting the sales of a furniture store. Numerical results showed that the LSTM model seemed to outperform all others, followed by the CNN model and the Prophet model. Zohdi, Rafiee, Kayvanfar, & Salamiraad (2022) used both machine learning and deep learning models to predict the demand for a supply chain management system. This study compared machine learning models such as K-nearest neighbors, decision trees, extreme gradient boosting and multi-layer perceptron

neural network. The numerical results of the study showed that multi-layer perceptron neural network had a stronger performance than the machine learning models mentioned above. Martínez, Charte, Frías, & Martínez-Rodríguez (2022) emphasized the importance of fast runtimes and developed a Generalized Regression Neural Networks (GRNNs) based framework for time series forecasting tasks. The main aim of this study was to create a fast and accurate neural network framework that was able to capture seasonality and trendiness from data. The numerical results of this study showed that GRNNs can be used in order to train time series forecasting models highly fast and accurately.

3. METHODOLOGY

In this section, mathematical details of the Prophet and LSTM algorithms are discussed along with selected performance measures.

3.1. Prophet

The Prophet procedure is a forecasting model developed by Facebook. There are libraries in Python and R for the implementation of Prophet. Meta's Prophet framework is an additive regression model that consists of four components: a trend function *g(t)* that models non-periodic changes such as the growth over time, a seasonality component *s(t)* that shows the periodic changes (i.e., monthly, yearly, weekly changes), a holiday component *h(t)* that merges the effects of holidays to the model and the *e(t)* parameter which represents the idiosyncratic changes that are not foreseen by the model (Taylor & Letham, 2017).

The calculation steps of the Prophet model are as follows (Taylor & Letham, 2017).

Meta's Prophet framework can be expressed by Equation (1).

$$y(t) = g(t) + s(t) + h(t) + e(t) \tag{1}$$

The Prophet has two trend models, one of them being a saturating growth model and the other one is a piecewise linear model. The nonlinear saturating growth model is a model that aims to model and forecast growth. As the authors indicate, the growth model is typically modeled using a logistic growth model. For example, in a case where the capacity of an online site users depends on the number of people that has access to internet, the typical growth model is given in Equation (2):

$$g(t) = \frac{C}{1 + exp\ (-k(t - m))} \tag{2}$$

Where *C* is the capacity of online site users, *k* is the growth rate, and *m* is the offset parameter.

However, this growth model is missing some aspects of growth. First, the capacity is not constant, so in the Prophet framework, constant *C* is replaced by a time-dependent *C* of *C(t)*. Second, the growth rate is not constant either. In order to catch the changes in trend, the model incorporates changepoints where the growth rate has the ability to change. In a case where there are S changepoints at times $s_{j,}$ where *j* represents the values from 1 to *S*, the authors define a vector of rate adjustments: $\delta \in \mathbb{R}^S$. Where δ_j is the change in rate at time s_j. The rate mentioned above at time *t* is the base rate *k*, with the adjustments up to that time *t,* and it can be defined by Equation (3).

$$k + \Sigma_{j:t>s_j}\delta_j \tag{3}$$

The equation above can be defined more clearly by using a vector as in Equation (4);

$$a(t) \in \{0, 1\}^S \tag{4}$$

where

$$a_j(t) = \begin{cases} 1, & if\ t \geq s_j, \\ 0, & otherwise \end{cases} \tag{5}$$

Rate *k* and offset parameter *m* both need to be adjusted when the former is adjusted in order to connect the endpoints of segments. The adjustment as change point *j* is as in Equation (6):

$$\gamma_j = \left(s_j - m - \sum_{l<j} \gamma_l\right) * \left(1 - \frac{\left(k + \sum_{l<j} \delta_l\right)}{\left(k + \sum_{l \leq j} \delta_l\right)}\right) \tag{6}$$

Then, the rate at time *t* becomes $k + a(t)^T \delta$ as seen in Equation (7). So, the growth model is reached as follows:

$$g(t) = \frac{C(t)}{1 + \exp\left(-(k + a(t)^T \delta)\left(t - (m + a(t)^T \gamma)\right)\right)} \tag{7}$$

The second trend model which is called the linear trend with changepoints is a simple linear model with a fixed rate of growth. This model is more suited for problems where there is not a saturating growth, and can be defined by Equation (8).

$$g(t) = (k + a(t)^T \delta)t + (m + a(t)^T \gamma) \tag{8}$$

Where *k* is the growth rate, δ has the rate adjustments and *m* is the offset parameter. To make function continuous, the growth rate can be expressed by Equation (9):

$$\gamma_j = -s_j \delta_j \tag{9}$$

The changepoints can be defined by the user or may be automatically selected by the model. The literature studies indicate it is advised to define a large number of changepoints. The changepoints can be shown as in Equation (10):

$$\delta_j \sim \mathcal{L}(0, \tau) \tag{10}$$

In this equation, τ directly controls the flexibility of the model.

When the model is going to start making predictions about the future using the data from the past, the *g(t)* will have a constant rate. The uncertainty in the forecast trend will be solved by extending the generative model forward. This generative model is composed of *S* changepoints over a history of *T* in which each has a rate change given in Equation (10).

Simulation of future changes is done by replacing τ with variance as in Equation (11).

$$\lambda = \frac{1}{S} \sum_{j=1}^{S} |\delta_j| \tag{11}$$

Future changepoints are randomly sampled in a way that the average frequency of changepoints in history is the same as the ones in the future as in Equation (12).

$$\forall j > T, \begin{cases} \delta_j = 0 \;\; w.p. & \frac{T-S}{T} \\ \delta_j \sim \mathcal{L}(0, \lambda) \;\; w.p. & \frac{S}{T} \end{cases} \tag{12}$$

Thus, uncertainty in the trend is calculated by making the assumption that the average frequency and magnitude of rate changes in history will be the same as in the future.

The second component which is the seasonality component is modeled using the Fourier series. Where *P* is the time period we want our time series to have (e.g. *P* = 7 for a weekly data or *P* = 365.25 for yearly data in daily-scaled data). The seasonality component of the model can be expressed by Equation (13).

$$s(t) = \sum_{n=1}^{N} \left(a_n \cos\left(\frac{2\pi n t}{P}\right) + b_n \sin\left(\frac{2\pi n t}{P}\right)\right) \tag{13}$$

Fitting this seasonality component would require the construction of a matrix of seasonality vectors for each value of t in our data (?). For example, when looking at a yearly seasonality with *N*=10*,* Equation (14) is obtained.

$$X(t) = \left[\cos\left(\frac{2\pi(1)t}{365.25}\right), \ldots, +\sin\left(\frac{2\pi(10)t}{365.25}\right)\right] \tag{14}$$

In this case, the seasonal component can be expressed by Equation (15).

$$s(t) = X(t)\beta \tag{15}$$

In the generative model, Prophet takes $\beta \sim Normal(0,\sigma^2)$ in order to implement a smoothing prior to the seasonality. The more N is increased the more the model's ability to fit seasonality, which changes more rapidly, increases.

The third component ($h(t)$) is for including predictable instances in the model such as national holidays, Valentine's Day, etc. In order to use this feature, the user needs to provide the dates to the model by also specifying the source country of the holiday. For each holiday i, D_i represents the past and future occurrences of the holiday i which then is added by an indicator function that denotes whether the time t is during holiday i. As the last step, the model assigns a parameter of k_i to holiday i that represents the change in the forecast. This procedure is conducted in a similar way as seasonality components which is by generating a matrix of regressors as in Equation (16):

$$Z(t) = [1(t \in D_1), \ldots, 1(t \in D_L)] \tag{16}$$

The holiday component of the Prophet model can be expressed by Equation (17).

$$h(t) = Z(t) * \text{К} \tag{17}$$

and using a prior $\text{К} \sim Normal(0,v^2)$.

3.2. Long-Short Term Memory (LSTM)

Long-Short Term Memory (LSTM) is a particular type of Recurrent Neural Network that is constructed to deal with sequential data. They are generally used for speech recognition, handwriting recognition, and time series forecasting (Weytjens et al. 2021). However, in the traditional RNN there is a gradient vanishing and exploding problem. To overcome gradient vanishing and exploding problems, gated RNNs are introduced which have the ability to learn when and how to forget past data (Goodfellow et al. 2016). One of the most famous gated RNN frameworks is Long-short term memory and the flow of this network is illustrated in Fig. 1.

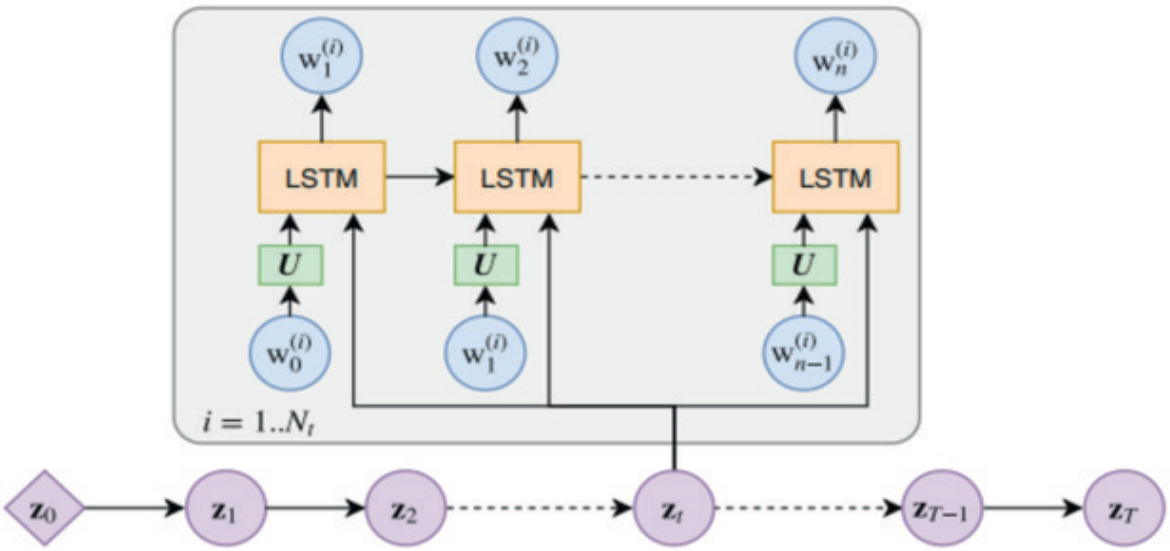


Figure 1. LSTM flow (Delasalles, Lamprier, & Denoyer, 2019)

LSTM includes three stages; forget gate, input gate, and output gate. At the forget gate of f_t,what percentage of the Long-term memory to remember is decided via the sigmoid activation function. Then, at the input gate i_t, short-term memory and the input are used to calculate the potential Long-term memory with hyperbolic tangent function (*tanh*). Similar to the forget gate, the percentage of potential long-term memory to remember is determined with the sigmoid activation function. As a result, Long term memory is updated. At the last stage, the new long-term memory is the input of the *tanh* function to find the potential short-term memory, again the percentage to remember is calculated with the sigmoid function. For a single time step, the model is operated using Equations (18)-(23) (Goodfellow et al. 2016).

In these equations, f_t, i_t, $o_t$represent the forget gate, the input gate and the output gate, respectively. Also, c_t and h_t indicate the cell state and the hidden state, respectively. is used for the sigmoid activation function, while is used for the hyperbolic tangent function.

$$f_t = \sigma_g\,(W_f \times x_t + U_f \times h_{t-1} + b_f) \tag{18}$$

$$i_t = \sigma_g(W_i \times x_t + U_i \times h_{t-1} + b_i) \tag{19}$$

$$o_t = \sigma_g(W_O \times x_t + U_o \times h_{t-1} + b_o) \tag{20}$$

$$c'_t = \sigma_c(W_c \times x_t + U_c \times h_{t-1} + b_c) \tag{21}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot c'_t \tag{22}$$

$$h_t = o_t \cdot \sigma_c(c_t) \tag{23}$$

3.3. Performance Measures

In this study, Root Mean Squared Error (RMSE), Weighted Mean Absolute Percentage Error (wMAPE) and R-squared were used as performance measures which are clearly explained in this section.

Mean squared error (MSE) is computed as the average of the squared differences between the predicted and observed values. Root Mean Squared Error (RMSE) is the square root of MSE. These two performance measures have the same units as the target variable. The main difference between these measures is that MSE effectively penalizes larger errors more severely. Hence, an overshoot may result in quite a high MSE value while RMSE smooths it out. The formulas for MSE and RMSE are given in Equations (24)-(25), respectively.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{24}$$

$$RMSE = \sqrt{MSE} \tag{25}$$

In these equations, Y_i is the observed output vector at *i* th instance and $\hat{Y}_i$ is the predicted output vector at *i* th instance.

wMAPE is another metric used in this study to compare the performance of the proposed approaches. This decision is based on the limitations of commonly used metrics, such as the Mean Absolute Error (MAE) or L1 loss, which is computed as the mean of the differences between the actual measurements and the predictions. While MAE can be useful, it is not a meaningful metric by itself, as it depends on the magnitude of the data. To address these limitations, this study chose to use the Mean Absolute Percentage Error (MAPE) as a more meaningful metric. However, even MAPE has its own limitations, as it tends to become less reliable as the sales increase. For instance, forecasting a sale of $130 instead of $100 may not be as unacceptable as forecasting a sale of $13,000 instead of $10,000.

In this study, to overcome these limitations, wMAPE was chosen as the performance measurement metric. It weighs the error by adding the total sales, making it a more suitable metric for the study. MAE, MAPE and wMAPE can be calculated using Equations (26)-(28), respectively.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i\ | \tag{26}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} |\frac{(y_i - \hat{y}_i)}{y_i}\ | \tag{27}$$

$$WAPE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{\sum_{i=1}^{n} |y_i|} \tag{28}$$

Another performance metric used in this study is R-squared. R-squared is one of the goodness of fit measures. It measures the proportion of variability in *Y* that can be explained using *X*. In other words, it represents the relationship between dependent variables and independent variables. It is equal to the squared correlation between the model and the response (dependent) variable (James et al, 2013). The R-squared can be calculated using Equations (29)-(31).

$$RSS = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{29}$$

$$TSS = \frac{1}{n}\sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{30}$$

$$R^2 = 1 - \frac{RSS}{TSS} \tag{31}$$

In these equations, RSS gives the unexplainable part of the error left after performing the model, while TSS measures the total sum of squares.

4. APPLICATION

In this study, the proposed LSTM, Prophet and SARIMA approaches are applied for hourly sales forecasting with real-life data from an e-commerce site. In this section, the application steps are detailed.

4.1. Data Set

In this study, the real-world dataset that is used to train the models consisted of both numerical features and ordinal features. A sample of the dataset is shown in Fig. 2 and Table 1 with the features of total sales, day of the week, holiday, temperature and weather conditions per timestamp. The dataset used in the study consists of dates per hour starting from December 14, 2022, at 9 a.m. and ending on January 23, 2023, at 9 a.m. with a total of 961 timestamps. In order to maintain data privacy, the values in the data set were scaled.

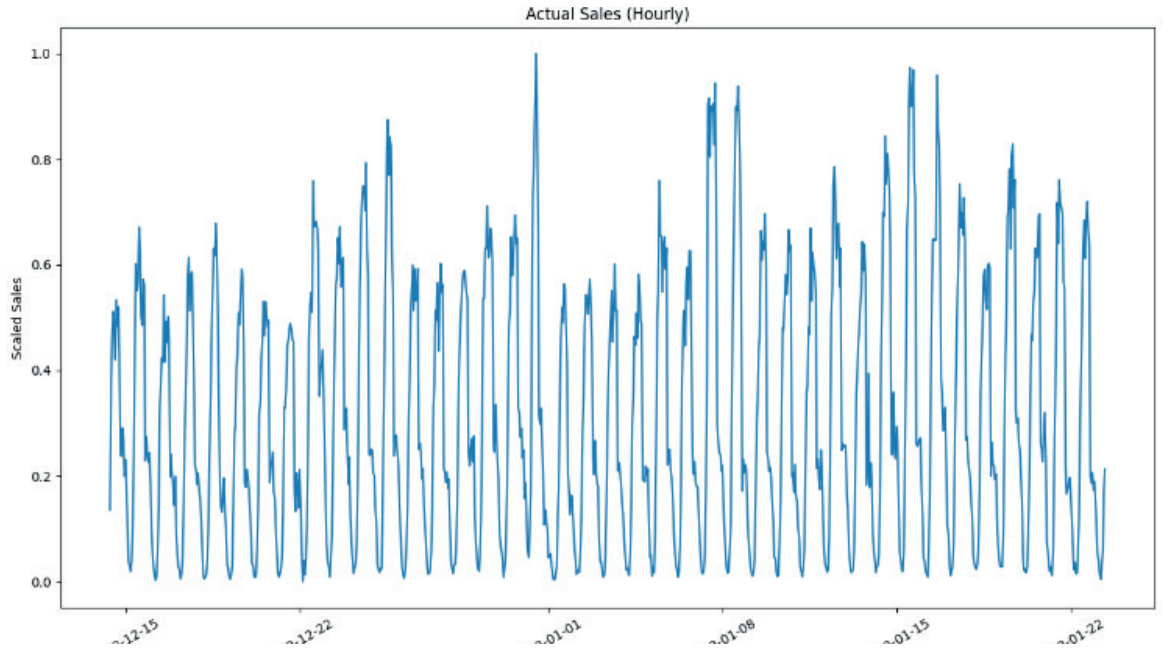


Figure 2. Time series plot of the hourly sales used in this study

Table 1
A sample of the dataset used in this study

Date	Day	Total Sales	Holiday	Temperature	Condition
2022-12-14 9:00:00	4	0.1360	0	10.5	3
2022-12-14 10:00:00	4	0.3830	0	12	4
2022-12-14 11:00:00	4	0.4602	0	12	4
2022-12-14 12:00:00	4	0.5116	0	11.7	3
2022-12-14 13:00:00	4	0.4838	0	12	3
...	...	...	...	...	...
2023-01-23 5:00:00	2	0.0043	0	9.8	2
2023-01-23 6:00:00	2	0.0279	0	10.1	2
2023-01-23 7:00:00	2	0.0581	0	10.5	2
2023-01-23 8:00:00	2	0.1682	0	11	2
2023-01-23 9:00:00	2	0.2129	0	11.5	2

4.2. Short Term Sales Forecasting Using Proposed Models

In this study, the hold-out validation method is used for training the proposed models. 80% of the dataset is used for training the model and the remaining 20% dataset is used for testing. The forecasting performance of the proposed models is compared using R-squared, RMSE, and WMAPE metrics. Since the sales data is scaled using a min-max scaling method, the R-squared calculation is implemented using the observed sales data rather than the scaled values. However, RMSE and WMAPE are calculated using the scaled values due to interpretability reasons.

This study used the Python programming language to train the models. In order to train the models, different Python libraries and/or dependencies are used. For the LSTM model, Tensorflow's Keras API is used; for the Prophet model, the Meta's Prophet package is used, and for the SARIMA model, the statsmodels package is used.

As indicated before, in this study, the SARIMA model is used as a comparison model to evaluate the prediction performance of the LSTM and PROPHET models. Due to the model's univariate structure, only the total sales feature is used for the training process, and other features are ignored. 54 combinations of different parameter values are tried in order to find the

optimal SARIMA model according to AIC values. After the training process, the best parameters (0,1,0), (1,1,1,24) are adopted, and the final model is formed using these best parameters.

After taking the SARIMA model as the baseline model, the Prophet model is trained using all of the features in the data set. Like the SARIMA model, the total sales column is also scaled; however, some additional steps are taken in the training process of the Prophet model. The holiday column of the data is wrapped into the Prophet model's holiday argument. Also, additional regressors (i.e., day of the week, temperature, and weather condition regressors) are added in order to include all of the features in our dataset. The model is trained by implementing both the linear and logistic growth trends.

The LSTM model is trained using all of the features in the data set. After scaling the total sales column, ordinal columns of weather condition and day of the week are encoded as binary features by adding new columns to the data set. Before fitting the data set into the LSTM input layer, the dataset is restructured into a 3D-array in which the timestamps are grouped by days using a shifting method. For instance, the first sample in the modified data set started from the 1st timestamp and finished with the 24th timestamp in which the 25th timestamp's sales data is used as the target variable. This process is replicated for all the timestamps, meaning that the second sample of the modified data set started from the 2nd timestamp and finished with the 25th timestamp where the 26th timestamp's sales data is used as the target variable. So, the shape of the data that included only the independent variables turned into (937, 24, 16) where 937 represented the number of samples (i.e., the grouped days) and the target array is turned into an array with the shape (937,). Representation of the reshaped data can be seen in Fig. 3 in which **t** represents the timestamps at time 1,2,3, etc. and **f** represents the feature index that is between 1 and 16 inclusively.

[[[t1f1, ..., t1f16], [t2f1, ..., t2f16], [t3f1, ..., t3f16], ..., [t24f1, ..., t24f16]] → [t25f1]

[[t2f1, ..., t2f16], [t3f1, ..., t3f16], [t4f1, ..., t4f16], ..., [t25f1, ..., t25f16]] → [t26f1]

...

[[t937f1, ..., t937f16], [t938f1, ..., t938f16], ..., [t960f1, ..., t960f16]]] → [t961f1]

Figure 3. LSTM data structure

In this study, the neural network structure included a Bidirectional LSTM layer, a global pooling layer, and 2 dense layers with relu and linear activation functions. This model is trained for 100 epochs and optimized using the Adam optimizer. In order to prevent overfitting, early stopping callback is added with a monitoring interval of 10 epochs. Different methods are implemented in order to prevent overfitting such as adding a dropout layer, adding l2 regularization, etc. However, adding the early stopping callback showed the best results. The LSTM model is also tested with different architectures, such as using a single dense layer, not using a bidirectional LSTM layer etc. Best results are achieved using the architecture described above.

5. RESULTS

The prediction performance of the proposed models is given in Table 2 for training and testing data. Numerical results showed that the proposed LSTM model outperformed the Prophet and the SARIMA in terms of R-squared, RMSE and WMAPE measures. The results of this study confirmed the argument that deep learning frameworks do perform better for sales forecasting tasks. As mentioned before, metrics for the training set is calculated using 80% of the data, and for the test set's metrics, 20% of the data is used. Amongst the trained models, Prophet performed the worst with a testing R-squared of 0.8243; however, having lower RMSE and WMAPE than the SARIMA model. The lowest error metrics for the test set are achieved with the LSTM model where R-squared, RMSE and WMAPE were 0.9113, 0.0763 and 0.1623, respectively. As this study trained the SARIMA model as the baseline model, implementation of the Prophet model showed worse results than expected. However, since the LSTM model performed the best, the study confirmed the superiority of deep learning frameworks over statistical forecasting methods. The time series plot of the predicted values of the proposed models for the test data is presented in Figure 4.

Table 2
Prediction performance of the proposed models for training and testing data

	Train Set			Test Set		
	R-squared	**RMSE**	**wMAPE**	**R-squared**	**RMSE**	**wMAPE**
LSTM	0.9643	0.0456	0.1040	0.9113	0.0763	0.1623
SARIMA	0.9388	0.0592	0.1314	0.8791	0.0952	0.2216
Prophet	0.8724	0.0805	0.1849	0.8243	0.0941	0.1939

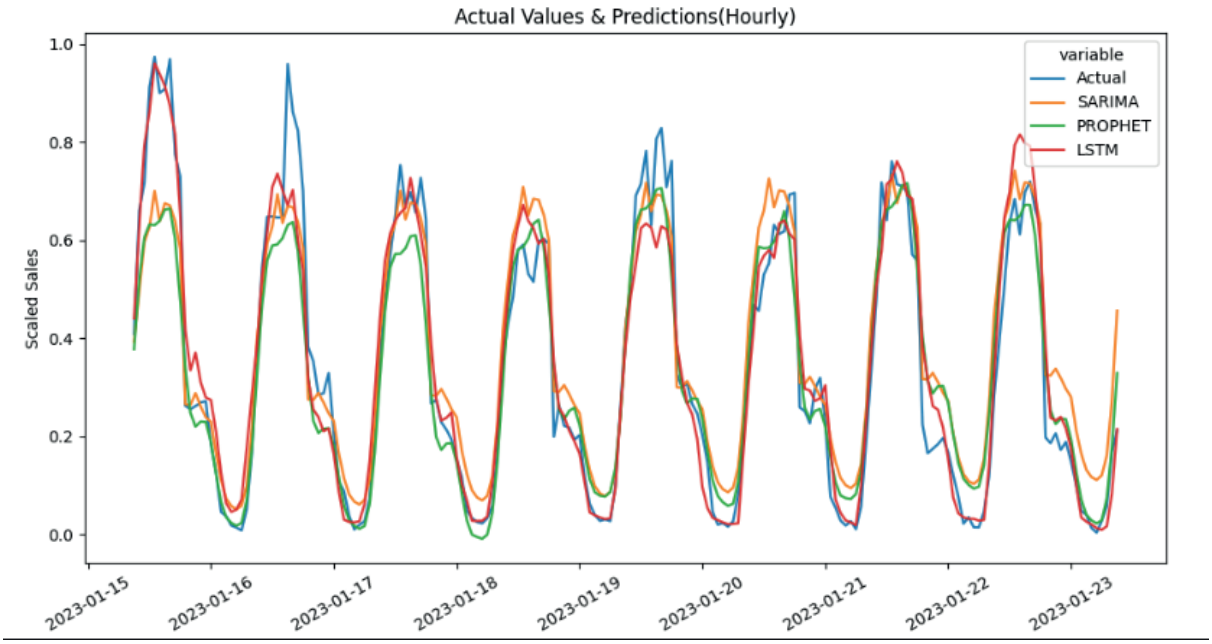


Figure 4. Time series plot of the proposed models for testing data

6. CONCLUSION AND DISCUSSION

In this study, Long-Short Term Memory (LSTM) Networks and the Prophet algorithm are compared with classic statistical approaches in terms of the aforementioned performance metrics using a real-world e-commerce data set. According to the numerical results, the LSTM model outperformed the other models in terms of prediction accuracy. These results showed the superiority of deep learning frameworks over statistical forecasting methods. Regarding the future work on this topic, despite the fact that parameter tuning is already done at the application stage, there is still some room for improvement in optimizing hyper-parameters. Future studies can also focus on using different metaheuristic algorithms and/or increasing the size of the dataset. Especially for deep learning models, a larger dataset tends to improve the model performance. The results of this study may be improved by using a bigger dataset and a more complex neural networks architecture. In this study, weekdays, holidays, weather conditions and temperature are chosen as external variables; however, new features can be added, like campaign and price information, to model extreme points in the data which may lead to an increase in the model's performance. The same method used in this study can also be used to predict the sales of particular product items. Predicting the sales of certain products might have a higher business value which might help industries to plan the supply and stock levels of their product items.

Peer-review: Externally peer-reviewed.
Author Contributions: Conception/Design of Study- A.E., İ.Ö., M.D., T.Ö.; Data Acquisition- A.E., İ.Ö., M.D.; Data Analysis/Interpretation- A.E., İ.Ö., M.D., T.Ö.; Drafting Manuscript- A.E., İ.Ö., M.D.; Critical Revision of Manuscript- T.Ö.; Final Approval and Accountability- A.E., İ.Ö., M.D., T.Ö.; Material and Technical Support- A.E., İ.Ö., M.D.; Supervision- T.Ö.

Conflict of Interest: The authors have no conflict of interest to declare.
Grant Support: The authors declared that this study has received no financial support.

REFERENCES

Arunraj, N. S., & Ahrens, D. (2015). A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting. *International Journal of Production Economics*, *170*, 321–335. https://doi.org/10.1016/j.ijpe.2015.09.039

Bandara, K., Shi, P., Bergmeir, C., Hewamalage, H., Tran, Q., & Seaman, B. (2019). Sales Demand Forecast in E-commerce Using a Long Short-Term Memory Neural Network Methodology. In T. Gedeon, K. W. Wong, & M. Lee (Eds.), *Neural Information Processing* (pp. 462–474). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-36718-3_39

Chandriah, K. K., & Naraganahalli, R. V. (2021). RNN / LSTM with modified Adam optimizer in deep learning approach for automobile spare parts demand forecasting. *Multimedia Tools and Applications*, *80*(17), 26145–26159. https://doi.org/10.1007/s11042-021-10913-0

Chang, P.-C., Liu, C.-H., & Fan, C.-Y. (2009). Data clustering and fuzzy neural network for sales forecasting: A case study in printed circuit board industry. *Knowledge-Based Systems*, *22*(5), 344–355. https://doi.org/10.1016/j.knosys.2009.02.005

Choi, T.-M., Hui, C.-L., Liu, N., Ng, S.-F., & Yu, Y. (2014). Fast fashion sales forecasting with limited data and time. *Decision Support Systems*, *59*, 84–92. https://doi.org/10.1016/j.dss.2013.10.008

Delasalles, E., Lamprier, S., & Denoyer, L. (2019). Dynamic Neural Language Models. In T. Gedeon, K. W. Wong, & M.

Lee (Eds.), *Neural Information Processing* (pp. 282–294). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-36718-3_24

Ensafi, Y., Amin, S. H., Zhang, G., & Shah, B. (2022). Time-series forecasting of seasonal items sales using machine learning – A comparative analysis. *International Journal of Information Management Data Insights*, *2*(1), 100058. https://doi.org/10.1016/j.jjimei.2022.100058

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Sequence Modeling: Recurrent and Recursive Nets. In *Deep learning* (pp. 373–420). Cambridge, Massachusetts: The MIT Press.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). Linear Regression. In *An introduction to statistical learning: With applications in R* (pp. 59–128). New York: Springer.

Ji, S., Wang, X., Zhao, W., & Guo, D. (2019). An Application of a Three-Stage XGBoost-Based Model to Sales Forecasting of a Cross-Border E-Commerce Enterprise. *Mathematical Problems in Engineering*, *2019*, 1–15. https://doi.org/10.1155/2019/8503252

Jing, X., & Lewis, M. (2011). Stockouts in Online Retailing. *Journal of Marketing Research*, *48*(2), 342–354. https://doi.org/10.1509/jmkr.48.2.342

Loureiro, A. L. D., Miguéis, V. L., & da Silva, L. F. M. (2018). Exploring the use of deep neural networks for sales forecasting in fashion retail. *Decision Support Systems*, *114*, 81–93. https://doi.org/10.1016/j.dss.2018.08.010

Martínez, F., Charte, F., Frías, M. P., & Martínez-Rodríguez, A. M. (2022). Strategies for time series forecasting with generalized regression neural networks. *Neurocomputing*, *491*, 509–521. https://doi.org/10.1016/j.neucom.2021.12.028

Punia, S., Nikolopoulos, K., Singh, S. P., Madaan, J. K., & Litsiou, K. (2020). Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail. *International Journal of Production Research*, *58*(16), 4964–4979. https://doi.org/10.1080/00207543.2020.1735666

Sun, Z.-L., Choi, T.-M., Au, K.-F., & Yu, Y. (2008). Sales forecasting using extreme learning machine with applications in fashion retailing. *Decision Support Systems*, *46*(1), 411–419. https://doi.org/10.1016/j.dss.2008.07.009

Taylor, S. J., & Letham, B. (2017). *Forecasting at scale* (No. e3190v2). PeerJ Inc. https://doi.org/10.7287/peerj.preprints.3190v2

Weytjens, H., Lohmann, E., & Kleinsteuber, M. (2021). Cash flow prediction: MLP and LSTM compared to ARIMA and

Prophet. *Electronic Commerce Research*, *21*(2), 371–391. https://doi.org/10.1007/s10660-019-09362-7

Yu, Q., Wang, K., Strandhagen, J. O., & Wang, Y. (2018). Application of Long Short-Term Memory Neural Network to

Sales Forecasting in Retail—A Case Study. In K. Wang, Y. Wang, J. O. Strandhagen, & T. Yu (Eds.), *Advanced Manufacturing and Automation VII* (pp. 11–17). Singapore: Springer. https://doi.org/10.1007/978-981-10-5768-7_2

Yu, Y., Choi, T.-M., & Hui, C.-L. (2011). An intelligent fast sales forecasting model for fashion products. *Expert Systems with Applications*, *38*(6), 7373–7379. https://doi.org/10.1016/j.eswa.2010.12.089

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159-175. https://doi.org/10.1016/S0925-2312(01)00702-0

Zhao, K., & Wang, C. (2017, August 26). *Sales Forecast in E-commerce using Convolutional Neural Network*. arXiv. https://doi.org/10.48550/arXiv.1708.07946

Zohdi, M., Rafiee, M., Kayvanfar, V., & Salamiraad, A. (2022). Demand forecasting based machine learning algorithms on customer information: An applied approach. *International Journal of Information Technology*, *14*(4), 1937–1947. https://doi.org/10.1007/s41870-022-00875-3

ACTA INFOLOGICA 2023;7(1):71-80
dergipark.org.tr/acin

İSTANBUL UNIVERSITY PRESS

DOI: 10.26650/acin.1234106

RESEARCH ARTICLE

Detection of Orienting Response to Novel Sounds in Healthy Elderly Subjects: A Machine Learning Approach Using EEG Features

Sağlıklı Yaşlı Bireylerde Yeni Seslere Yönlendirme Yanıtının Tespiti: EEG Özelliklerini Kullanan Bir Makine Öğrenimi Yaklaşımı

Emine Elif Tülay[1]

ABSTRACT

Unexpected events in the environment elicit the orienting response that protects humans from dangerous situations and there is great importance in identifying these events, especially in aging. The aims of the current study are attempting to find which classification model exhibits the best performance by means of event-related spectral perturbation (ERSP) features based on EEG and to understand which frequency bands, and time windows, contribute most to the classification of external stimuli. The data of 20 healthy elderly participants were included in the study and the 3-Stimulation auditory oddball paradigm was applied to participants. Different classifiers including Support Vector Machine (SVM) with Linear and Polynomial kernels, Linear Discriminant Analysis (LDA), and Naive Bayes were fed by ERSP features obtained from varying frequency bands and time domains. The classification process was fulfilled using custom-written scripts via the FieldTrip Toolbox (version no: 20220104) integrated with the MVPA-light toolbox running under Matlab R2018b. The best performance was obtained by linear SVM which was fed by theta response (4 – 8 HZ) in the early time window (0.1 – 0.5 s) with 90% accuracy in the case of standard stimuli distinguished from novel stimuli. Delta responses also exhibit distinctive characteristics for standard and novel stimuli by running LDA (87% accuracy) and polynomial SVM (86% accuracy). These findings show that the delta and theta responses have contributed to detecting standard and novel sounds with remarkable performances of SVM and LDA.

Keywords: Delta, theta, auditory stimuli, machine learning

ÖZ

Çevrede meydana gelen beklenmedik olaylar, insanı tehlikeli durumlardan koruyan yönlendirici tepkiyi ortaya çıkarır ve bu olayların tespit edilmesi özellikle yaşlanma sürecinde büyük önem taşır. Mevcut çalışmanın amacı, EEG'ye dayalı olaya ilişkin spektral pertürbasyon (ERSP) özellikleri aracılığıyla hangi sınıflandırma modelinin en iyi performansı gösterdiğini bulmaya çalışmak ve hangi frekans bantlarının ve zaman pencerelerinin dış uyaranın sınıflandırılması için en çok katkıda bulunduğunu anlamaktır. 20 sağlıklı yaşlı katılımcının verileri çalışmaya dahil edilmiştir ve katılımcılara 3-Stimülasyon işitsel oddball paradigması uygulanmıştır. Lineer ve Polinom çekirdek fonksiyonlu Destek Vektör Makinesi (DVM), Lineer Diskriminant Analizi (LDA) ve Naive Bayes gibi farklı sınıflandırıcılar, değişen frekans bantlarından ve zaman alanlarından elde edilen ERSP öznitelikleri ile beslenmiştir. Sınıflandırma işlemi, Matlab R2018b altında çalışan MVPA-light araç kutusu ile entegre FieldTrip Toolbox (sürüm no: 20220104) aracılığıyla özel yazılmış komutlar kullanılarak gerçekleştirilmiştir. En iyi performans erken zaman penceresinde (0.1 – 0.5 s) teta yanıtı (4 – 8 HZ) ile beslenen lineer DVM tarafından standart uyaranların yeni uyaranlardan ayırt edilmesi durumunda %90 doğrulukla elde edilmiştir. Delta yanıtları ayrıca LDA (%87 doğruluk) ve polinom DVM (%86 doğruluk) çalıştırarak standart ve yeni uyaranlar için ayırt edici özellikler sergilemektedir. Bu bulgular, delta ve teta yanıtlarının, DVM ve LDA'nın dikkate değer performanslarıyla standart ve yeni seslerin algılanmasına katkıda bulunduğunu göstermektedir.

Anahtar Kelimeler: Delta, teta, işitsel uyaran, makine öğrenmesi

[1] (Assist. Prof.) Mugla Sitki Kocman University, Faculty of Engineering, Department of Software Engineering, Muğla, Turkiye

ORCID: E.E.T. 0000-0003-0150-5476

Corresponding author:
Emine Elif TÜLAY
Mugla Sitki Kocman University, Faculty of Engineering, Department of Software Engineering, Muğla, Turkiye
E-mail address: eliftulay@mu.edu.tr

Submitted: 16.01.2023
Revision Requested: 16.01.2023
Last Revision Received: 16.03.2023
Accepted: 14.03.2023
Published Online: 25.04.2023

Citation: Tulay, E.E. (2023). Detection of orienting response to novel sounds in healthy elderly subjects: A machine learning approach using EEG features. *Acta Infologica, 7*(1), 71-80. https://doi.org/10.26650/acin.1234106

1. INTRODUCTION

Human brains are affected by different types of stimuli and produce stimuli-specific responses. According to the literature, unexpected novel events like a car horn in traffic elicit the Orienting Response (OR), which leads to the automatic detection of sudden changes in the environment (Debener, Makeig, Delorme, & Engel, 2005; Berti, Vossel, & Gamer, 2017) by disrupting ongoing thoughts and actions, and prepares the human on a physiological, behavioral, and cognitive level (Lynn, 1966). In other words, the sensory change detection mechanism based on the OR protects humans from dangerous situations despite the fact that a considerable number of studies support that OR is influenced by habituation (Barry, 2009; Cavanagh, Kumar, Mueller, Richardson, & Mueen, 2018).

Although there are many neuroimaging techniques like fMRI, PET/SPECT, or MEG, EEG is the most widely used to investigate brain responses in the cognitive domain due to its being technically and economically more practical. Moreover, different neurophysiological features like Event-related Potential (ERP) and Event-related oscillations (EROs) were investigated to understand the mechanism of attentional processes (Başar-Eroğlu, Başar, Demiralp, & Schürmann, 1992; Basar, Demiralp, Schürmann, Basar-Eroglu, & Ademoglu, 1999; Başar, Başar-Eroglu, Karakaş, & Schürmann, 2001; Başar-Eroğlu & Demiralp, 2001; Berti et al., 2017; Behforuzi et al., 2019) of rare stimuli in various paradigms. Especially event-related delta and theta oscillations have been associated with perception and attention (Harmony, 2013; Güntekin & Başar, 2016; Karakaş, 2020). There are a number of feature extraction methods available for measuring ERO. Event-related spectral perturbation (ERSP) (Delorme & Makeig, 2004) is one of the most important measurements of oscillatory activities to understand the mechanisms of cognitive processes (Makeig, 1993; Wei, Zhao, Yan, Duan, & Li, 1998).

Neuro-cognitive processes for unexpected events are affected by age (Li & Lindenberger, 2002; Berti et al., 2017). According to the studies based on ERO in the literature, the decrease of delta ERO in a visual oddball paradigm (Emek-Savaş, Güntekin, Yener, & Başar, 2016) and cued Go/Nogo Paradigm (Schmiedt-Fehr & Başar-Eroglu, 2011) was associated with aging. On the other hand, the findings of Huizeling, Wang, Holland, & Kessler (2021) revealed that older adults present different ERO patterns during attentional control to compensate for cognitive decline. Moreover, Schmiedt-Fehr, Dühl, & Basar-Eroglu (2011) supported that there may be modality-specific changes with age, and the brain responses of older adults could be affected more during visual stimuli in comparison to auditory stimuli in early stages, whereas Ho et al., (2012) found that the healthy elderly group showed higher delta power than the young group during auditory stimuli.

When the findings in the literature were considered, it could be said that there is no consensus about the mechanism of attentional processing of elderly subjects among researchers. Especially with the rise in machine learning (ML) approaches and methods that have outstanding robustness and adaptability, researchers can objectively and efficiently differentiate neural responses to external stimuli (Saeidi et al., 2021). In the last decade, various ML techniques have been applied to EEG signals for understanding affective processing (Alarcao & Fonseca, 2017; Wang & Wang, 2021; Rahman et al., 2021), attentional processing (Lotte et al., 2018), and even for detecting medical conditions (Hosseini, Hosseini, & Ahi, 2021; Chung & Teo, 2022). Among the different features of EEG data used in the literature, ERO in the Time-Frequency domain provides more information about temporal, spectral, and spatial dynamics of cognitive processes (Aliakbaryhosseinabadi, Kamavuako, Jiang, Farina, & Mrachacz-Kersting, 2019). However, various studies showed that ERP features also had promising classification performances for stimuli classification (Parvar et al., 2014; Tjandrasa & Djanali, 2018; Akhter, Lawal, Tanvir, & Ahmed, 2020; Borra & Magosso, 2021).

The current study aimed to examine mainly two aspects of the classification of three sound stimuli applied to healthy elderly subjects. First, to attempt to find which classification model exhibits the best performance by means of ERSP features. Second, to understand which frequency bands, and time windows, contribute most to the classification of external stimuli. For this purpose, different classifiers were run on delta and theta responses at different time windows (please see Section 2.3). It was hypothesized that novel sound stimuli would be differentiated from the other type of sound stimuli with the delta and theta powers, especially in the early time window. Also, it was estimated that higher performances would be obtained by means of Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA) classifiers.

2. METHODS AND MATERIALS

2.1 Dataset

In the current study, a public dataset (Cavanagh, 2021) was adopted (available at https://openneuro.org/datasets/ds003490/versions/1.1.0) including EEG signals of 25 healthy elderly subjects. The data of 20 participants (8 female, 12 male) were included in the study, disregarding the remaining 5 participants due to persistent artifacts or noise in their recordings. Out of these 5 participants, the epoch numbers of 2 participants were below 20, and the data of 3 participants were too noisy (muscle artifacts and spikes). The mean age was 68.8 (Standard deviation: 17.5) years. The inclusion criteria for the participants were as follows; Mini-Mental State Exam (MMSE) score ≥ 26. All participants provided written informed consent (Cavanagh et al., 2018).

The 3-auditory oddball paradigm was applied to participants during EEG recording. In the paradigm, there were three stimuli named standard (440 Hz sinusoidal tones, and 80 dB), target (660 Hz sinusoidal tones, and 80 dB), and novel distractors that are naturalistic sounds (Bradley & Lang, 1999), varying with each presentation (65 dB with an inter-quartile range of +/− 6.5 dB). The total number of stimuli is 200, including 140 standard, 30 target, and 30 novel. Each stimulus was presented for 200 ms and a random inter-trial interval (ITI) was selected from a uniform distribution of 0.5 to 1 second for the novel condition, and 950 to 1450 ms for both the standard and target conditions. The subjects mentally counted the target sounds (Cavanagh et al., 2018).

The EEGs of participants were recorded using the 64 channel Brain Vision system with a sampling rate of 500 Hz. During the recording, a CPz electrode served as reference, and an AFz electrode served as ground.

2.2 EEG Data Analysis

All the steps of EEG data analysis including preprocessing, feature extraction, and classification were fulfilled using custom-written scripts via the FieldTrip Toolbox (version no: 20220104) (Oostenveld, Fries, Maris, & Schoffelen, 2011) running under Matlab R2018b (MathWorks, Natick, MA, U.S.A.).

2.2.1 Preprocessing

Unlike the reference study by Cavanagh et al. (2018), a different pre-processing pipeline has been applied to data. The processing steps were described below;

1- *Importing data*: All conditions for each subject were imported to the Matlab platform by segmenting the data around the stimulus onset (−2000 to 2000 ms) and selecting 30 channels (F7, F5, F3, F1, Fz, F2, F4, F6, F8, T7, C5, C3, C1, Cz, C2, C4, C6, C8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, O1, Oz, O2) apart from 64 channels. During the importing process, the data were re-referenced to an average reference of selected channels. Also, each epoch was baseline corrected according to the mean amplitude of−200 to 0 ms pre-stimulus time window. Finally, the discrete Fourier transform (DFT) filter was applied for removing the line noise.

2- *Removing artifacts:* The trials that contain fast muscular artifacts, jumps, and uncommon patterns were eliminated manually to enhance the efficiency of the next step, Independent Component Analysis (ICA).

3- *Applying ICA:* To remove eye movements, heartbeat effects, and spiky patterns, the fast ICA method[1] was used. After decomposing the data into sources, components that spectrally and topographically correspond to related artifacts were selected and removed, and the cleaned data was reconstructed.

4- *Detrending:* In this step, slow low-frequency drifts were removed per trial.

5- *Divide data into separate conditions:* The data that include all types of stimuli (target, standard, novel) were divided into separate sub-data based on the conditions per participant.

1 https://github.com/fieldtrip/fieldtrip/blob/master/external/fastica/fastica.m

6- *Removing residual bad trials:* The trials in each sub-data were checked by visual inspection, and bad trials were manually removed from the data.

7- *Equalizing the number of epochs among conditions:* The trials were then matched between conditions for each participant separately. The minimum number of trials was determined across all conditions per participant and after calculating the number of remaining trials for any conditions that were large, the calculated number of randomly selected trials was removed from the sub-data. In case a participant had less than 20 trials for a condition, the data were excluded from the dataset.

2.2.2 Feature Extraction: ERSP Analysis

To obtain the feature sets based on ERSP values for the classifier, the sensor-level time-frequency decomposition with complex Morlet wavelets was conducted on each trial of clean data for each condition, each selected channel (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz, O2), and for all participants. A Morlet function with three cycles was used to calculate the time-frequency transform between 1 and 15 Hz with 0.5 Hz frequency resolution, and the wavelet coefficients were estimated for 10 ms steps between −2000 s and 2000 ms. Time-frequency representations (TFRs) were baseline-corrected with respect to the −1000 to −500 ms pre-stimulus period for slow oscillations (delta and theta bands).

2.3 Classification Analysis

The Fieldtrip integrated MVPA-light toolbox (Treder, 2020) was used to assess whether the ERSP features of EEG could be used to detect novel stimuli in healthy elderly subjects. For this purpose, several classifiers were employed, including Linear SVM, Polynomial SVM, LDA, and Naive Bayes (NB) with default hyperparameters[2] of the MVPA-light toolbox. These hyperparameters can be listed as follows: For SVM, a default search grid was used to automatically determine the best c parameter. When the polynomial kernel was used, gamma was set to 1/(number of features), coef0 was set to 1 and degree was set to 2. For LDA, shrinkage regularization was used and the shrinkage regularization parameter (lambda) was calculated automatically using the Ledoit-Wolf formula[3]. For NB, probabilities were modeled using Gaussians and every class had an equal probability. The class means and variances were estimated for every feature for training. At testing time, the maximum a posteriori (MAP) rule was applied to assign a sample to the class with the maximum posterior probability.

Prior to training the classifier, nested z-scoring was used to avoid the flow of information from the test set flowing into the processing of the train set. Also, a 10-fold cross-validation method was used to train classifiers that were applied for the average of time-frequency points in a specific time and frequency ranges using selected EEG channels as features. The training process with cross-validation was repeated 5 times with new randomly assigned folds to obtain robust results. After that, all test folds and repetitions were averaged to reach the final result. The performance of the classifier was evaluated using the confusion matrix and accuracy metrics.

During the classification analysis, the spectral-temporal searchlight analysis was implemented by using different frequency bands, and time of interest to distinguish novel stimuli from target and standard stimuli. For each classification model, EEG channels were included as features in the delta (1.5 – 4 Hz) frequency band with 0.1– 0.7 s time windows of interest, theta (4 – 8 Hz and 5.5 – 8 Hz) with 0.1 – 0.5 s and 0.5-0.7 s time windows of interest. The time windows and frequency ranges were determined by visual inspection of grand averages (please see Fig. 1 and Fig. 2).

3. RESULTS

Fig. 1 and Fig. 2 depict the grand averages of ERSP values in the delta and theta frequency bands, respectively, over the average across all electrodes (please see Section 2.2.2) for all types of sound stimuli applied to healthy elderly subjects. In both Figures, the left plot represents target stimuli, the middle plot represents standard stimuli, and the right box represents novel stimuli.

2 https://github.com/fieldtrip/fieldtrip/blob/master/ft_statistics_mvpa.m
3 https://github.com/treder/MVPA-Light/blob/master/external/cov1para.m

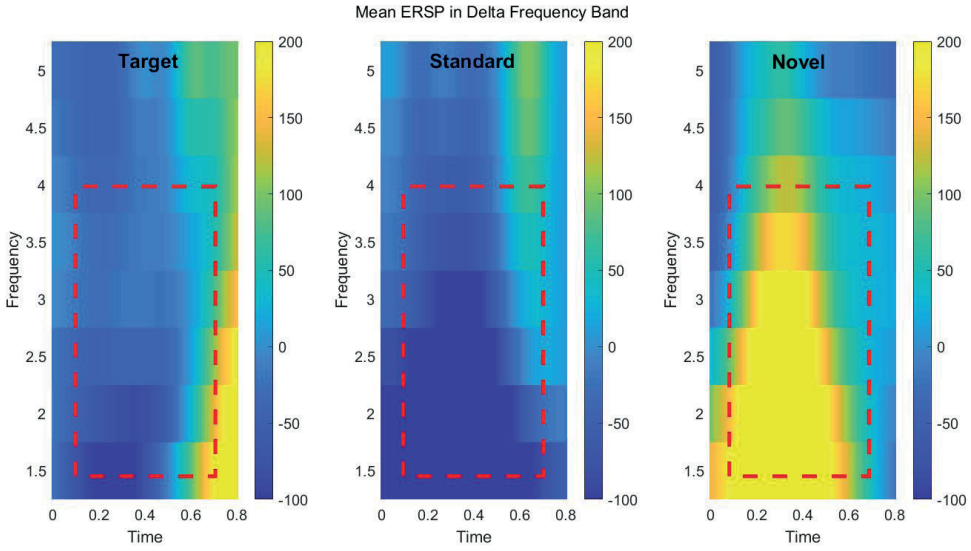


Figure 1. Time-Frequency Representations for all stimuli: Grand average of ERSP values in Delta frequency band over the averaged across all electrodes. Red dashed lines represent the ranges for frequency (1.5 – 4 Hz) and time (0.1 – 0.7 s) that were used by the classifiers as features.

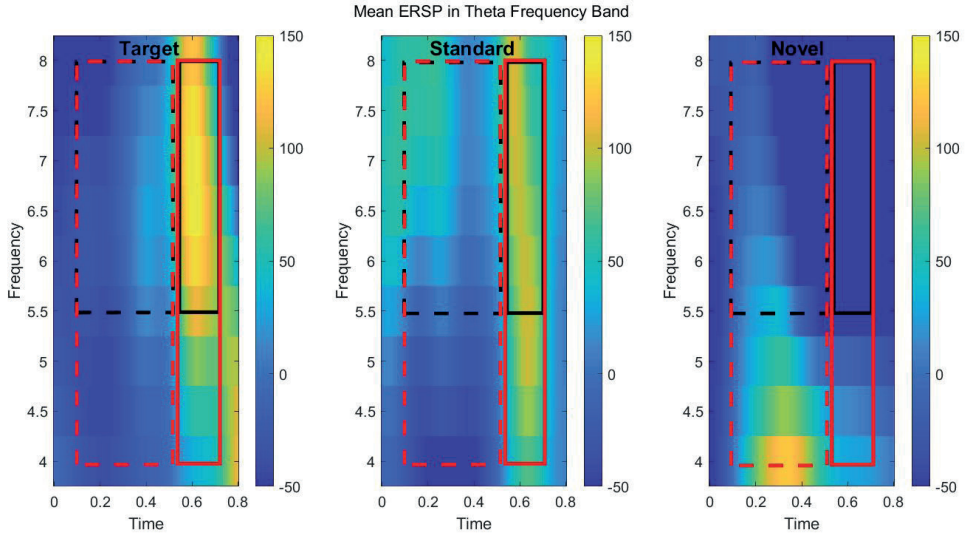


Figure 2. Time-Frequency Representations for all stimuli: Grand average of ERSP values in Theta frequency band over the averaged across all electrodes. Red dashed lines represent the ranges for frequency (4 – 8 Hz) and time (0.1 – 0.5 s), black dashed lines represent the ranges for frequency (5.5 – 8 Hz) and time (0.1 – 0.5 s), red solid lines represent the ranges for frequency (4 – 8 Hz) and time (0.5 – 0.7 s), black solid lines represent the ranges for frequency (5.5 – 8 Hz) and time (0.5 – 0.7 s). The time-frequency points in all these ranges were used by the classifiers as features separately.

The plots in Fig. 1 and Fig. 2 give us clues for the frequency and time ranges in order to perform spectral-temporal searchlight analyses for the classification process. Whereas one time-frequency range was used in the delta band (Fig. 1), four time-frequency ranges were determined in the theta band (Fig. 2). The ranges were also determined in light of the literature given in Section 1.

Fig. 3 depicts the results of the classification analyses by means of the accuracy metric for Linear SVM, Polynomial SVM, LDA, and NB classifiers (the bars from left to right) where different time-frequency points were used as features. The first (front) row shows the performances for delta (1.4-4 Hz) as the frequency of interest at 0.1-0.7 s times of interest, the second

row shows the performances for theta (4-8 Hz) as the frequency of interest at 0.1-0.5 s times of interest, the third row shows the performances for theta (4-8 Hz) as the frequency of interest at 0.5-0.7 s times of interest, the fourth row shows the performances for theta (5.5-8 Hz) as the frequency of interest at 0.1-0.5 s times of interest, and the last (at the back) row shows the performances for theta (5.5-8 Hz) as the frequency of interest at 0.5-0.7 s times of interest. Fig. 3A represents the accuracy values for classifying target and novel stimuli, whereas Fig. 3B represents the accuracy values for classifying standard and novel stimuli.

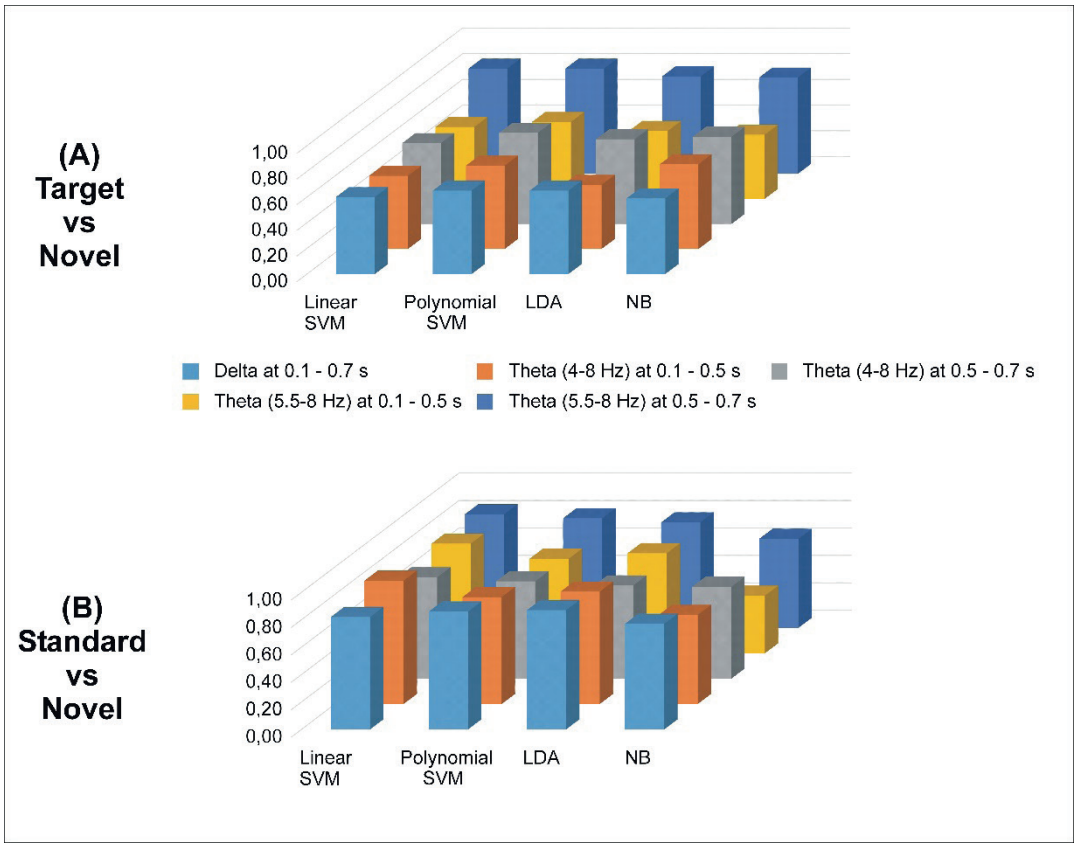


Figure 3. The classification performances by means of accuracy metric for Linear SVM, Polynomial SVM, LDA, and NB classifiers where different times and frequencies were used as features. (a) The performances to distinguish target and novel stimuli (b) The performances to distinguish standard and novel stimuli.

As seen from the bar graph (Fig. 3A), the best feature of EEG to distinguish novel stimuli from target stimuli is late theta (5.5 – 8 Hz at 0.5-0.7s) via all classifier models. Fig. 2 also supports this finding. However, early theta (0.1 – 0.5 s) in the same frequency range could not show the same performance. Moreover, when the theta frequency band was taken wider (4 – 8 Hz), the performances could not reach higher levels at both time windows. To distinguish novel stimuli from standard stimuli, whereas the performances of linear SVM, polynomial SVM, and LDA were close to each other, NB had a lower performance for all feature sets (Fig. 3B). Fig. 3A and 3B clearly showed that the performances of all classifiers were higher for distinguishing novel stimuli from standard stimuli than distinguishing novel stimuli from target stimuli by means of almost all feature sets, except late theta (5.5 – 8 Hz at 0.5-0.7s). All accuracies and confusion matrices (TP and TN rates) of the classification analysis obtained by running searchlight analyses by means of the temporal and spectral features of ERSP are summarized in Table 1.

Table 1

Classification metrics after running spectral-temporal searchlight analyses

Classes	Frequency Bands	Time Windows	Linear SVM		Polynomial SVM		LDA		NB	
			TP/TN (%)	Acc	TP/TN (%)	Acc	TP/TN (%)	Acc	TP/TN (%)	Acc
Target vs Novel	Delta	0.1 - 0.7 s	70/55	0.60	82/40	0.65	80/57	0.65	48/67	0.59
	Theta 4-8 Hz	0.1 - 0.5 s	61/58	0.57	78/54	0.65	67/36	0.50	36/89	0.66
		0.5- 0.7 s	63/66	0.63	67/76	0.71	77/52	0.66	53/83	0.68
	Theta 5.5-8 Hz	0.1 - 0.5 s	59/53	0.56	75/46	0.60	66/43	0.53	47/48	0.50
		0.5- 0.7 s	75/85	0.82	75/87	0.82	76/80	0.76	60/85	0.75
Standard vs Novel	Delta	0.1- 0.7 s	88/69	0.82	89/86	0.86	90/84	0.87	74/83	0.77
	Theta 4-8 Hz	0.1- 0.5 s	95/87	0.90	84/73	0.78	94/72	0.82	43/79	0.65
		0.5- 0.7 s	68/75	0.74	66/86	0.71	80/52	0.68	46/86	0.67
	Theta 5.5-8 Hz	0.1 - 0.5 s	80/79	0.80	87/52	0.69	85/59	0.73	44/36	0.42
		0.5- 0.7 s	81/79	0.83	77/85	0.80	95/66	0.77	44/88	0.65

Acc: Accuracy, TP/TN: True Positive/True Negative

The findings revealed that the best performance (0.90 accuracy) was achieved with Linear SVM for the classification of novel stimuli and standard stimuli by means of early theta (4 – 8 Hz at 0.1 – 0.5 s). Also, the TP and TN values of the confusion matrix showed that the classifier was better at predicting standard stimuli (95%) than it was at predicting novel stimuli (87%). When the delta frequency band was used as a feature set, the performance of LDA (0.87 accuracy) and polynomial SVM (0.86 accuracy) surpassed the performance of linear SVM. Among all confusion matrix values for the classification of novel and standard stimuli, the most balanced prediction rates were achieved by polynomial SVM where the delta frequency band was used as the feature set (Table 1).

Moreover, both SVM classifiers were quite successful to differentiate two rare stimuli, target and novel. Linear SVM and polynomial SVM were much better at predicting novel stimuli (85%, and 87% respectively) than at predicting target stimuli (75% for both) with a 0.82 accuracy rate.

4. DISCUSSION AND CONCLUSION

The current study evaluated the delta and theta responses upon application of the 3-Stimulation auditory oddball paradigm by means of ERSP measurement, and the classification analyses were performed with varying selected features in frequency and time domains to understand which features have a higher contribution at identifying novel sounds. Moreover, different classifiers were used including SVM with Linear and Polynomial kernels, LDA, and NB to reveal which classifier exhibits good distinction for unexpected stimuli with which features. To the best of my knowledge, this is the first study that attempts to detect novel sounds with different traditional machine learning techniques fed with the ERSP features of EEG in various time-frequency ranges.

The most remarkable performances of the current study were obtained in the case that standard stimuli distinguished from novel stimuli and SVM, with both kernel types showing better performances followed by LDA. Furthermore, the other important observations of the spectral-temporal searchlight analyses revealed that the best classification performance was achieved by linear SVM (0.90 accuracy) which was fed by theta response (4 – 8 HZ) in the early time window (0.1 – 0.5 s). Delta responses also have distinctive characteristics for standard and novel stimuli. However, LDA (0.87 accuracy) and polynomial SVM (0.86 accuracy) were more powerful than the rest of the classifiers with delta response features. On the other hand, in the case to classify both rare stimuli, target, and novel, the best prominent feature was late theta (5.5-8 Hz), where SVM (both kernels) was used (0.82 accuracy). In general, the results revealed that novel sounds were distinguished from standard tones better than from target tones since whereas both novel and target tones were rare sounds which were the attended and/or oriented noises, standard tones were frequent sounds.

The findings in this study were in line with the previous studies that show the relation between delta (Güntekin and Başar, 2016) and theta (Karakaş, 2020) oscillatory responses and attentional processes. According to the literature, for stronger orientation coding, the main effects (peak latency) of the stimulus were found mostly at 300–500 ms post-stimulus onset for delta oscillations; however, general stimulus effects were found at time localization from 190 to 960 ms. For theta oscillations, whereas the main effects (peak latency) of the stimulus were found at 320–400 ms post-stimulus onset, a significant difference between target and non-target processing was obtained within 60–700 ms (Demiralp et al., 1999). In another study by Başar-Eroğlu et al. (1992), in responses to 3rd attended tones, there was a significant increase in the theta frequency band (frontal and parietal locations; 0-250 ms).

There were also common points with the studies that applied the ML approach. Aliakbaryhosseinabadi et al. (2019) ran the LDA classifier with three types of feature sets (time domain, frequency domain, and the combination of these two sets) obtained by the application of three modalities (auditory, visual, and audiovisual). However, the study focused on young adults rather than the elderly group. According to the results of this study, spectro-temporal features (combined feature set) had a higher accuracy than the other two feature sets for all modalities. Moreover, spectro-temporal features obtained upon application of auditory stimuli enable the classifier to attain remarkable performances for all brain regions, separately, in contrast to the current study which considered all the brain regions in one feature set. However, brain oscillations are selectively distributed in the whole brain (Başar, 2006). Therefore, feeding the classifier with a single feature set that includes the features obtained from all brain regions could provide more robust results.

Another study that attempted to classify standard and novel sounds by means of spatio-temporal patterns of discriminant electrophysiological responses to auditory stimuli was done by Aellen, Göktepe-Kavis, Apostolopoulos, & Tzovara (2021). In the study, the performances of deep learning (convolutional neural networks (CNN) with different techniques) and 'traditional' machine learning algorithms (Logistic Regression and SVM with 'rbf' kernel) were compared. The findings revealed that the AUC scores of the different CNN architectures (Shallow CNN: 0.75; Deep CNN: 0.73; ResNet: 0.72) were significantly higher than the AUC of logistic regression (0.63) and the AUC of SVM (0.58). Although the study applied deep-learning techniques, the performance of the current study where EROs in the time-frequency domain were used as features in the 'traditional' machine learning approach was much better than in this study where only time-domain features were used. Moreover, in the current study, the preprocessing step that makes the data more suitable for ML (Akhter et al., 2020) was more comprehensive than in the study by Aellen et al. (2021).

There is great importance in identifying rare stimuli, which have an impact on attentional processing in the environment (Liebherr et al., 2021), especially as adults grow older (Riis et al., 2008). The current study has revealed promising classification results to identify the novel sounds. Also, unlike many classification studies in the literature, exhaustive preprocessing steps were applied before feature extraction, including various artifact removal methods for robust classification results. However, the performances would be improved by overcoming several limitations of the study and using different feature sets. As a major problem, it is not possible to understand the aging effect in the identification of novel stimuli. To overcome this limitation, it absolutely would be better to include the young healthy group in the study, even dividing them into subgroups. This inclusion will also be helpful to decide the best model in cases of different ages.

Moreover, undoubtedly, the faster brain oscillations, like alpha, beta, and gamma, in response to sound stimuli are also related to cognitive processes (Mäkinen, May, & Tiitinen, 2004; Başar & Güntekin, 2012; Başar, 2013; Villena-González, Palacios-García, Rodríguez, & López, 2018). For example, beta oscillations in response to novel stimuli elicited an OR (Haenschel, Baldeweg, Croft, Whittington, & Gruzelier, 2000). Therefore, it would be informative to examine the distinctive features of high frequencies for the identification of sound stimuli in future studies. It should also be acknowledged that the current study addressed the sound stimuli classification with a small sample set. It obviously would be better to increase the number of subjects for more robust results and to use different learning techniques such as deep learning.

Peer-review: Externally peer-reviewed.
Conflict of Interest: The author has no conflict of interest to declare.
Grant Support: The author declared that this study has received no financial support.

REFERENCES

Aellen, F. M., Göktepe-Kavis, P., Apostolopoulos, S., & Tzovara, A. (2021). Convolutional neural networks for decoding electroencephalography responses and visualizing trial by trial changes in discriminant features. *Journal of Neuroscience Methods*, *364*, 109367. https://doi.org/10.1016/j.jneumeth.2021.109367

Akhter, R., Lawal, K., Tanvir, Md., & Ahmed, S. (2020). Classification of Common and Uncommon Tones by P300 Feature Extraction and Identification of Accurate P300 Wave by Machine Learning Algorithms. *International Journal of Advanced Computer Science and Applications*, *11*(10). https://doi.org/10.14569/ijacsa.2020.0111080

Alarcao, S. M., & Fonseca, M. J. (2019). Emotions Recognition Using EEG Signals: A Survey. *IEEE Transactions on Affective Computing*, *10*(3), 374–393. https://doi.org/10.1109/taffc.2017.2714671

Aliakbaryhosseinabadi, S., Kamavuako, E. N., Jiang, N., Farina, D., & Mrachacz-Kersting, N. (2019). Classification of Movement Preparation Between Attended and Distracted Self-Paced Motor Tasks. *IEEE Transactions on Biomedical Engineering*, *66*(11), 3060–3071. https://doi.org/10.1109/tbme.2019.2900206

Barry, R. J. (2009). Habituation of the orienting reflex and the development of Preliminary Process Theory. *Neurobiology of Learning and Memory*, *92*(2), 235–242. https://doi.org/10.1016/j.nlm.2008.07.007

Bradley, M.M., & Lang, P.J. (1999). International Affective Digitized Sounds (IADS-1): Stimuli, instruction manual, and affective ratings. Technical Report No B-2. University of Florida, Center for Research in Psychophysiology: Gainesville, FL, USA.

Başar, E. (2006). The theory of the whole-brain-work. *International Journal of Psychophysiology*, *60*(2), 133–138. https://doi.org/10.1016/j.ijpsycho.2005.12.007

Başar, E. (2013). A review of gamma oscillations in healthy subjects and in cognitive impairment. *International Journal of Psychophysiology*, *90*(2), 99–117. https://doi.org/10.1016/j.ijpsycho.2013.07.005

Başar, E., Başar-Eroglu, C., Karakaş, S., & Schürmann, M. (2001). Gamma, alpha, delta, and theta oscillations govern cognitive processes. *International Journal of Psychophysiology*, *39*(2-3), 241–248. https://doi.org/10.1016/s0167-8760(00)00145-8

Basar, E., Demiralp, T., Schürmann, M., Basar-Eroglu, C., & Ademoglu, A. (1999). Oscillatory Brain Dynamics, Wavelet Analysis, and Cognition. *Brain and Language*, *66*(1), 146–183. https://doi.org/10.1006/brln.1998.2029

Başar, E., & Güntekin, B. (2012). A short review of alpha activity in cognitive processes and in cognitive impairment. *International Journal of Psychophysiology*, *86*(1), 25–38. https://doi.org/10.1016/j.ijpsycho.2012.07.001

Başar-Eroğlu, C., Başar, E., Demiralp, T., & Schürmann, M. (1992). P300-response: possible psychophysiological correlates in delta and theta frequency channels. A review. *International Journal of Psychophysiology*, *13*(2), 161–179. https://doi.org/10.1016/0167-8760(92)90055-g

Başar-Eroğlu, C., & Demiralp, T. (2001). Event-related theta oscillations: an integrative and comparative approach in the human and animal brain. *International Journal of Psychophysiology*, *39*(2-3), 167–195. https://doi.org/10.1016/s0167-8760(00)00140-9

Behforuzi, H., Feng, N. C., Billig, A. R., Ryan, E., Tusch, E. S., Holcomb, P. J., … Daffner, K. R. (2019). Markers of Novelty Processing in Older Adults Are Stable and Reliable. *Frontiers in Aging Neuroscience*, *11*. https://doi.org/10.3389/fnagi.2019.00165

Berti, S., Vossel, G., & Gamer, M. (2017). The Orienting Response in Healthy Aging: Novelty P3 Indicates No General Decline but Reduced Efficacy for Fast Stimulation Rates. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01780

Borra, D., & Magosso, E. (2021). Deep learning-based EEG analysis: investigating P3 ERP components. *Journal of Integrative Neuroscience*, *20*(4), 791–811. https://doi.org/10.31083/j.jin2004083

Cavanagh, J.F. (2021). EEG: 3-Stim Auditory Oddball and Rest in Parkinson's. OpenNeuro; [Dataset] doi: 10.18112/openneuro.ds003490.v1.1.0

Cavanagh, J. F., Kumar, P., Mueller, A. A., Richardson, S. P., & Mueen, A. (2018). Diminished EEG habituation to novel events effectively classifies Parkinson's patients. *Clinical Neurophysiology*, *129*(2), 409–418. https://doi.org/10.1016/j.clinph.2017.11.023

Chung, J., & Teo, J. (2022). Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges. *Applied Computational Intelligence and Soft Computing*, *2022*, e9970363. https://doi.org/10.1155/2022/9970363

Debener, S., Makeig, S., Delorme, A., & Engel, A. K. (2005). What is novel in the novelty oddball paradigm? Functional significance of the novelty P3 event-related potential as revealed by independent component analysis. *Cognitive Brain Research*, *22*(3), 309–321. https://doi.org/10.1016/j.cogbrainres.2004.09.006

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Demiralp, T., Yordanova, J., Kolev, V., Ademoğlu, A., Devrim, M., & Samar, V.J. (1999). Time–frequency analysis of single-sweep event-related potentials by means of fast wavelet transform. Brain Lang. 66, 129–145. https://doi:10.1006/brln.1998.2028

Emek-Savaş, D. D., Güntekin, B., Yener, G. G., & Başar, E. (2016). Decrease of delta oscillatory responses is associated with increased age in healthy elderly. *International Journal of Psychophysiology*, *103*, 103–109. https://doi.org/10.1016/j.ijpsycho.2015.02.006

Güntekin, B., & Başar, E. (2016). Review of evoked and event-related delta responses in the human brain. *International Journal of Psychophysiology*, *103*, 43–52. https://doi.org/10.1016/j.ijpsycho.2015.02.001

Haenschel, C., Baldeweg, T., Croft, R. J., Whittington, M., & Gruzelier, J. (2000). Gamma and beta frequency oscillations in response to novel auditory stimuli: A comparison of human electroencephalogram (EEG) data with *in vitro* models. *Proceedings of the National Academy of Sciences*, *97*(13),

7645–7650. https://doi.org/10.1073/pnas.120162397

Harmony, T. (2013). The functional significance of delta oscillations in cognitive processing. *Frontiers in Integrative Neuroscience*, *7*. https://doi.org/10.3389/fnint.2013.00083

Ho, M.-C., Huang, C.-F., Chou, C.-Y., Lin, Y.-T., Shih, C.-S., Wu, M.-T., … Liu, C.-J. (2012). Task-related brain oscillations in normal aging. *Health*, *04*(09), 762–768. https://doi.org/10.4236/health.2012.429118

Hosseini, M.-P., Hosseini, A., & Ahi, K. (2021). A Review on Machine Learning for EEG Signal Processing in Bioengineering. *IEEE Reviews in Biomedical Engineering*, *14*, 204–218. https://doi.org/10.1109/rbme.2020.2969915

Huizeling, E., Wang, H., Holland, C., & Kessler, K. (2021). Changes in theta and alpha oscillatory signatures of attentional control in older and middle age. *European Journal of Neuroscience*, *54*(1), 4314–4337. https://doi.org/10.1111/ejn.15259

Karakaş, S. (2020). A review of theta oscillation and its functional correlates. *International Journal of Psychophysiology*. https://doi.org/10.1016/j.ijpsycho.2020.04.008

Li, K. Z. H., & Lindenberger, U. (2002). Relations between aging sensory/sensorimotor and cognitive functions. *Neuroscience & Biobehavioral Reviews*, *26*(7), 777–783. https://doi.org/10.1016/s0149-7634(02)00073-8

Liebherr, M., Corcoran, A. W., Alday, P. M., Coussens, S., Bellan, V., Howlett, C. A., … Bornkessel-Schlesewsky, I. (2021). EEG and behavioral correlates of attentional processing while walking and navigating naturalistic environments. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-01772-8

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., & Yger, F. (2018). A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *Journal of Neural Engineering*, *15*(3), 031005. https://doi.org/10.1088/1741-2552/aab2f2

Lynn, R. (1966). *Attention, Arousal, and the Orientation Reaction*. Oxford: Pergamon Press.

Makeig, S. (1993). Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalography and Clinical Neurophysiology*, *86*(4), 283–293. https://doi.org/10.1016/0013-4694(93)90110-h

Mäkinen, V. T., May, P. J. C., & Tiitinen, H. (2004). Human auditory event-related processes in the time-frequency plane. *NeuroReport*, *15*(11), 1767–1771. https://doi.org/10.1097/01.wnr.0000134841.48507.5a

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

Parvar, H., Sculthorpe-Petley, L., Satel, J., Boshra, R., D'Arcy, R. C. N., & Trappenberg, T. P. (2014). Detection of event-related potentials in individual subjects using support vector machines. *Brain Informatics*, *2*(1), 1–12. https://doi.org/10.1007/s40708-014-0006-7

Rahman, Md. M., Sarkar, A. K., Hossain, Md. A., Hossain, Md. S., Islam, Md. R., Hossain, Md. B., … Moni, M. A. (2021). Recognition of human emotions using EEG signals: A review. *Computers in Biology and Medicine*, *136*, 104696. https://doi.org/10.1016/j.compbiomed.2021.104696

Riis, J. L., Chong, H., Ryan, K. K., Wolk, D. A., Rentz, D. M., Holcomb, P. J., & Daffner, K. R. (2008). Compensatory neural activity distinguishes different patterns of normal cognitive aging. *NeuroImage*, *39*(1), 441–454. https://doi.org/10.1016/j.neuroimage.2007.08.034

Saeidi, M., Karwowski, W., Farahani, F. V., Fiok, K., Taiar, R., Hancock, P. A., & Al-Juaid, A. (2021). Neural Decoding of EEG Signals with Machine Learning: A Systematic Review. *Brain Sciences*, *11*(11), 1525. https://doi.org/10.3390/brainsci11111525

Schmiedt-Fehr, C., & Basar-Eroglu, C. (2011). Event-related delta and theta brain oscillations reflect age-related changes in both a general and a specific neuronal inhibitory mechanism. *Clinical Neurophysiology*, *122*(6), 1156–1167. https://doi.org/10.1016/j.clinph.2010.10.045

Schmiedt-Fehr, C., Dühl, S., & Basar-Eroglu, C. (2011). Age-related increases in within-person variability: Delta and theta oscillations indicate that the elderly are not always old. *Neuroscience Letters*, *495*(2), 159–163. https://doi.org/10.1016/j.neulet.2011.03.062

Tjandrasa, H., & Djanali, S. (2018). Classification of P300 event-related potentials using wavelet transform, MLP, and soft margin SVM. *2018 Tenth International Conference on Advanced Computational Intelligence (ICACI)*. https://doi.org/10.1109/icaci.2018.8377481

Treder, M.S. (2020). MVPA-Light: a classification and regression toolbox for multi- dimensional data. *Frontiers in Neuroscience 2020*; 14: 289. https://doi.org/10.3389/ fnins.2020.00289.

Villena-González, M., Palacios-García, I., Rodríguez, E., & López, V. (2018). Beta Oscillations Distinguish Between Two Forms of Mental Imagery While Gamma and Theta Activity Reflects Auditory Attention. *Frontiers in Human Neuroscience*, *12*. https://doi.org/10.3389/fnhum.2018.00389

Wang, J., & Wang, M. (2021). Review of the emotional feature extraction and classification using EEG signals. *Cognitive Robotics*, *1*, 29–40. https://doi.org/10.1016/j.cogr.2021.04.001

Wei, J., Zhao, L., Yan, G., Duan, R., & Li, D. (1998). The temporal and spatial features of event-related EEG spectral changes in 4 mental conditions. *Electroencephalography and Clinical Neurophysiology*, *106*(5), 416–423. https://doi.org/10.1016/s0013-4694(97)00161-2

ACTA INFOLOGICA 2023;7(1):81-93
dergipark.org.tr/acin

DOI: 10.26650/acin.1096774

RESEARCH ARTICLE

Application of Integer Response Optimization Models for the Healthcare Resources

Sağlık Kaynakları için Tam Sayılı Yanıt Optimizasyon Modellerinin Uygulanması

Abdulkadir Atalan[1]

[1](Assist. Prof.) Gaziantep Islam Science and Technology University, Faculty of Engineering and Natural Sciences, Department of Industrial Engineering, Gaziantep, Turkiye

ORCID: A.A. 0000-0003-0924-3685

Corresponding author:
Abdulkadir ATALAN
Gaziantep Islam Science and Technology University, Faculty of Engineering and Natural Sciences, Department of Industrial Engineering, Gaziantep, Turkiye
E-mail address: abdulkadir.atalan@gibtu.edu.tr

Submitted: 31.03.2022
Revision Requested: 12.12.2022
Last Revision Received: 15.03.2023
Accepted: 14.03.2023
Published Online: 07.04.2023

Citation: Atalan, A. (2023). Application of integer response optimization models for the healthcare resources. *Acta Infologica, 7*(1), 81-93. https://doi.org/10.26650/acin.1096774

ABSTRACT

This study investigated the relationship between the number of hospitals and healthcare professionals in the Turkish healthcare system. The study data covers the years 1967-2018. In the study, statistical analysis was performed, and the integer response optimization models were created using the linear regression response optimization models. The Pearson correlation test was applied to show the relationship between decision variables. The percentage of variation of the response variable specified by the linear regression model was calculated as 97.85. The numbers of nurses (p value=0.001, f value=21.36, t value=-4.62), midwives (p value=0.001, f value=24.88, t value=4.99), and OHP (Other health professionals) (p value=0.001, f value=24.16, t value=4.92) were found to be effective on the number of hospitals, doctors and pharmacists were not effective on the number of hospitals with a margin of error of 1.00% or 5.00%. The optimum number of healthcare workers required to provide health services in 1798 hospitals was calculated as 153128 doctors, 165244 nurses, 56351 midwives, 32032 pharmacists, and 177409 OHP. As a result of this study, we confirmed that the Turkish government should be able to provide health services in more hospitals with the existing healthcare system resources.

Keywords: Turkish healthcare system, healthcare professionals, hospitals, integer response optimization models

ÖZ

Bu çalışma, Türk sağlık sistemindeki hastane sayısı ile sağlık profesyonelleri arasındaki ilişkiyi araştırmak amacıyla yapılmıştır. Çalışma verileri 1967-2018 yıllarını kapsamaktadır. Çalışmada istatistiksel analizler yapılmış ve lineer regresyon yanıt optimizasyon modelleri kullanılarak tamsayılı yanıt optimizasyon modelleri oluşturulmuştur. Karar değişkenleri arasındaki ilişkiyi göstermek için Pearson korelasyon testi uygulanmıştır. Doğrusal regresyon modeli ile belirlenen yanıt değişkeninin varyasyon yüzdesi 97,85 olarak hesaplanmıştır. Hemşire (p değeri=0,001, f değeri=21,36, t değeri=-4,62), ebe (p değeri=0,001, f değeri=24,88, t değeri=4,99) ve diğer sağlık çalışanları (DSÇ) (p değeri=0,001, f değeri = 24.16, t değeri=4.92) hastane sayısı üzerinde etkili bulunmuş, doktor ve eczacı hastane sayısı üzerinde %1.00 veya %5.00 hata payı ile etkili olmamıştır. 1798 hastanede sağlık hizmeti vermek için gereken optimum sağlık çalışanı sayısı 153128 doktor, 165244 hemşire, 56351 ebe, 32032 eczacı ve 177409 DSÇ olarak hesaplanmıştır. Bu çalışmanın sonucunda, Türk hükümetinin sağlık sisteminin mevcut kaynakları ile daha fazla hastanede sağlık hizmeti sunabilmesi gerektiği doğrulanmıştır.

Anahtar Kelimeler: Türk sağlık sistemi, sağlık profesyonelleri, hastaneler, tamsayılı yanıt optimizasyon modelleri

1. INTRODUCTION

The healthcare systems of countries consist of two cornerstones: the human factor and physical structures. The main point is the human factor rather than the physical structures when health systems components are examined in detail (Leonard 2004; Taylor 1976). The human factor, health workers such as doctors, nurses, technicians, assistants, etc., is at the center of health services (Khatri et al. 2006). Healthcare system sections are divided into employees and organizations (Atalan and Donmez 2020; Mihaylova et al. 2011). While employees create the human factor, doctors, midwives, nurses, technicians, etc., structures such as hospitals, clinics, etc., are classified as healthcare institutions. Some researchers have handled these classifications as healthcare components and health resources (Siciliani, Stanciole, and Jacobs 2009). For example, while hospitals, private health insurance companies, patients, and rule-making states are considered health components, doctors, nurses, number of beds, etc., are defined as healthcare resources (Lovink et al. 2017; Shinjo and Aramaki 2012).

The components or resources that researchers define make up health systems as variables, factors, elements, etc. (Maleki, Majlesinasab, and Sepehri 2014; Mihaylova et al. 2011). These dependent and independent variables affect the dependent variable (Ahmed and Alkhamis 2009; Thapa et al. 2018). In this study, five independent variables and one dependent variable were considered. While the number of hospitals was defined as the dependent variable, four independent variables were determined as the number of doctors, nurses, midwives, pharmacists, and other healthcare workers. Researchers have used many methods to reveal the link between dependent and independent variables (Atalan 2021; ATALAN 2020). Since healthcare systems have complex and dynamic structures, a single process is insufficient to solve the problem (Ramudhin, Chan, and Mokadem 2006). For this reason, many methods specific to healthcare problems are used at the same time. Statistical analysis approaches are used at the beginning of an examination of healthcare problems (Dönmez, Atalan, and Dönmez 2020; Malehi, Pourmotahari, and Angali 2015; Pellegrini, Rodriguez-Monguio, and Qian 2014). Also, the correlation values between dependent and independent variables defined in this study were computed using the linear regression analysis method. In addition to the correlation values of the health resources considered for health management, the correlation values were calculated indicating the types of diseases seen which are useful in health resource planning (Cheng et al. 2022). However, in this study, data on disease types were not considered in health resource management.

This study examined the health resources affecting Turkey's healthcare economy using linear regression statistical analysis. Using the data for the years 1990-2016, per capita income, and population showed that the number of doctors was statistically significant, while the number of nurses and beds was insignificant (Atalan 2018). This study did not use the healthcare economy and population of Turkey data. Kan et al. estimated future healthcare costs for the elderly and compared the results obtained from using two methods, standard and linear regression (Kan et al. 2019). Data from 2,911 patients were analyzed using 'chi-square tests, one-way ANOVA, multiple linear regression, and binominal logistic regression to determine statistically whether outpatients received medical attention from the same physicians (Lee, Chang, and Du 2017). Another study used the linear regression method to examine patient safety in a regional hospital, considering doctor and nurse variables. As a result, the effect of doctors and nurses on hospital safety culture was statistically emphasized in this study (Chi et al. 2017). Comparisons were made between these three occupations by measuring job satisfaction among doctors, nurses, and assistants with a linear regression model (Krogstad et al. 2006). Krogstad has preferred predictive logistic regression models to examine the perception of cooperation between 551 doctors and 2,050 nurses (Krogstad 2004). Researchers used a multiple logistic regression method to examine doctors' attitudes toward new nursing practices by expanding nurses' job descriptions (Brodsky and Van Dijk 2008). Regression analysis and one-sample t-test were used to characterize nurses' organizational commitment and job qualifications according to different dimensions in another study (Gabrani et al. 2016). Ishikawa used logistic regression analysis to estimate relevant factors resulting from the geographic distribution of physicians according to health policies (Ishikawa 2020). Regression analysis models have been widely used to solve many problems in the field of healthcare (Atalan 2014). The mathematical optimization models were developed as a second method to calculate this study's optimum healthcare component and resources.

Researchers in studies generally use optimization models in the field of healthcare to determine the number of healthcare resources, organize the shift system of healthcare resources, and eliminate unnecessary steps in the patient flow diagram

(Abdalkareem et al. 2021; Daldoul et al. 2018; Goienetxea Uriarte et al. 2017; Lin 2008). The optimization models contributed to an increase in the number of patients treated, a decrease in patient waiting times, and an increase in the efficiency of hospital resources (Atalan and Dönmez 2020; Daldoul et al. 2018; Schmid 2012). Optimization models such as linear, nonlinear, integer, dynamic, and static differ according to the problem addressed in healthcare (Agarana and Olokunde 2015; Batun and Begen 2013; Cabrera et al. 2012; Özcan and Tüysüz 2018; Steiner et al. 2015). This study used a linear optimization model since there was no interaction between independent variables.

As a result of optimizing the number of healthcare resources employed in emergency services using the discrete event simulation model with linear optimization, the patient waiting time was reduced, and the number of patients treated increased (Ahmed and Alkhamis 2009; Atalan 2022). In another study, optimum healthcare resource numbers were calculated to provide better healthcare by optimizing patient flow (Ansari et al. 2020). Ramudhin et al. developed a medBPM (medical Business Process Modeling) optimization model for the treatment process to examine patient flow movements, decision points, and waiting for queues in the healthcare system (Ramudhin et al. 2006). To efficiently plan the healthcare resources required for the operating rooms, a dynamically structured optimization model has been developed for healthcare institutions with limited healthcare resources, using the assignment problem method, one of the optimization types (Zhao and Wen 2021). It has been improved by multi-objective optimization models to regulate nurses' shift systems, which are essential in increasing healthcare service quality, reducing hospital budget, and improving staff work satisfaction (Chiang et al. 2019). The optimization models created for this study have two objective functions. The objective functions of the optimization models developed were not evaluated within the scope of the multi-objective optimization model, and the models were run separately.

This study consists of four parts. Literature information about the methods used for dependent and independent variables affecting healthcare systems are given in the first section. The data used in this study and the applied linear regression response optimization model method are provided in the second part. The statistical and optimization results obtained in the study are evaluated in the third part. The results and suggestions for future studies are stated in the conclusion.

2. DATA AND METHODOLOGY

The methodology part of this study consists of three parts. The study data and variables are given in the first part. Two methods were used in the study. The effect of independent variables on the dependent variable was statistically measured using linear regression analysis. The optimum results of dependent and independent variables were obtained by considering two different optimization models in the second method. Developed optimization models were created according to linear mathematical modeling since there is no interaction between the input variables.

2.1. Data and Variables

The study data belongs to the Turkish Healthcare System, and the data between 1967-2018 were obtained from the Turkish Statistical Institution (TUIK 2021). The data set was considered dependent (the number of hospitals) and independent (medical doctors, nurses, midwives, pharmacists, and other healthcare professionals), including six variables. Descriptive statistical data of these variables are shown in Table 1.

Table 1
Descriptive Statistics Data of Healthcare Resources

Variables Parameters	Medical Doctors	Nurses	Midwives	Pharmacists	Other Healthcare Professionals[1]	Hospitals
Types	independent	independent	independent	independent	independent	dependent
Notation	X_d	X_n	X_m	X_p	X_o	Y_{it}
Sample Size	52.00000	52.00000	52.0000	52.000	52.0000	52.00
Mean	67418.00	62160.00	31995.0	6686.0	48178.0	1024.9
Mean of Std. Dev.	6118.000	6584.000	2161.00	1150.00	6514.0	40.90
Std. Dev.	44114.00	47474.00	15587.0	8295.0	46974.0	294.70
Variance	19460413	22538122	242945	68814	220659	86866.1
Variance Coeff.	65.43000	76.37000	48.7200	49.720	97.500	28.760

Minimum	11875.00	6161.000	5621.00	2203	5997.00	646.00
Q_1	26534.00	24568.00	16949.0	11369	11846.0	776.30
Median	59018.00	52362.00	35680.0	17145	26468.0	945.00
Q_3	103570.0	81515.00	44248.0	23044	66144.0	1213.50
Maximum	153128.0	190499.00	56351.0	32032	177409.0	1534.0
Skewness	0.45.000	0.920000	-0.0600	-0.220	1.260000	0.4800
Kurtosis	-1.14.00	0.060000	-1.4900	-1.050	0.47000	-1.200

[l]Technicians, Biologists, ‘ Physiotherapists, Technicians, Technologists, Secretaries, etc. For detailed information: (TUIK 2021); Abbreviations: Coeff., coefficient; Std. Dev., standard deviation; Q_1, the first quartile of data; Q_3, the third quartile of data

The number of physicians among the independent variables is one of the cornerstones of healthcare systems. The number of doctors increased over the years, with a maximum of 153,128 doctors employed. The lowest level was in 1967 when 11,875 doctors were employed. An average of 64,417 physicians was employed during the years for which the data were taken into account. While a maximum of 190,499 nurses were employed in 2018, only 6,161 nurses were employed in 1967. The average number of nurses between these years was 62,160. Five thousand six hundred twenty-one midwives were employed in 1967, while a maximum of 56,351 midwives were employed in 2018. The average number of midwives between these years isapproximately 32,000.

Pharmacists play an essential role in the health system. The number of pharmacists employed in hospitals is deficient when compared with the number of pharmacists operating outside the hospital. This study considered the number of pharmacists working inside and outside the hospital. This study evaluated the number of pharmacists factor as an independent variable. While a maximum of 32,032 pharmacists worked in 2018, only 2,203 pharmacists worked in 1967. The average number of pharmacists between these years wasas 16,686. The average number of other health professionals (OHP) considered between these years was estimated as 48,178. While a maximum of 177,409 other health workers were employed in 2018, only 5,997 other health workers were employed in 1967. The average number of hospitals considered between these years was approximately 1,025. A maximum of 1,534 hospitals provided healthcare services in 2018, but only 646 hospitals provided services in 1981.

The number of hospitals fluctuated over the years. While an increase was observed from 1967 to 1980, the number of hospitals providing healthcare services decreased dramatically in 1981 and 1982. Although there was an increase in the number of hospitals from 1982 to 2001, there was a slight decrease until 2008. Although there was an increase in the number of hospitals providing health services from 2009 to 2014, there was a stagnation in the number of hospitals until 2018; even though there were fluctuations in the number of hospitals, the employment of healthcare personnel increased over the years. This study statistically analyzed whether healthcare personnel was affected by the fluctuations in the number of hospitals.

The correlation data of dependent and independent variables are given in Correlation values are measured according to the values between -1.00 and +1.00 between two variables. As the correlation values get closer to -1.00, they create a strong negative correlation, while as they get closer to +1.00, they create a strong positive correlation. This study has positive correlation values between all variables. Correlation values between the variables ranged between 0.858 and 0.983. According to this table, there is a strong correlation value between the number of other health workers and the number of nurses. The correlation values between the dependent and independent variables were calculated as 0.978 with the number of doctors, 0.956 with the number of nurses, 0947 with the number of midwives, 0.918 with the pharmacist, and 0.946 with the number of OHP.

Table 2
Correlation values between dependent and independent variables

Variables	**Medical Doctors**	**Nurses**	**Midwives**	**Pharmacists**	**OHP**
Hospitals	0.978	0.956	0.947	0.918	0.946
OHP	0.952	0.983	0.870	0.858	1.000
Pharmacists	0.961	0.929	0.981	1.000	
Midwives	0.972	0.931	1.000		
Nurses	0.979	1.000			

2.2. Linear Regression Model

There are two types of variables in regression analysis: dependent and independent. In short, regression analyzes are widely used to statistically detect the effects of independent variables on dependent variables. The symbol usually denotes the independent variables. The dependent variables affected by the independent variables are expressed with the symbol . A regression equation with one independent variable is formulated as follows (Cattaneo, Jansson, and Newey 2018):

$$y_{i,n} = \beta_i' x_{i,n} + \gamma_n' w_{i,n} + \mu_{i,n}, \qquad i = 1, \dots, n \tag{1}$$

where, $y_{i,n}$ represents the dependent variable, and $x_{i,n}$ denotes the independent variable. The $w_{i,n}$ symbolizes the vector of covariates with the coefficient of γ_n' in the regression model, and β_i' signifies the coefficient of influence of the independent variable. The $\mu_{i,n}$ indicates the number of errors between the observed value and the predicted value. Since then, $w_{i,n} = 0$, $w_{i,n}$, was excluded from the formulation in this study. The following formula expresses linear regression models with multidirectional fixed effects according to data distributions and in non-interactive independent variable environments.

$$Y_{it} = \alpha_i + \beta_i' X_{it} + U_{it}, \qquad i = 1, \dots, N, \;\; t = 1, \dots, T \tag{2}$$

where Y_{it} represents the dependent variables and suppose that $Y_{it} = y_{(i-1)T+t,n}$ for $1 \le i \le N$ and $1 \le t \le T$. The constant value of the regression model is $\alpha_i, \alpha_i \in \mathbb{R}$. X_{it} represents the independent regressors and suppose that $X_{it} = x_{(i-1)T+t,n}$ for $X_{it} = X_{i1}, \dots, X_{iT} : 1 \le i \le n$. $U_{it} = U_{i1}, \dots, U_{iT}$ are the values of errors equal to the i^{th} unit variables of dimension N and $U_{it} = U_{i1}, \dots, U_{iT}$, $X_{it} = X_{i1}, \dots, X_{iT} : 1 \le i \le n$. There are multiple independent variables in this study. A regression equation with a large number of independent variables is expressed (Montgomery, Peck, and Vining 2012):

$$Y_{it} = \alpha_i + \beta_1' X_{1t} + \beta_2' X_{2t} + \beta_3' X_{3t} + \cdots + \beta_i' X_{it} + U_{it}, \quad i = 1, \dots, N, \;\; t = 1, \dots, T \tag{3}$$

In this study, the regression equation contains five independent variables and one dependent variable, according to the notations of the variables shown in Table 1. The equation is expressed as $Y_{it} = \alpha_i + \beta_d' X_{dt} + \beta_n' X_{nt} + \beta_m' X_{mt} + \beta_p' X_{pt} + \beta_o' X_{ot} + U_{it}$ for $i = 1, \dots, N, \;\; t = 1, \dots, T$. The regression equation resulting from the absence of interaction between the independent variables has a linear equation structure. In other words, the regression equation obtained in this study is an example of a linear regression equation.

2.3. Integer Optimization Models

Integer optimization models are used if the decision variables are integers; and the input and output decision variables have indivisibility problems (Öztürk 2016). Integer optimization models are generally defined as linear optimization models that adopt the constraint of integer decision variables. Linear mathematical modeling is created as follows (Bradley, Hax, and Magnanti 1977):

$$Maximize \text{ or } Minimize \;\; \sum_{i=j}^{T} C_j x_j$$

Subject to

$$\sum_{j=1}^{T} A_{ij} x_j \le K_i \;\; for \; i = 1, 2, 3, \dots, N_1 \tag{4}$$

$$\sum_{j=1}^{T} A_{ij} x_j = K_i \;\; for \; i = N_1 + 1 \dots, N_2$$

$$\sum_{i=j}^{T} A_{ij} x_j \ge K_i \;\; for \; i = N_2 + 1 \dots, N_3$$

$$X_j \ge 0, and \; j = 1, 2, 3, \dots, T$$

where x_j represents the decision variable. A_{ij} and C_j are decision variable coefficients and constants. The value of K_i represents the right-hand-side value of the optimization model's constraints. The integer mathematical model is formulated as follows (Bradley et al. 1977):

$$Maximize\ or\ Minimize\ \sum_{i=j}^{T} C_j x_j$$

Subject to

$$\sum_{j=1}^{T} A_{ij} x_j \leq K_i \ \ for\ i = 1, 2, 3, \dots, N_1$$

$$\sum_{j=1}^{T} A_{ij} x_j = K_i \ \ for\ i = N_1 + 1 \dots, N_2$$

$$\sum_{i=j}^{T} A_{ij} x_j \geq K_i \ \ for\ i = N_2 + 1 \dots, N_3 \tag{5}$$

$$X_j \geq 0, and \ \ integer, for\ j = 1, 2, 3, \dots, T$$

If the decision variables are binary, the decision variable limits are expressed as $X_j = 0$ or $X_j = 1$. In such a case, the model constraint is described as follows:

$$\sum_{i=j}^{T} A_{ij} x_j = 1 \tag{6}$$

Optimization models created with such definitions are defined as integer assignment models (Winston and Venkataramanan 2002).

The second method of this study is the creation of integer optimization models. Optimization models are formed by determining the decision variables, objective function, constraints and limits, and the signs or types of the decision variables. While the decision variables of this study are independent factors, the objective function consists of the regression equation. Two types of optimization models were created for this study. The integer optimization method was used since the values of the decision variables in both optimization models are integers. In the first optimization model, the objective function equals an average number. The maximum value of the constraints is defined as the average value they have for 52 years. This optimization model is built as follows:

$$Y_{it}\{\alpha_i + \beta_d' X_{dt} + \beta_n' X_{nt} + \beta_m' X_{mt} + \beta_p' X_{pt} + \beta_o' X_{ot} + U_{it}\} = \overline{Y}_{it}$$

Subject to the constraints:

$$l_d^t \leq X_{dt} \leq \overline{u}_d^t$$

$$l_n^t \leq X_{nt} \leq \overline{u}_n^t$$

$$l_m^t \leq X_{mt} \leq \overline{u}_m^t \tag{7}$$

$$l_p^t \leq X_{pt} \leq \overline{u}_p^t$$

$$l_o^t \leq X_{ot} \leq \overline{u}_o^t$$

$$t = \{1967, 1968, \dots, 2018\}$$

$$X_{dt}, X_{nt}, X_{mt}, X_{pt}, X_{ot} \geq 0, and\ integer$$

where, y_{hs} represents the mean of the values of the objective function over 52 years. $l_d^t, l_n^t, l_m^t, l_p^t, l_o^t;\ t = \{1967, 1968, \dots, 2018\}$ are the minimum values of the variables. $\bar{u}_d^t, \bar{u}_n^t, \bar{u}_m^t, \bar{u}_p^t, \bar{u}_o^t;\ t = \{1967, 1968, \dots, 2018\}$ data represents the average of the values of the independent variables over 52 years. The second optimization model developed for this study, the objective function, was created to calculate the optimum values the independent variables can have when the number of hospitals at the current maximum level is obtained. The second optimization model is expressed as mathematical equations as follows:

$$Maximize\ Y_{it} = \left\{\alpha_i + \beta_d' X_{dt} + \beta_n' X_{nt} + \beta_m' X_{mt} + \beta_p' X_{pt} + \beta_o' X_{ot} + U_{it}\right\},$$

Subject to the constraints:

$$\begin{gathered} l_d^t \le X_{dt} \le \bar{u}_d^t \\ l_n^t \le X_{nt} \le \bar{u}_n^t \\ l_m^t \le X_{mt} \le \bar{u}_m^t \\ l_p^t \le X_{pt} \le \bar{u}_p^t \\ l_o^t \le X_{ot} \le \bar{u}_o^t \\ t = \{1967, 1968, \dots, 2018\} \\ X_{dt}, X_{nt}, X_{mt}, X_{pt}, X_{ot} \ge 0, and\ integer \end{gathered} \tag{8}$$

where, $u_d^t, u_n^t, u_m^t, u_p^t, u_o^t;\ t = \{1967, 1968, \dots, 2018\}$ are the maximum values of the independent variables. Minitab 18.01 computer statistics program was used both for statistical analysis and for the solution of optimization models. The findings obtained from the study methods are included in the third part.

3. FINDINGS OF THE STUDY

A two-stage method was used for this study. Regression analysis, the first method, was performed and tested whether the independent factors affected the dependent factor. In addition, the accuracy percentages of the statistical analysis (variation percentages of the response variable specified by the linear regression model) were calculated as R^2 = 97.85%, $R^2_{adjusted}$ =97.62%, and $R^2_{predicted}$ =97.62%, respectively. While predicted R^2 is used in statistical regression models due to the complexity of the models with more than one variable, adjusted R^2 is used to compare the reliability power between regression models containing a different number of variables and univariate models. The R^2 values obtained in this study are high, and the reliability of the obtained results has been tested. Statistical effects of dependent and independent variables are given in

Table 3
The results of statistical analysis

Variables	Coefficient of Regression	Sum of Squares	Mean of Squares	t-value	f-value	p-value
X	0.00029	48,000	48,000	0.15	0.020	0.880
X_n	-0.00970	44245	44245	-4.62	21.36	0.001
X_m	0.01681	51518	51518	4.99	24.88	0.001
X_p	0.00315	572,00	572,00	0.53	0.280	0.602
X_o	0.00999	50034	50034	4.92	24.16	0.001

According to , the numbers of the nurse (p value=0.001, f value=21.36, t value=-4.62), midwives (p value=0.001, f value=24.88, t value=4.99), and OHP (p value=0.001, f value=24.16, t value=4.92) were found to be effective on the number of hospitals. Still, the number of doctors and pharmacists was inadequate for the number of hospitals, with a margin of error of 1.00% or 5.00%. However, if the number of doctors and pharmacists is evaluated within the margin of error of 10%, we can assume that they affected the outcome. While the increase in the number of health institutions also causes an increase in some healthcare resources, some of them should not be employed too much. By obtaining the regression coefficients of the dependent and independent variables, the regression equation has emerged as follows:

$$Y_{it} = 536{,}6 + 0{,}00029X_{dt} - 0{,}00970X_{nt} + 0{,}01681X_{mt} + 0{,}00315X_{pt} + 0{,}00999X_{ot} + U_{it} \tag{9}$$

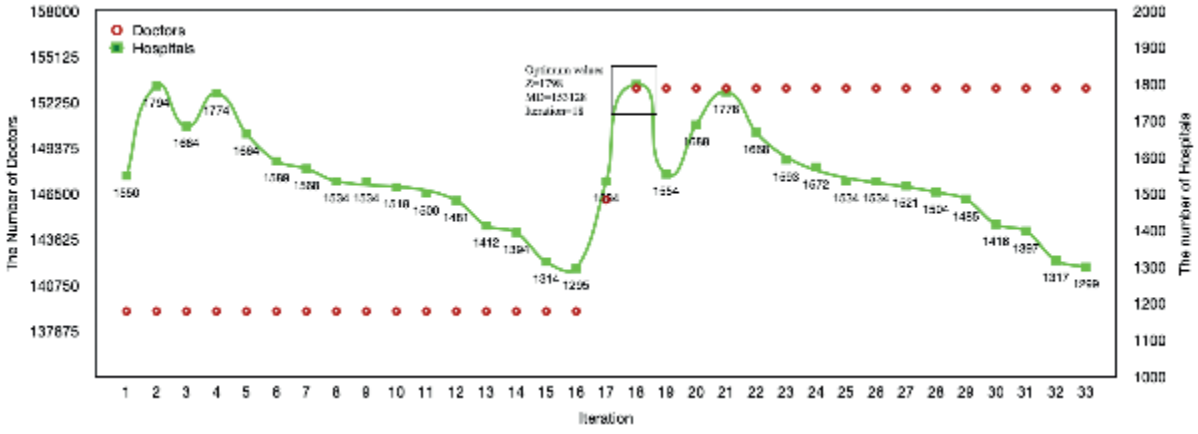
According to this equation, of the five independent variables, only the number of nurses, had a negative effect on the number of hospitals. It is understood that other independent variables positively impact the number of hospitals. This regression equation is also considered the objective function of the optimization models developed. The optimum values obtained by running the first and second optimization models are shown in Table 4.

Table 4
The optimum results of dependent and independent variables

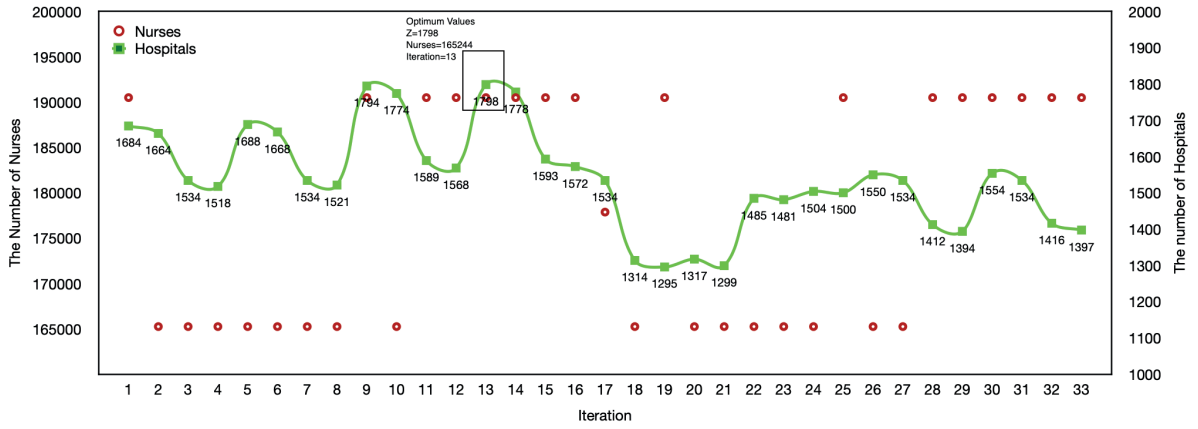
Optimization Models	Objective functions	Decision Variables				
		X_d	X_n	X_m	X_p	X_o
Model 1,	1263,99 ~ 1264	60999,29 ~ 60999	16545,90 ~ 16546	15145,25 ~ 15145	12887.26 ~ 12887	57581.173 ~ 57581
Model 2,	1458,09 ~ 1458	139105,617 ~ 139106	165244,26 ~ 165244	48796,89 ~ 48797	25500,35 ~ 25500	156824,39 ~ 156824

Table 4 was converted to integers when healthcare resources could not be expressed as fractions. According to optimization model 1, the objective function was calculated as 1,264. The optimum health resource numbers to be employed in 1,264 hospitals were calculated as 60,999 doctors, 16,546 nurses, 15,145 midwives, 12,887 pharmacists, and 57,581 OHP. According to the second optimization model, the optimum number of hospitals was 1,458. According to the optimum number of hospitals, 13,9106 doctors, 16,5244 nurses, 48,797 midwives, 25,500 pharmacists, and 156,824 OHP should be employed. The closest value to the number of 1,458 hospitals is the data for 2011. While the number of hospitals operating in 2011 is 1,453, the number of doctors employed is 126,029, the number of nurses is 124,982, the number of midwives is 51,905, the number of pharmacists is 26,089, and the number of OHP is 110,862. As a result of the optimum values obtained, we determined that the employment of doctors, nurses, and OHP is low, while the employment of midwives and pharmacists is high compared to 2011 data.

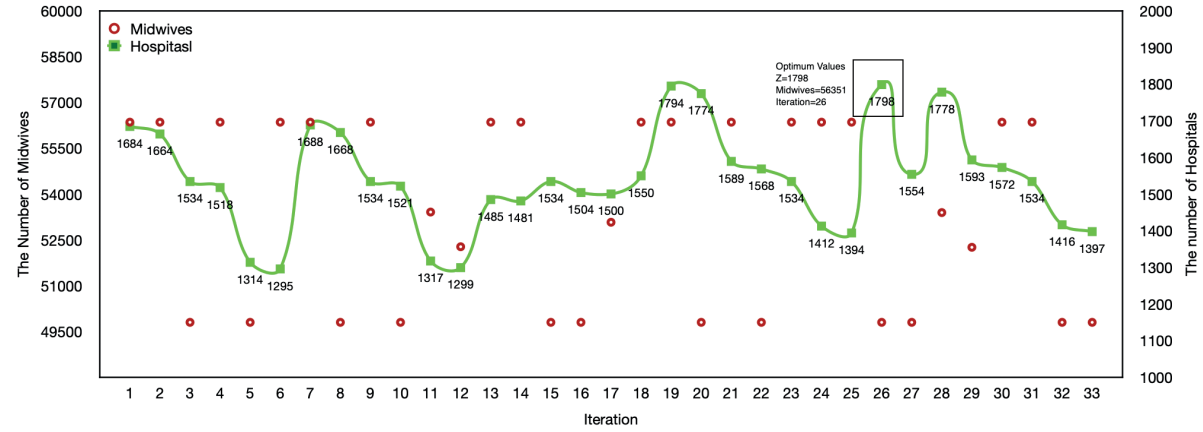
Calculating how many hospitals should provide healthcare services according to the maximum and minimum values of the existing independent variables is shown in . According to these figures, the maximum number of hospitals was calculated as 1,798 without limiting the objective function. While maximizing the number of hospitals, the limits of the decision variables did not change. In other words, how many hospitals will operate with the existing health workers (the numbers may vary within limits) without employing extra health workers has been optimized. The optimum number of health workers required to provide health services in 1,798 hospitals was calculated as 153,128 doctors, 165,244 nurses, 56,351 midwives, 32,032 pharmacists, and 177,409 OHP. They obtain optimum results by arranging the numbers of each healthcare worker from the lowest to the highest, corresponding to different iterations. According to these results, 25,255 nurses are overworked in the Turkish Healthcare System.



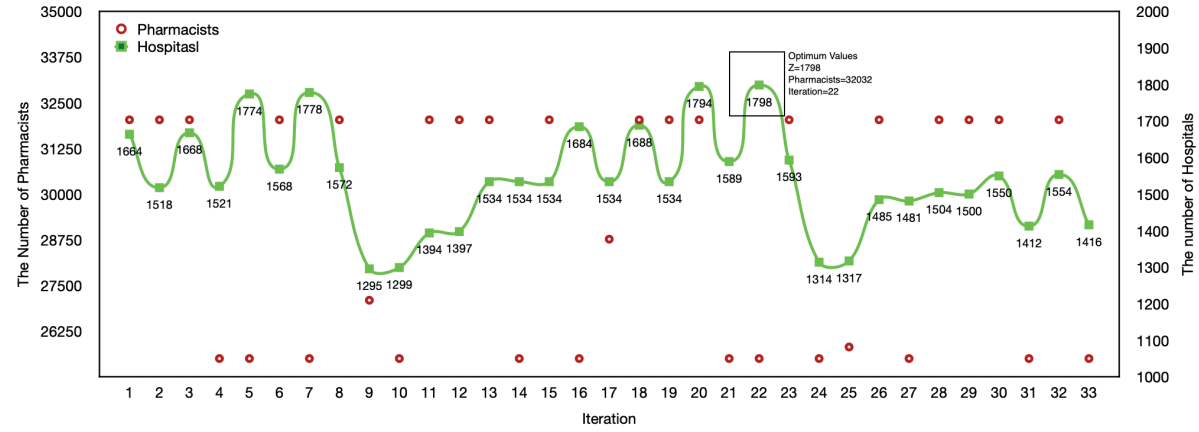
(a)


Nurses
Hospitals
Optimum Values
Z=1798
Nurses=165244
Iteration=13
The Number of Nurses
The number of Hospitals
Iteration

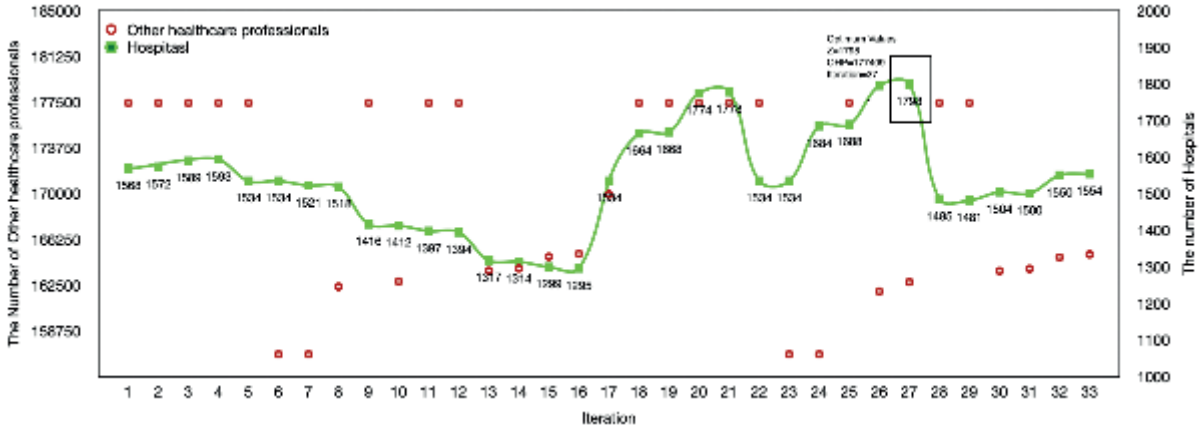
(b)


Midwives
Hospitasl
Optimum Values
Z=1798
Midwives=56351
Iteration=26
The Number of Midwives
The number of Hospitals
Iteration

(c)


Pharmacists
Hospitasl
Optimum Values
Z=1798
Pharmacists=32032
Iteration=22
The Number of Pharmacists
The number of Hospitals
Iteration

(d)



(e)

Figure 7. Number of Healthcare Professionals for the Maximum Number of Hospitals (a-the number of doctors, b-the number of nurses, d-the number of midwives, d-the number of pharmacists, e-the number of OHP) (Z: objective function)

The optimization model developed in this study will contribute to the determination of the optimum number of healthcare resources needed according to the number of hospitals that operating in the future. Thus, efficient and sufficient use of health resources will be ensured in the healthcare systems of countries. Efficient use of healthcare resources is one of the critical issues in health system management. While health resource planning is done in health management, there are more effective ways to use available resources. The employment of health personnel in existing health institutions does not disrupt the service or recruitment of personnel for new health institutions. In new health institutions, there are two situations that can affect the employment of health personnel and their utilization.. In the first situation, the institution must determine how best to utilize existing health personnel, but the main issue of health management is to deal with the second case, which causes problems in terms of costs. In other words, it is preferred in health management to keep existing resources and expense costs constant and to get more efficiency (Ceylan and Atalan 2021). Although this situation may seem appropriate for small-scale communities or health institutions, if it is examined in terms of patient population and number of institutions, the lack of many health resources can cause disadvantages (for example, long waiting times, staff and patient dissatisfaction, a small number of patients being examined or treated, etc.). Balancing the two situations should be decided by looking at the changes in healthcare systems.

4. DISCUSSION AND CONCLUSION

The integer optimization models were created using the response optimization method to facilitate healthcare resource planning by using the data on the number of health resources from 1967 to 2018 in Turkey in this study. The number of hospitals was considered the dependent variable, while the number of doctors, nurses, midwives, pharmacists, and OHP were defined as the independent variable. Linear regression and integer optimization methods were used in this study. In regression analysis, the number of nurses, midwives, and OHP was effective on the number of hospitals, but the number of doctors and pharmacists was only effective on the dependent variable, according to a 10% margin of error. Two optimization models were developed by determining the limits of the decision variables with the obtained regression equation. Optimum results were obtained for the number of healthcare resources employed by taking the average number of hospitals operating in the first model 52 years. In the second optimization model, the optimum values of the healthcare resources used were calculated by maximizing the number of hospitals. The number of doctors, nurses, and midwives from health resources is low, but the number of other health resources is high, according to the results obtained from the second optimization model.

The number of hospitals is increasing day by day to meet human needs. There may be many reasons for this increase, such as the population, the desire to live longer, etc. However, an increase in the number of healthcare resources is considered normal only after ensuring that efficiency from the existing healthcare resources is sufficient. Increasing the number of healthcare resources in a healthcare institution does not mean that that institution will obtain higher efficiency. Advantages such as the increase in the number of patients treated, reduced patient waiting time, and decreased patient examination/treatment costs are not directly proportional to increasing the number of healthcare resources. Many processes, such as patient flow diagrams, examination/treatment procedures, and communication between resources, must be examined well to reveal such benefits.

For future studies, the number of healthcare resources can be calculated using the data used in this study and the estimation methods such as machine learning and time series. The number and types of dependent and independent variables can be increased or decreased by considering the healthcare cost data. Another research topic is the optimum number of health resources calculated by calculating the labor costs of healthcare resources for a hospital. For this reason, this study will contribute significantly to the literature and be an essential source for future studies.

Peer-review: Externally peer-reviewed.
Conflict of Interest: The author has no conflict of interest to declare.
Grant Support: The author declared that this study has received no financial support.

REFERENCES

Abdalkareem, Zahraa A., Amiza Amir, Mohammed Azmi Al-Betar, Phaklen Ekhan, and Abdelaziz I. Hammouri. 2021. "Healthcare Scheduling in Optimization Context: A Review." *Health and Technology* 11(3):445–69. doi: 10.1007/s12553-021-00547-5.

Agarana, M. C., and T. O. Olokunde. 2015. "Optimization of Healthcare Pathways in Covenant University Health Centre Using Linear Programming Model." *Journal of Applied Mathematics* 91(3):215–28.

Ahmed, Mohamed A., and Talal M. Alkhamis. 2009. "Simulation Optimization for an Emergency Department Healthcare Unit in Kuwait." *European Journal of Operational Research* 198(3):936–42.

Ansari, Aseer Ahmad, Sabyasachi Chakraborty, Satyabrata Aich, Beom Su Kim, and Hee-Cheol Kim. 2020. "Optimization of Healthcare Network for Improved Service Delivery." Pp. 314–18 in *2020 22nd International Conference on Advanced Communication Technology (ICACT)*. IEEE.

Atalan, Abdulkadir. 2014. "Central Composite Design Optimization Using Computer Simulation Approach." *Flexsim Quarterly Publication* 5–19.

Atalan, Abdulkadir. 2018. "Türkiye Sağlık Ekonomisi Için İstatistiksel Çok Amaçlı Optimizasyon Modelinin Uygulanması." *İşletme Ekonomi ve Yönetim Araştırmaları Dergisi* 1(1):34–51.

Atalan, Abdulkadir. 2021. "EFFECT OF HEALTHCARE EXPENDITURE ON THE CORRELATION BETWEEN THE NUMBER OF NURSES AND DOCTORS EMPLOYED." *International Journal of Health Management and Tourism* 6(2):515–25. doi: 10.31201/ijhmt.949500.

Atalan, Abdulkadir. 2022. "A Cost Analysis with the Discrete-event Simulation Application in Nurse and Doctor Employment Management." *Journal of Nursing Management* 30(3):733–41. doi: 10.1111/jonm.13547.

ATALAN, Abdulkadir. 2020. "Forecasting for Healthcare Expenditure of Turkey Covering the Years of 2018-2050." *Gümüşhane Üniversitesi Sağlık Bilimleri Dergisi* 9(1):8–16. doi: 10.37989/gumussagbil.538111.

Atalan, Abdulkadir, and Cem Cagri Donmez. 2020. "DEVELOPING OPTIMIZATION MODELS TO EVLUATE HEALTHCARE SYSTEMS." *Sigma Journal of Engineering and Natural Sciences* 38(2):853–73.

Atalan, Abdulkadir, and Cem Cağrı Dönmez. 2020. "Optimizing Experimental Simulation Design for the Emergency Departments." *Brazilian Journal of Operations & Production Management* 17(4):1–13. doi: 10.14488/BJOPM.2020.026.

Batun, Sakine, and Mehmet A. Begen. 2013. "Handbook of Healthcare Operations Management. Optimization in Healthcare Delivery Modeling: Methods and Applications." Pp. 75–119 in Vol. 184. New York, NY: Springer.

Bradley, Stephen P., Arnoldo C. Hax, and Thomas L. Magnanti. 1977. *Applied Mathematical Programming*. Addison-Wesley.

Brodsky, Eithan, and Dina Van Dijk. 2008. "Advanced and Specialist Nursing Practice: Attitudes of Nurses and Physicians in Israel." *Journal of Nursing Scholarship* 40(2):187–94. doi: 10.1111/j.1547-5069.2008.00225.x.

Cabrera, Eduardo, Emilio Luque, Manel Taboada, Francisco Epelde, and Ma Luisa Iglesias. 2012. "ABMS Optimization for Emergency Departments." in *Proceedings - Winter Simulation Conference*.

Cattaneo, Matias D., Michael Jansson, and Whitney K. Newey. 2018. "Inference in Linear Regression Models with Many Covariates and Heteroscedasticity." *Journal of the American Statistical Association* 113(523):1350–61. doi: 10.1080/01621459.2017.1328360.

Ceylan, Zeynep, and Abdulkadir Atalan. 2021. "Estimation of Healthcare Expenditure per Capita of Turkey Using Artificial Intelligence Techniques with Genetic Algorithm-based Feature Selection." *Journal of Forecasting* 40(2):279–90. doi: 10.1002/for.2747.

Cheng, Feng-Chou, Ming-Chung Lee, Ling-Hsia Wang, Wen-Juain Lin, Tzu-Chiang Lin, Shiang-Yao Liu, and Chun-Pin Chiang. 2022. "Correlation between the Beverage Serving Activities and the Dental Use of Health Care Resources of National Health Insurance for Common Oral Diseases." *Journal of Dental Sciences* 17(4):1510–19. doi: 10.1016/j.jds.2022.06.004.

Chi, Chih-Yi, Hsin-Hung Wu, Chih-Hsuan Huang, and Yii-Ching Lee. 2017. "Using Linear Regression to Identify Critical Demographic Variables Affecting Patient Safety Culture From Viewpoints of Physicians and Nurses." *Hospital Practices and Research* 2(2):47–53. doi: 10.15171/hpr.2017.12.

Chiang, An Jen, Angus Jeang, Po Cheng Chiang, Po Sheng Chiang, and Chien-Ping Chung. 2019. "Multi-Objective Optimization for Simultaneous Operating Room and Nursing Unit Scheduling." *International Journal of Engineering Business Management* 11:184797901989102. doi: 10.1177/1847979019891022.

Daldoul, Dorsaf, Issam Nouaouri, Hanen Bouchriha, and Hamid Allaoui. 2018. "A Stochastic Model to Minimize Patient Waiting Time in an Emergency Department." *Operations Research for Health Care* 18:16–25. doi: 10.1016/j.orhc.2018.01.008.

Dönmez, Nergis Feride Kaplan, Abdulkadir Atalan, and Cem Çağrı Dönmez. 2020. "Desirability Optimization Models to Create the Global Healthcare Competitiveness Index." *Arabian Journal for Science and Engineering* 45(8):7065–76. doi: 10.1007/s13369-020-04718-w.

Gabrani, Adriatik, Adrian Hoxha, Jonila Gabrani (CYCO), Elizana Petrela (ZAIMI), Edmond Zaimi, and Endrit Avdullari. 2016. "Perceived Organizational Commitment and Job Satisfaction among Nurses in Albanian Public Hospitals: A Cross-Sectional Study." *International Journal of Healthcare Management* 9(2):110–18. doi: 10.1179/2047971915Y.0000000019.

Goienetxea Uriarte, Ainhoa, Enrique Ruiz Zúñiga, Matías Urenda Moris, and Amos H. C. Ng. 2017. "How Can Decision Makers Be Supported in the Improvement of an Emergency Department? A Simulation, Optimization and Data Mining Approach." *Operations Research for Health Care* 15:102–22. doi: 10.1016/J.ORHC.2017.10.003.

Ishikawa, Masatoshi. 2020. "Time Changes in the Geographical Distribution of Physicians and Factors Associated with Starting Rural Practice in Japan." *The International Journal of Health Planning and Management* 35(2):558–68. doi: 10.1002/hpm.2964.

Kan, Hong J., Hadi Kharrazi, Hsien-Yen Chang, Dave Bodycombe, Klaus Lemke, and Jonathan P. Weiner. 2019. "Exploring the Use of Machine Learning for Risk Adjustment: A Comparison of Standard and Penalized Linear Regression Models in Predicting Health Care Costs in Older Adults" edited by G. Stiglic. *PLOS ONE* 14(3):e0213258. doi: 10.1371/journal.pone.0213258.

Khatri, Naresh, Jack Wells, Jeff McKune, and Mary Brewer. 2006. "Strategic Human Resource Management Issues in Hospitals: A Study of a University and a Community Hospital." *Hospital Topics* 84(4):9–20. doi: 10.3200/HTPS.84.4.9-20.

Krogstad, U. 2004. "Doctor and Nurse Perception of Inter-Professional Co-Operation in Hospitals." *International Journal for Quality in Health Care* 16(6):491–97. doi: 10.1093/intqhc/mzh082.

Krogstad, Unni, Dag Hofoss, Marijke Veenstra, and Per Hjortdahl. 2006. "Predictors of Job Satisfaction among Doctors, Nurses and Auxiliaries in Norwegian Hospitals: Relevance for Micro Unit Culture." *Human Resources for Health* 4(1):3. doi: 10.1186/1478-4491-4-3.

Lee, I. Chen, Chao-Sung Chang, and Pey-Lan Du. 2017. "Do Healthier Lifestyles Lead to Less Utilization of Healthcare Resources?" *BMC Health Services Research* 17(1):243. doi: 10.1186/s12913-017-2185-4.

Leonard, M. 2004. "The Human Factor: The Critical Importance of Effective Teamwork and Communication in Providing Safe Care." *Quality and Safety in Health Care* 13(suppl_1):i85–90. doi: 10.1136/qshc.2004.010033.

Lin, Chin-I. 2008. "Optimization Models for Capacity Planning in Health Care Delivery." University of Florida, Florida.

Lovink, Marleen H., Anke Persoon, Raymond T. C. M. Koopmans, Anneke J. A. H. Van Vught, Lisette Schoonhoven, and Miranda G. H. Laurant. 2017. "Effects of Substituting Nurse Practitioners, Physician Assistants or Nurses for Physicians Concerning Healthcare for the Ageing Population: A Systematic Literature Review." *Journal of Advanced Nursing* 73(9):2084–2102. doi: 10.1111/jan.13299.

Malehi, Amal Saki, Fatemeh Pourmotahari, and Kambiz Ahmadi Angali. 2015. "Statistical Models for the Analysis of Skewed Healthcare Cost Data: A Simulation Study." *Health Economics Review* 5:11. doi: 10.1186/s13561-015-0045-7.

Maleki, Mohammad, Nahidsadat Majlesinasab, and Mohammad Mehdi Sepehri. 2014. "Two New Models for Redeployment of Ambulances." *Computers & Industrial Engineering* 78:271–84. doi: 10.1016/J.CIE.2014.05.019.

Mihaylova, Borislava, Andrew Briggs, Anthony O'Hagan, and Simon G. Thompson. 2011. "Review of Statistical Methods for Analysing Healthcare Resources and Costs." *Health Economics* 20(8):897–916. doi: 10.1002/hec.1653.

Montgomery, Douglas C., Elizabeth A. Peck, and G. Geoffrey Vining. 2012. *Introduction to Linear Regression Analysis*. 5th ed. Wiley.

Özcan, Tuncay, and Fatih Tüysüz. 2018. "Healthcare Expenditure Prediction in Turkey by Using Genetic Algorithm Based Grey Forecasting Models." Pp. 159–90 in *International Series in Operations Research and Management Science*.

Öztürk, Ahmet. 2016. *Yöneylem Araştırması*. 16th ed. Ekin.

Pellegrini, Lawrence C., Rosa Rodriguez-Monguio, and Jing Qian. 2014. "The US Healthcare Workforce and the Labor Market Effect on Healthcare Spending and Health Outcomes." *International Journal of Health Care Finance and Economics*. doi: 10.1007/s10754-014-9142-0.

Ramudhin, Amar, Eric Chan, and Abdelkader Mokadem. 2006. "A Framework for the Modelling, Analysis and Optimization of Pathways in Healthcare." Pp. 698–702 in *2006 International Conference on Service Systems and Service Management*. IEEE.

Schmid, Verena. 2012. "Solving the Dynamic Ambulance Relocation and Dispatching Problem Using Approximate Dynamic Programming." *European Journal of Operational Research* 219(3):611–21. doi: 10.1016/j.ejor.2011.10.043.

Shinjo, Daisuke, and Toshiharu Aramaki. 2012. "Geographic Distribution of Healthcare Resources, Healthcare Service Provision, and Patient Flow in Japan: A Cross Sectional Study." *Social Science & Medicine* 75(11):1954–63. doi: 10.1016/j.socscimed.2012.07.032.

Siciliani, Luigi, Anderson Stanciole, and Rowena Jacobs. 2009. "Do Waiting Times Reduce Hospital Costs?" *Journal of Health Economics* 28:771–80.

Steiner, Maria Teresinha Arns, Dilip Datta, Pedro José Steiner Neto, Cassius Tadeu Scarpin, and José Rui Figueira. 2015. "Multi-Objective Optimization in Partitioning the Healthcare System of Parana State in Brazil." *Omega* 52(Supplement C):53–64. doi: https://doi.org/10.1016/j.omega.2014.10.005.

Taylor, Carl E. 1976. "The Doctor's Role in Rural Health Care." *International Journal of Health Services* 6(2):219–30. doi: 10.2190/Q33J-0QC7-B2VA-L345.

Thapa, Chhitij, Abdul Ahad, Mohd. Aqil, Syed Sarim Imam, and Yasmin Sultana. 2018. "Formulation and Optimization of Nanostructured Lipid Carriers to Enhance Oral Bioavailability of Telmisartan Using Box–Behnken Design." *Journal of Drug Delivery Science and Technology* 44:431–39. doi: 10.1016/J.JDDST.2018.02.003.

TUIK. 2021. "Sağlık İstatistikleri, Istatistiksel Tablolar ve Dinamik Sorgulama." *Türkiye İstatistik Kurumu*. Retrieved (https://tuikweb.tuik.gov.tr/PreTablo.do?alt_id=1095).

Winston, Wayne L., and Munirpallam Venkataramanan. 2002. *Introduction to Mathematical Programming*. Fourth Edi. Thomson Learning.

Zhao, Jingtong, and Hanqi Wen. 2021. "Dynamic Planning with Reusable Healthcare Resources: Application to Appointment Scheduling." *Flexible Services and Manufacturing Journal*. doi: 10.1007/s10696-021-09411-0.

ACTA INFOLOGICA 2023;7(1):95-106
dergipark.org.tr/acin

İSTANBUL UNIVERSITY PRESS

DOI: 10.26650/acin.1235455

RESEARCH ARTICLE

User Experience Analysis of an Original Website Designed with Simplicity from the Perspective of Technology Acceptance Model

Sadelikle Tasarlanmış Özgün Bir Web Sitesinin Teknoloji Kabul Modeli Açısından Kullanıcı Deneyimi Analizi

Deniz Canca[1], **Ziya Nazım Perdahçı**[2]

ABSTRACT

This study aims to analyze the user experience (UX) of a plainly designed website on a comprehensive subject, to determine its effect and efficiency on a target group and to find out if it is a "must" to form a detailed design for such a mission. The analysis is conducted on an original website on school preference for university selection guidance. The website differs from others by bringing in the concept of presenting comparative comments about the schools of students who studied at two different universities, with no additional information. The research is conducted by using the Technology Acceptance Model (TAM), which is based on Information Architecture (IA), for UX analyses. The info processed by relational screening technique was obtained from 256 high school students through the internet by a survey regarding their opinions about the website. Data were analyzed by Structural Equation Modeling (SEM), in the context of five hypotheses. Findings confirmed that the IA factors predicted TAM factors, which showed that such a website holds a positive effect and efficiency for the measured factors. Moreover, a result other than those estimated was also achieved, which was the absence of an effect of one of the TAM factors to another, very probably due to the design's extreme plainness. These results present key messages both to web designers and the academy by showing that using plainness at the limits may possess extra positive outcomes and may reveal unexpected evidence in models.

Keywords: User experience, information architecture, context design, visual design, technology acceptance model

ÖZ

Bu çalışma, sade tasarlanmış bir web sitesinin kullanıcı deneyimini (UX) kapsamlı bir konuda analiz etmeyi, hedef kitle üzerindeki etkisini ve verimliliğini belirlemeyi, böyle bir web sitesi için detaylı bir tasarım oluşturmanın "zorunluluk" olup olmadığını ortaya çıkarmayı amaçlamaktadır. Analiz, üniversite seçim rehberliği için okul tercihi konulu orijinal bir web sitesinde gerçekleştirilir. Web sitesi, iki farklı üniversitede okuyan öğrencilerin okulları hakkında hiçbir ek bilgi olmaksızın karşılaştırmalı yorumlar sunma konseptini getirmesiyle diğerlerinden farklıdır. Araştırma, UX analizleri için Bilgi Mimarisini (IA) temel alan Teknoloji Kabul Modeli (TAM) kullanılarak yürütülmektedir. İlişkisel tarama tekniği ile işlenen bilgiler, internet üzerinden 256 lise öğrencisine site hakkındaki düşüncelerine yönelik bir anket yapılarak elde edilmiştir. Veriler, Yapısal Eşitlik Modellemesi (YEM) ile beş hipotez bağlamında analiz edilmiştir. Bulgular, IA faktörlerinin TAM faktörlerini öngördüğünü doğruladı ve bu, böyle bir web sitesinin ölçülen faktörler için olumlu bir etkiye ve verimliliğe sahip olduğunu gösterdi. Ayrıca tahmin edilenden farklı bir sonuca da ulaşılmıştır, bu da büyük olasılıkla tasarımın aşırı sadeliğinden dolayı TAM faktörlerinden birinin diğerine etkisinin olmaması şeklindedir. Bu sonuçlar, sadeliğin sınırlarda kullanılmasının ekstra olumlu sonuçlar doğurabileceğini ve modellerde beklenmedik kanıtları ortaya çıkarabileceğini göstererek hem web tasarımcılarına hem de akademiye önemli mesajlar sunmaktadır.

Anahtar Kelimeler: Kullanıcı deneyimi, bilgi mimarisi, bağlam tasarımı, görsel tasarım, teknoloji kabul modeli

[1](Dr.), Independent researcher
[2](Assoc. Prof.), Mimar Sinan Fine Arts University, Department of Informatics, Istanbul, Turkiye

ORCID: D.C. 0000-0002-5302-3877; Z.N.P. 0000-0002-1210-2448

Corresponding author:
Ziya Nazım PERDAHÇI
Mimar Sinan Fine Arts University, Department of Informatics, Istanbul, Turkiye
E-mail address: nz.perdahci@msgsu.edu.tr

Submitted: 24.01.2023
Revision Requested: 11.03.2023
Last Revision Received: 11.03.2023
Accepted: 02.04.2023
Published Online: 10.05.2023

Citation: Canca, D., Perdahci, Z.N. (2023). User experience analysis of an original website designed with simplicity from the perspective of technology acceptance model. *Acta Infologica.* Advance Online Publication. https://doi.org/10.26650/acin.1235455

1. INTRODUCTION

Information architecture (IA), which has a prominent place in structuring today's huge information load, refers to organizing data according to the minds of potential users and making the complex understandable. When it comes to information architecture in the internet; it is defined as the art and science of shaping information for usability through structural design in websites and mobile applications (Morville & Rosenfeld, 2006).

When developing a product in the context of IA, evaluating the users' physical and emotional reactions to that product is extremely important in optimizing the product. Herein, user experience (UX) is seen as a crucial component and can be defined as the users' perceptions and reactions resulting from the use of a product, system, or service. According to this definition, UX includes users' emotions, preferences, perceptions, physical and psychological reactions, behaviors, and achievements before, during and after the usage. In this context, three factors affecting the user experience are product, user, and use (ISO, 2018). UX enables designers to develop the product correctly by showing how easy and seamless users can perform the operations in the product and feel about it (Norman, 2002).

In this extent, context and visual design come forward. Usability, which is the level of ease of use of a product or system in IA, and content, which is the information itself, are the essential elements of the architecture. In addition, aesthetics has a strong effect on the entire structure and is an important side element in this context. Content design and usability, together with the visual design that forms aesthetics, affect user satisfaction in websites and mobile applications.

In websites, the content is the main element that distinguishes the site from its competitors and attracts the user. The content should have a rich and understandable structure that meets the needs of the target group and is suitable for user characteristics, such as age and gender. Content on websites is also shown as the number one factor for search engine optimization (Sparktoro, 2019). While the success of websites is mainly determined by their content, acceptance of good content by users is only possible with good visuals. Usability in websites and mobile applications is also mainly based on visual design. The usability of a product is determined by the visual control features of that product and their functionality (Telek, 2013). In addition, aesthetics based on visual design also affect user interest and satisfaction, as well as content and ease of use. Tranctinsky *et al*., (2000) and Norman (2002) emphasize that aesthetic elements can have such strong effects that they can make learning and using the application easier. Aesthetic elements sometimes even surpass functionality, and some even pave the way to become cult brands.

Simplicity, which has separately important places in content design and visual design, is another key factor behind a favored design (Nielsen, 1999), meaning that users on the web are able to get what they came for. Simplicity brings in clearness for context, ease of use for visual design, and a pleasant and attractive interface for aesthetics (Karvonen, 2000).

In this scope of design variables, the purpose of the study is investigating the user experience for an original website designed with absolute simplicity and revealing the efficiency of the design factors in such a design. It is obvious that the information to be obtained with UX on a website based on a simple design will make a good contribution to the literature.

For this study, the newly developed website "O Okul", a Turkish website meaning "That School" in English, is set as the subject material. The site, unlike other preference guidance sites, has been planned with a concept that gives guidance by comparing schools with the comments of students who have studied at those schools. In addition to this different point of view of the site, the principle of simplicity has been acted on throughout the editing and construction of the site. The website is presented to students who will enter the nationwide qualifying exam, which is the condition for enrollment to all universities in the country.

The model of the research is based on the Technology Acceptance Model (TAM), which is an effective model to reveal the factors that affect the adoption of the developments in today's technology. While the main factors of the model, perceived usefulness and perceived ease of use, are affected by external variables, they can affect users' intention to use and ultimately form the actual usage output. Data were analyzed by Structural Equation Modeling (SEM), the statistical technique to test and evaluate multivariate causal relationships of observed and latent variables. The study model covers 5 hypotheses regarding the interaction between five factors.

Within this scope, the research questions of the study are written as follows:

Does the content design and visual design at the website have an impact on users' perceptions of the ease of use and usefulness of the site?

If the designs in question influence user perceptions, do these perceptions have a positive effect on each other and on usage intention?

To find answers to these questions, the website is presented to a target group together with a survey on the context design and visual design of the site and the internal factors of the TAM. The analysis on the obtained data set of the UX, is implemented by the Analysis of Moment Structures (AMOS) through the software Statistical Package for the Social Sciences (SPSS).

2. RESEARCH BACKGROUND

2.1 The Original Website

The product put forth during the design process should serve a specific purpose, be functional because of conscious thought, and have at least one unique feature which is extraordinary in the sense that it has not been made before or is different from similar ones in functionality or aesthetics (Önlü, 2004). Design is a process that includes problem solving, creativity and aesthetics. It is the organization of a whole consisting of ideas, materials, and forms to serve a certain purpose (Whitely, 1993; Brooks Jr., 2010).

In this context, besides the school choice guidance websites on the Internet, the "O Okul" website in the study is a design that approaches the display of the schools from a different perspective and in a very simple way. The editing and construction of the site are made for functionality in a logical and theoretical integrity. While the schools are introduced separately with various sub-plugins on the existing sites, the schools are introduced by comparing them in "O Okul". Other websites involve comparison based on enrollment scores of the schools, which are most accurately available on the test center's website. Each pair of schools discussed in "O Okul" however, is compared by a student who studied at both schools. Here, while on one hand the fact that the most accurate comments about a school can be obtained from the experiences of the students from that school, on the other hand this factor is combined with the comparison method that will prevent students from being biased in their comments for their own school.

Since simplicity is based on both the contextual and visual design of the site (Karvonen, 2000; Maeda, 2006), this factor has been prioritized throughout the site setup. Except for the "About" section where the site ID is shared, there are only comparisons on the site. Comparisons are demonstrated on the main page in a box loop called "Carousel", which stops at certain time intervals. The carousel boxes contain only school logos and names, showing which schools are being compared, together with comment headings. When the boxes are clicked, the page with the relevant comment opens. Under the carousel, all comments are listed in pairs. Apart from the comment pages on the site, there is only one separate section for the option to search by school names. While configuring the site, the basic element in the page layout was again simplicity. The simple setup of the site is given as a flowchart in Figure 1 omitting the return to the home page link.

Figure 1. The setup of the website

2.2. Technology Acceptance Model (TAM)

IA is a design-based interdisciplinary and applied field for accessing and using today's information load (Resmini & Rosati, 2011). In this field, one of the most effective models developed on the factors affecting adoption of the rapid and major developments in technology is the Technology Acceptance Model (TAM). This model is aimed at understanding user behaviors

related to the presented product. The basis of TAM, formed by Fred Davis in 1986, constitutes the three variables "perceived usefulness" and "perceived ease of use", with the "attitude toward using" influenced by them. This foundation forms the structure of "user motivation". While external variables affect "perceived usefulness" and "perceived ease of use", "attitude toward using" directly determines "actual system use" output. The model is shown in Figure 2 with these variables (Davis, 1986).

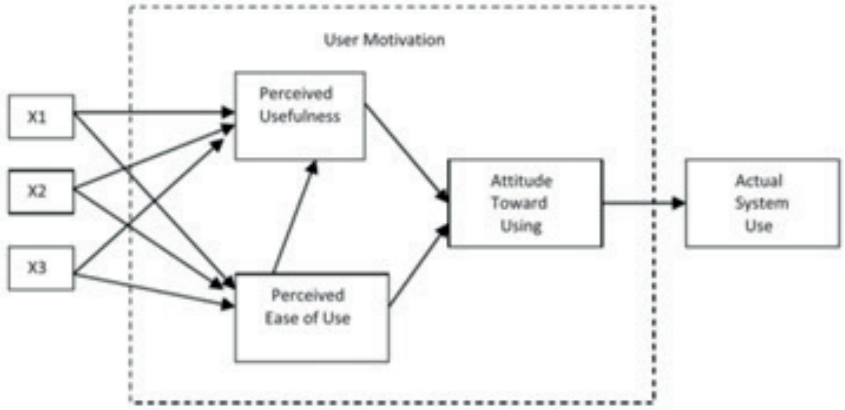


Figure 2. The main structure of Technology Acceptance Model

Numerous studies on the model have shown that it is superior to other models and makes meaningful predictions for both experienced and inexperienced users (Taylor & Todd, 1995). TAM has been subject to many studies since its emergence and has developed over time. The first model explained 40% of the intention to use (Davis, 1989). In 2000 TAM 2 was formed which increased the variance to 37%-52% by adding the factors that affect "perceived usefulness" to the model (Vankatesh & Davis, 2000). In 2008, by adding the factors that affect "perceived ease of use" to the model, TAM 3 was formed, which increased the variance to 40%-53% (Vankatesh & Davis, 2000). TAM is still being used effectively in assessing an existing or newly developed product, as well as in research to develop new models.

2.3 Research Hypotheses

There are five factors and five hypotheses in the research model. Three of the factors come from the original model as internal variables, while the other two influence these factors externally as independent variables. The factors are defined and the hypotheses are formulated as follows:

Contextual Design: The setting of the information in a product to meet the user's needs and present it in the best viable way (IDF, 2021).

Visual Design: Establishing the aesthetic appeal and usability of a product with appropriate elements, typography, space, layout, and colors (IDF, 2020).

Perceived Usefulness: The degree to which a person believes that using a particular system will improve work performance (Davis, 1986).

Perceived Ease of Use: The degree to which a person believes that using a particular system will not require physical and mental effort (Davis, 1986).

Attitude towards Using: The degree to which a person considers doing or not doing a certain future behavior (Warshaw & Davis, 1985).

H1: The contextual design of the website has a positive effect on perceived usefulness.

H2: The visual design of the site has a positive effect on perceived ease of use.

H3: Perceived ease of use for the website has a positive effect on perceived usefulness.

H4: Perceived usefulness for the website has a positive effect on the intention to use.

H5: Perceived ease of use for the website has a positive effect on the intention to use.

3. METHOD

3.1. The Research Model

This research, which aims to determine the effectiveness of a constructed website by applying it to a sample target audience, bears a relational design model. In this type of research, it is aimed to determine the existence and degree of a common change between two or more variables (Karasar, 2000).

The original model developed by Davis (1986) was directly taken as the model of the research. The dependent internal variables in this model are perceived usefulness, perceived ease of use, and intention to use, where contextual and visual design are reflected as independent external variables. In this context, the hypotheses set on the model are defined relationally in Figure 3.

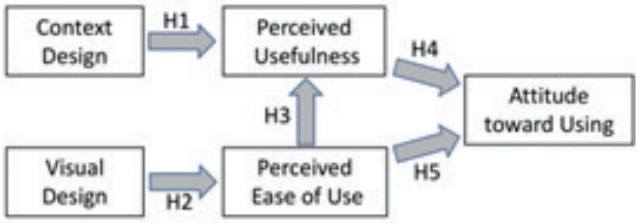


Figure 3. Model of the research

3.2. Population and Sample

The population of the research is students from public schools with high base points, and the sample is students from two well-known high schools within them; BAAL and SSAL, in Istanbul. The sample was determined by convenient sampling, where it is chosen from easily accessible and applicable places due to limitations in terms of time, money, and labor (Büyüköztürk, 2018). The two schools were selected for the ease of the application process as both schools had teachers who had been our classmates and had worked with us previously.

3.3. Data Collection Tools

The questionnaire consisted of 16 items for the basic components of TAM and 8 items for the IA factors affecting them. The 5-point scale lines up as; totally agree (4), partly agree (3), not sure (2), partly disagree (1), and totally disagree (0).

3.3.1 The Questionnaire

In the study, a 5-point Likert type questionnaire was prepared with 24 items, in which the questions obtained from the literature based on the "Technology Acceptance Model" were used. The visual form of the questionnaire and the limit on the number of items used were set by using the information obtained from the studies (Canca & Altun, 2011; Canca, 2011).

Items Related to IA

Content design:

- In my opinion, the content on the website is up-to-date and accurate.
- In my opinion, the content on the website is suitable for user needs.
- In my opinion, the content on the website is practical and useful.
- In my opinion, the content on the website is reliable.

Visual design:

- In my opinion, the website has been structured for easy use.
- In my opinion, the website is very good in terms of speed.
- In my opinion, the website is clearly designed for the organization of information.
- In my opinion, the website has a nice and high-quality look.

Items Related to TAM

Perceived usefulness:

- Using this website can enable me to make a quicker decision on my school choice.
- Using this website can enable me to make a better decision in choosing a school.
- Using this website would increase my productivity.
- Using this website would positively affect my effectiveness in my school choice.
- Using this website would make it easier for me to make a choice.
- Using this website would be helpful in making a school selection.

Perceived ease of use:

- Regarding this website, learning to use it is easy for me.
- Regarding this website, I can get what I want with the commands I give while using it.
- Regarding this website, I can use it comfortably.
- Regarding this website, I found it clear and understandable.
- Regarding this website, I can use it skillfully in a short time.
- Regarding this website, using it will be easy.

Attitude toward using:

- I would use this website in the coming months.
- I do not think I would want to use this website.
- I would recommend this website to my friends whom I want to help.
- I think that this website will be used by students because it will be liked.

3.3.2. Data Collection Process

For data collection, firstly, the questionnaire was transferred to an internet platform. Selection of the platform was made by uploading the questionnaire to three of the most preferred sites worldwide, and Survey Planet was chosen as the most suitable one for the study. We obtained ethics approval from the Provincial Directorate of National Education before inviting individuals to fill out the survey.

Personal information of users except for the school and class information was not collected due to the General Data Protection Regulation (GDPR) in EU law on personal data protection. Data from those who marked the same options in all the items or made a constant repetition or gave similar answers to the 21st and 22nd items, which have opposite statements, were excluded from the research.

3.3.3. Data Analysis Process

The descriptive and content analysis methods were used in data analysis. While in descriptive analysis the data set is evaluated in cause-effect relationships with previously formed themes, in content analysis the relations between the conceptualized data are determined (Karasar, 2000). The data set was analyzed with the software SPSS-AMOS.

Before the analysis, the adequacy of the sample size was checked. For SEM analyses, there are two approaches for the minimum number of participants: one considering only the number of participants, the other considering the ratio according to the number of items. These can be very briefly summarized as: 100, 150 and 200 responses in direct participant number, and at least 5 or 10 responses for each item in proportional participant number (O'Reilly, 2020). Considering these numbers, we note that 256 students who participated in the study and shared reliable information were above the required number.

3.4. Statistical Bases

Before leading to the testing of the hypotheses, statistical bases were checked by assessing the central distribution, reliability tests, and confirmatory factor analysis. After this overview, the structural equation of the model was setup and activated.

3.4.1. Central Distribution

In the study where responses ranged between 0 and 4, the mean and standard deviation values are presented in Table 1.

Table 1
Central distribution values of the questionnaire responses

Factors	Item heading	Mean	Standard deviation
Contextual design	Up-to-date and accurate	2.996	0.874
	Suitable for user needs	3.055	0.944
	Practical and useful	3.164	0.914
	Reliable	3.004	0.874
Visual design	Structured for easy use	3.219	0.982
	Fast	3.047	0.989
	Clear design	3.027	1.000
	Nice and quality look	2.730	1.145
Perceived usefulness	Enabling quicker decision	2.695	1.029
	Enabling better decision	2.902	1.026
	Increasing productivity	2.758	1.079
	Increasing effectiveness	2.922	1.007
	Easy in making a choice	2.879	1.028
	Helping in deciding	3.039	0.993
Perceived ease of use	Easy to learn to use	3.395	0.952
	Responds well to commands	3.203	0.928
	Easy interaction	3.230	0.906
	Clear and understandable	3.227	0.922
	Quick mastering in usage	3.324	0.877
	Easy usage	3.402	0.862
Attitude to use	Will use in the future	2.672	1.089
	Will not use in the future	1.242	1.076
	Will recommended	2.840	1.049
	Will be liked by others	2.977	0.998

It is observed that the views on the external variables are gathered around the "partly agree" option, which is expressed as three on the 0-4 scale. Distribution of the views on internal variables show that the means gather slightly below the "partly agree" value for "perceived usefulness", and slightly above the "partly agree" value for "perceived ease of use". For the output factor, "attitude towards using", opinions are gathered slightly below the value for "partly agree".

3.4.2. Reliability Analysis

After these findings, the model consisting of 24 items of five factors was tested for reliability (see Table 2.).

Table 2
Factor estimates and convergent validity results of the model.

Factors	Item no	Unstandardized estimates	Standardized estimates	Explained variance	Cronbach's Alpha	Composite reliability	Average variance extracted
Contextual design	1	1.256	0.736	0.542	0.815	0.727	0.401
	2	1.387	0.753	0.568			
	3	1.473	0.826	0.683			
	4	1.000	0.586	0.344			
Visual design	5	1.100	0.718	0.516	0.720	0.677	0.352
	6	0.799	0.518	0.268			
	7	1.166	0.748	0.559			
	8	1.000	0.560	0.313			
Perceived usefulness	9	1.000	0.797	0.635	0.935	0.906	0.618
	10	1.032	0.825	0.681			
	11	1.055	0.801	0.642			
	12	1.100	0.896	0.803			
	13	1.035	0.826	0.682			
Perceived ease of use	14	0.999	0.825	0.680	0.892	0.874	0.537
	15	1.040	0.776	0.602			
	16	0.892	0.683	0.467			
	17	0.996	0.781	0.609			
	18	1.028	0.792	0.627			
	19	0.836	0.677	0.459			
Attitude to use	20	1.000	0.824	0.678	0.885	0.773	0.440
	21	1.000	0.791	0.625			
	22	-0.994	-0.796	0.633			
	23	0.932	0.765	0.585			
	24	0.975	0.841	0.708			

All standardized estimates in the model were above 0.5 (Table 2) and Cronbach's Alpha values of all factors were found to be high within the scope of the reliability (Bademci, 2006) of the application.

3.4.3. Confirmatory Factor Analysis

Confirmatory Factor Analysis is used in scale development and validity analyses or to confirm a predetermined structure (Yaşlıoğlu, 2017). To examine the suitability of the data set in the study for the analysis, firstly correlations between the factors were checked (Table 3).

Table 3
Connection between the factors ($p < 0.001$)

Factors	CD	VD	PU	PEU	ATU
CD		0.561	0.705	0.563	0.592
VD	0.561		0.573	0.739	0.471
PU	0.705	0.573		0.520	0.687
PEU	0.563	0.739	0.520		0.499
ATU	0.592	0.471	0.687	0.499	

While there is a significant ($p < 0.001$) relation between all factors, the correlations between VD-PEU, CD-PU, and PU-ATU factors were found to be high, which respectively correspond to the hypotheses H2, H1, and H4.

4. FINDINGS

4.1. Data of the First Model

Numerical data on structural relations were obtained by setting up the factors according to the original TAM in AMOS with the SEM. Considering that the sample size is larger than 250 and the number of variables is between 12 and 30, the fit values of the first model according to the reference values for the ideal fit (Byrne, 2011) are given in Table 4.

Table 4
Fit values of the first model

Goodness of fit statistic	Reference values	Fit values of the first model
CMIN/DF	< 5	2.948
GFI	> 0.90	0.819
CFI	> 0.92	0.887
NFI	> 0.90	0.839
TLI	> 0.90	0.873
RMSEA	< 0.07	0.087

4.2. Data of the Renewed Model

The changes that AMOS indicates can be understood by looking at the central distribution values and item contents. While no two-way relation was established between any of the items in the external variables, similarity between the linked item pairs appears to be the reason for this change. These changes had a positive effect on factor relations and fit values in general (Table 5). The renewed model is presented in Figure 4.

Table 5
Fit values of the renewed model

Goodness of fit statistic	Reference values	Fit values of the renewed model
CMIN/DF	< 5	1.829
GFI	> 0.90	0.875
CFI	> 0.92	0.953
NFI	> 0.90	0.902
TLI	> 0.90	0.946
RMSEA	< 0.07	0.049

Except for the GFI value in the second model, all values show that the model has reached the level required for an ideal fit. After the arrangements suggested by the AMOS program were made, data of the latest model was obtained, and was proceeded to testing of the hypotheses.

4.3. Confirmatory Factor Analysis

At the end of the SEM analysis, relations between the variables and the regression estimates in the model are presented in Table 6.

Table 6
Predictive findings on the hypotheses

Hypotheses	Regression loads	Standard loads	t values	p values	Result of the hypothesis
H1: CD > PU	1.347	0.183	7.377	0.000	Verified
H2: VD > PEU	1.018	0.119	8.528	0.000	Verified
H3: PEU > PU	-	-	-	0.850	Falsified
H4: PU > ATU	0.831	0.076	10.926	0.000	Verified
H5: PEU >ATU	0.238	0.064	3.706	0.000	Verified

While all hypotheses are presented in Table 6, only the rejection of H3 draws attention. Regarding this finding, additional tests were conducted on the model (Sen, 2019). Firstly, the correlational arrow between CD and PU has been removed from the additions indicated by AMOS. Separately, two predictive arrows have been added from CD to PEU and from VD to PU. Again, apart from these trials, each of the item pairs with correlational arrows was selected and removed according to the load change. In each case, while the fit values weakened, there was no significant change for H3. As a result, it is seen that all the tested hypotheses are supported except the one related to the effect of perceived ease of usee on perceived usefulness.

5. DISCUSSION AND CONCLUSION

The first hypothesis of the study (H1) is on the effect of contextual design of the website on perceived usefulness. Findings confirm that contextual design positively affects perceived usefulness, as expected. It is an expected situation that the content of a site that is consulted about choosing a school will affect the perceived benefit in this regard, as would be expected in almost every subject. The fact that the contextual design of the site predicts perceived usefulness in a positive way also shows that it is organized in accordance with the purpose.

The second hypothesis (H2) of the study is on the effect of the visual design of the website on perceived ease of use. Findings confirm that visual design positively affects perceived ease of use, as expected. It is expected that the visual design of a site to be used for school preference will affect the perceived ease of use in this regard, as would be expected in almost every subject. The fact that visual design predicts perceived usefulness in a positive way also shows that the visual design on the site is organized in accordance with the purpose.

The third hypothesis (H3) of the study is on the effect of perceived ease of use on perceived usefulness. The findings do not show that perceived ease of use has an effect on perceived usefulness, contrary to what is reported in the literature. The fact that perceived ease of use does not have a significant effect on perceived usefulness is explainable with respect to the interaction between them for such a website. TAM is a model developed for a wide variety of products in the field of technology. Therefore, of course, it is expected that ease of use will directly affect usefulness when certain products are used for certain benefits. Technological products such as smart home appliances, autonomous vehicles and facial recognition systems are examples of this. However, it cannot be said that ease of use has a direct effect on the usefulness for the designed website within the scope of its concept. Moreover, simplicity in its design is likely to keep this relation at a minimum level. The absence of interaction between these factors also shows that perceived usefulness and perceived ease of use can be considered quite independently of each other in view of the external factors of contextual and visual design.

According to TAM, individuals tend to use a new product to the extent of their perceptions of its ease of use and the benefits it will provide, while PEU has an impact on PU (Silva & Diaz, 2007). These two main factors remain the most influential variables in the model, although many new factors have been tested in studies (Avcı Yücel & Gülbahar, 2013). However, it is seen that many studies refer to TAM due to the simplicity of the model without considering the nature of the relations between factors [33]. Considering the intention to play digital games; finding them difficult or easy, and seeing them as a waste of time or mental gymnastics, is a good example of the use of TAM in such studies (Charness & Boot, 2016). Likewise, the two factors in this study may not have a predictive effect on one another. Another reason for the finding about H3 in practice may be the fact that the PEU remains at an extremely high level, as the product is found so easily by the user group. Naturally, one fact that minimizes the effect of PEU on PU can be the simplicity of the design. The rejection of H3 also does not change the fact that PEU and PU are the main variables in the model.

The fourth hypothesis (H4) of the study is on the effect of perceived usefulness on intention to use. Findings confirm that perceived usefulness positively affects intention to use, as stated in the literature. Intrinsically, the perceived usefulness of a website on school choice may have a direct effect on the intention to use, and so it is expected that it can predict this output factor.

The fifth hypothesis (H5) of the study is on the effect of perceived ease of use on intention to use. Findings confirm that perceived ease of use positively affects intention to use, as stated in the literature. For a website on school preference, a direct effect of perceived ease of use on intention to use is not expected, but it may still predict it indirectly.

In the results, it is seen that the information architecture factors within the scope of the research have a positive effect on the adoption of the sample website. This shows that the contextual and visual designs are formed at a good level according to the criteria within the scope of the research, considering the target group.

Considering the findings based on factors, it can be said that the success in contextual and visual design is at similar levels. Unlike this similarity between external factors, it is seen that there is a difference in favor of ease of use within the measured scope between the levels of perceived usefulness and perceived ease of use.

Considering the findings based on items, additional comments can be made about the overall design and certain points. The first is about the lack of names of graduates on the website. This was for the purpose of obtaining sincere shares from the authors on the website. Despite this lack, as can be seen in the findings of item #4, users assumed that the comments were reliable. Similarly, there was no info about the graduation year of the authors. Despite this fact, as can be seen in the findings of item #1, users assumed that the comments on the website were up to date. This shows that the test group considered the information on the website reliable despite these two shortcomings.

This study analyses the UX of an extremely plainly designed website on a special comprehensive subject and reveals that it is not a "must" to form a detailed design for such a mission for the website designers. Furthermore, findings confirmed that an extremely plain and well-designed website has a positive effect for the attitude towards using the website. Moreover, for literature, the study presents that, probably due to the extreme plainness of the design, the effect of perceived ease of use on perceived usefulness disappears. These results are pleasing, as showing positive facts of plain design on a particular topic, and these types of designs embody data that have not been explored yet.

To raise the original website design to even better levels, enhancements and follow-up new measurements should also be made according to open-ended comments from users. In addition, since perceived usefulness, perceived ease of use, and intention to use, which are the main variables of TAM, were discussed in the study; the actual use, which is directly affected by the intention to use, has not been measured. This factor, which requires more different and detailed tools for measurement, can be measured by allocating more time and resources in development processes.

Acknowledgement: This submission appeared as part of Deniz Canca's master's thesis, titled "Okul tercihine yönelik orijinal bir web sitesi tasarımının teknoloji kabul modeli ile kullanıcı deneyimi analizi" with advisor Ziya Nazim Perdahci.
Peer-review: Externally peer-reviewed.
Author Contributions: Conception/Design of Study- N.Z.P., D.C.; Data Acquisition- D.C.; Data Analysis/Interpretation- D.C., N.Z.P.; Drafting Manuscript- D.C.; Critical Revision of Manuscript- N.Z.P.; Final Approval and Accountability- N.Z.P., D.C.; Material and Technical Support- D.C.; Supervision- N.Z.P.
Conflict of Interest: The authors have no conflict of interest to declare.
Grant Support: The authors declared that this study has received no financial support.

REFERENCES

Ajibade, P. (2018). Technology Acceptance Model Limitations and Criticisms: Exploring the Practical Applications and Use in Technology-Related Studies, Mixed-Method, and Qualitative Researches. Library Philosophy and Practice, Summer 2018, 1-13.

Avcı Yücel, Ü., Gülbahar, Y. (2013). Technology Acceptance Model: A Review of the Prior Predictors. Ankara University, Journal of Faculty of Educational Sciences, 46(1), 89-109.

Bademci, V. (2006). Güvenirliği Doğru Anlamak ve Bazı Klişeleri Yıkmak: Bilinenlerin Aksine, Cronbach'ın Alfa Katsayısı, Negatif ve "–1"den Küçük Olabilir. İnönü Üniversitesi Eğitim Fakültesi Dergisi, 7(12), 3-26.

Byrne, B.M. (2011). Structural Equation Modeling with AMOS: Basic Concepts, Applications and Programming. Routledge, Taylor & Francis Group, Abingdon.

Brooks Jr, F.P. (2010). The Design of Design: Essays from a Computer Scientist. Pearson Education Inc., Boston MA.

Büyüköztürk, Ş. (2018). Bilimsel Araştırma Yöntemleri. Pegem Akademi Yayıncılık, Ankara.

Canca, D. (2011). Do They Really Agree When They Say "I Agree"? International Journal of Arts and Sciences, Prague Conference 2011 Proceedings.

Canca, D., Altun, S. (2011). An Approach to the Optimum Length of Research Questionnaires. International Journal of Arts and Sciences, Prague Conference 2011 Proceedings.

Charness, N., Boot, W.R. (2016). Technology, Gaming, and Social Networking. Handbook of the Psychology of Aging. Academic Press, Elsevier Inc.

Davis, F.D. (1986). A Technology Acceptance Model for Empirically Testing New End-User Information Sustems: Theory and Results. Ph.D. Thesis, MIT.

Davis, F.D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly, 13(3), 319-40.

IDF, (2020). IDF, (2020). What is Visual Design? Visual Design. interaction-design.org/literature/topics/visual-design /

IDF, (2021). Contextual Design. interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/contextual-design

ISO, (2018). Ergonomics of Human-System Interaction – Part 11: Usability: Definitions and Concepts. iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en

Karasar, N. (2000). Bilimsel Araştırma Yöntemi, Nobel Akademik Yayıncılık, Ankara.

Karvonen, K. (2000). The Beauty of Simplicity. Proceedings on the 2000 Conference on Universal Usability.

Maeda, J. (2006). The Laws of Simplicity: Design, Technology, Business, Life. The MIT Press, Cambridge MA and London.

Morville, P., Rosenfeld, L. (2006). Information Architecture for the World Wide Web: Designing Large-Scale Web Sites. O'Reilly Media, Sebastopol, CA.

Nielsen, J. (1999). Designing Web Usability. The Practice of Simplicity. New Riders Publishing, Indiana, 1999.

Norman, D. (2002). The Design of Everyday Things. Basic Books, New York

O'Reilly (2020). The Rules of Thumb for Sample Size Needed for SEM. oreilly.com/ library/view/structural-equation-modeling/9781118356302/c07anchor-1.html

Önlü, N. (2004). Tasarımda Yaratıcılık ve İşlevsellik. Atatürk Üniversitesi SBE Dergisi, 3(1), 85-95.

Resmini, A., Rosati, L. (2011). A Brief History of Information Architecture. Journal of Information Architecture, 3(2), 33-45.

Silva, P.M., Diaz, G.A. (2007). Theories About Technology Acceptance: Why The Users Accept or Reject the Information Technology

Sparktoro (2019). Google Ranking Factors 2019: Opinions from 1500+ SEO's. sparktoro.com/blog/resources/google-ranking-factors-2019/

Şen, S. (2019). Yapısal Eşitlik Modellemesi. Harran Üniversitesi Eğitim Fakültesi Ders Notları, sedatsen.files.wordpress.com/2019/04/sem-teorik-sunum.pdf

Taylor, S., Todd, P.A. (1995). Understanding Information Technology Usage: A Test of Competing Models. Information Systems Research, 6(2), 144-176.

Telek, C. (2013). Kullanılabilirlik Kavramı, Tasarım Süreci İçindeki Yeri ve Benzer Tasarım Yaklaşımları ile İlişkisi. Yüksek Lisans Tezi, MSGSÜ.

Tranctinsky, N., Katz, A.S., Ikar, D. (2000). What is Beatiful is Usable. Interacting With Computers, 13(2), 127-145.

Venkatesh, V., Bala, H. (2008). Technology Acceptance Model 3 and a Research Agenda on Interventions. Decision Sciences, 39(2), 273-315.

Vankatesh, V., Davis, F.D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science, 46(2), 186-204.

Warshaw P.R., Davis, F.D. (1985). Disentangling Behavioral Intention and Behavioral Expectation. Journal of Experimental Social Psychology, 21(3), 213–228.

Whiteley, N., (1993). Design for Society. Reaktion Books Ltd, London.

Yaşlıoğlu, M.M. (2017). Sosyal Bilimlerde Faktör Analizi ve Geçerlilik: Keşfedici ve Doğrulayıcı Faktör Analizlerinin Kullanılması. İstanbul Üniversitesi İşletme Fakültesi Dergisi, 46: 74-85.

ACTA INFOLOGICA 2023;7(1):107-123
dergipark.org.tr/acin

İSTANBUL UNIVERSITY PRESS

DOI: 10.26650/acin.1132510

RESEARCH ARTICLE

Analysis of City Demographics in Turkiye Using Data Mining Techniques

Türkiye'de Şehir Nüfus İstatistiklerinin Veri Madenciliği Yöntemleriyle İncelenmesi

Özge Doğuç[1], Kevser Şahinbaş[2], Gökhan Silahtaroğlu[1]

[1](Prof. Dr.) Medipol University, Faculty of Business and Management Sciences, Department of Management Information Systems, Istanbul, Turkiye
[2] (Asst. Prof.) Medipol University, Faculty of Business and Management Sciences, Department of Management Information Systems, Istanbul, Turkiye

ORCID: Ö.D. 0000-0002-5971-9218;
K.Ş. 0000-0002-8076-3678;
G.S. 0000-0001-8863-8348

Corresponding author:
Özge DOĞUÇ
Medipol University, Faculty of Business and Management Sciences, Department of Management Information Systems, Istanbul, Turkiye
E-mail address: odoguc@medipol.edu.tr

Submitted: 18.06.2022
Revision Requested: 05.09.2022
Last Revision Received: 14.03.2023
Accepted: 12.04.2023
Published Online: 13.06.2023

Citation: Doguc, O., Sahinbas, K. & Silahtaroglu, G.(2023). Analysis of city demographics in Turkiye using data mining techniques. *Acta Infologica, 7*(1), 107-123. https://doi.org/10.26650/acin.1132510

ABSTRACT

In this study, a data mining model has been developed and used to analyze how cities and regions in Turkey can be grouped, aiming to find similarities and differences between them. For this purpose, data is obtained from Turkish Statistical Institution (TUIK) and fuzzy c-means clustering algorithm was used to find categorizations. The data set contains 142 variables from 8 categories such as education, health, happiness, and development levels. The results showed that in all categories, the biggest 3 cities in Turkey, İstanbul, Ankara, and İzmir are different from the rest of the country. Also, cities located in the western and eastern regions of Turkey are mostly grouped among themselves, showing the clear distinction between those two regions. Finally, small cities with big neighbors are grouped with other big cities, showing the direct impact of big cities on their neighbors. This also implies that small cities with no big neighbors are often isolated, as their residents don't have access to the services provided in the big cities.

Keywords: Data mining, clustering analysis, data management, demographic analysis

ÖZ

Bu çalışmada, Türkiye'deki şehirlerin ve bölgelerin nasıl gruplanabileceğini analiz etmek, aralarındaki benzerlik ve farklılıkları bulmak amacıyla bir veri madenciliği modeli geliştirilmiş ve kullanılmıştır. Bu amaçla Türkiye İstatistik Kurumu'ndan (TÜİK) veriler elde edilmiş ve kategorizasyonları bulmak için bulanık c-ortalamalar kümeleme algoritması kullanılmıştır. Veri setinde eğitim, sağlık, mutluluk ve gelişmişlik düzeyleri gibi 8 kategoriden 142 değişken yer almaktadır. Sonuçlar, Türkiye'nin en büyük 3 şehri olan İstanbul, Ankara ve İzmir'in tüm kategorilerde ülkenin geri kalanından farklı olduğunu gösterdi. Ayrıca, Türkiye'nin batı ve doğu bölgelerinde yer alan şehirler çoğunlukla kendi aralarında gruplandırılarak bu iki bölge arasındaki açık fark görülmektedir. Son olarak, büyük komşuları olan küçük şehirler, büyük şehirlerin komşuları üzerindeki doğrudan etkisini göstererek diğer büyük şehirlerle gruplandırılmıştır. Bu aynı zamanda, sakinlerinin büyük şehirlerde sağlanan hizmetlere erişimi olmadığı için, büyük komşuları olmayan küçük şehirlerin genellikle izole edildiği anlamına gelir.

Anahtar Kelimeler: Veri madenciliği, kümeleme analizi, veri yönetim, demografik analiz

1. INTRODUCTION

Turkey has 81 provinces and 7 geographic regions. Although most cities have unique characteristics, cities in each region may have similar cultural characteristics due to regional similarities. Therefore, a city in the Marmara Region and a city in the Eastern Anatolia region may show similarities in some of their demographic characteristics. Knowing in which features the cities are similar can help with the decisions that are made at the state level. For example, a proposed investment in a city can be planned for other cities that have been experiencing similar problems or have similar needs. Areas that cities need to develop or investments they need can be observed in more detail, and it will be easier for the investments to spread throughout the country if these similarities and differences are well-analyzed.

Data analytics is a science that analyzes existing historical data with statistical and machine learning methods and extracts useful information (Silahtaroğlu, 2008). As its name suggests, the quality of the data to be provided to machine learning is extremely important. For example, providing the age of the person as input can help the method to discover patterns and make predictions, but also providing the sign of the same person, along with the age category (such as child, young, old, very old) will help the machine to reveal different and hidden information about that individual. Similarly, correlating provinces or cities to people, such as the person's place of birth and city of residence helps the data analytic methods to extract information and patterns about populations. At this point, like the age example, in addition to the name of the province, using variables such as its population, literacy, and patients per physician will help the method produce a more detailed result.

This study uses 142 demographic variables about Turkish cities, spanning 5 years. This is the first study in the literature that uses such as wide range of demographic variables in Turkish cities to perform data analytics; as the studies in the literature had much limited scopes. This study provides a general clustering analysis of the cities that was carried out by collecting data from a wide range of categories. Therefore, unsupervised algorithms are used to find the groups of attributes as well as clusters of the cities that provide the best resemblance (similarities). This study aims to help the local governments and institutions by providing a detailed analysis of how and why some of the cities exhibit similarities and discuss how these similarities can be used in budget and investment planning.

2. LITERATURE SURVEY

There are several studies in the literature on cluster analysis using data provided by the Turkish Statistical Institution (TUIK). In one of these studies, Bulut (2019) aimed to cluster Turkish citizens based on the life satisfaction index values. In the study, a total of 11 index values and indicators showing the life satisfaction rates of the citizens were used as variables. The index values used in the study were housing, health, environment, access to infrastructure services, social life, income and wealth status, security, business life, civic participation, education, and life satisfaction. The value range of index values is between 0 and 1, and values closer to 1 represent higher living standards. In the study, Expectation Maximization (EM) and k-means algorithms are used. As a result of the study, the number of interpreted clusters was obtained as 2. Cities in the 2nd cluster generally consist of the cities located in the Southeastern and Eastern regions of Turkey. With this result, it has been seen that the satisfaction or dissatisfaction of the citizens in Turkey is caused by a regional problem. (Bulut, 2019).

In his study, Yilanci (2010) classified the cities in Turkey from a socioeconomic point of view using fuzzy clustering analysis. In addition to the fuzzy cluster analysis method, the k-means method was also used in the study for comparison. The study used 11 socioeconomic variables such as population density, unemployment rate, number of insured people, public investment expenditures, rate of higher education graduates, total agricultural production values, population per physician, infant mortality rate, the total number of people receiving a pension, gross domestic per capita. The study revealed two clusters of cities where the cities in the first cluster can be classified as developed when evaluated in terms of socioeconomics, and the cities in the second cluster can be classified as underdeveloped cities when evaluated in terms of socioeconomics. (Yılancı, 2010).

In the study by İncekirik and Altin (2021), the transportation data of the cities in Turkey were evaluated and classified by cluster analysis. The data set used for the study includes the transportation statistics between the years 2004-2018 and

contained 40 variables. As a result, 5 clusters of cities were discovered; where the clustering algorithm placed the cities that are developed and located on the coast with significant export and import activities in different clusters than the cities that did not have advanced transportation alternatives and thus were open to further development. (İncekırık & Altın, 2021).

In another study done by Tekin (2015), health data from 81 provinces of Turkey were examined to identify similar provincial groups. The study used data from 2013 with 16 health indicators; and aimed to provide a comparison of development levels of different cities using health and socioeconomic indicators. The study revealed significant differences in the quality of health services provided in the western and eastern regions of Turkey. (Tekin, 2015).

In Kandemir's (2018) study, the clustering of provinces in Turkey was made according to accommodation statistics. The study aims to determine the cities with priority in tourism and discuss planning to allocate resources to the cities that do not have priority. Using the data from 2016, the study showed that Ankara, Antalya, and Istanbul do not have a definite cluster membership with other cities and the distance between cities does not affect this situation. When it comes to domestic and foreign tourists, it is seen that Antalya is the most preferred city, as expected. And for foreign tourists, Çorum province is the least preferred and according to the number of domestic tourists, Osmaniye is the least preferred province (Kandemir, 2018).

This study analyzes the demographics of Turkish cities across 142 attributes by using an unsupervised learning method, clustering. Unlike the previous studies in the literature, this study analyzes the demographic data from a broad perspective and finds demographic attributes that show a resemblance between Turkish cities. The sections discuss the clustering methods that are used in this study and introduce the data set that is used for generating results.

2.1 K-MEANS CLUSTERING ALGORITHM

K-means is a popular clustering algorithm that is used to classify data. The algorithm divides the data to be classified into k classes or clusters based on their properties. Classification is done by the distribution of the selected data around the center points of the clusters, which show the closest features in terms of similarity to each other. The name of the algorithm comes from the requirement for the number of clusters to be known and constant. The number of clusters is expressed with the letter k, and it is also the representation of the number of clusters that will be created according to the similarity between the data. (Silahtaroğlu, 2004).

Figure 1 illustrates how the k-means algorithm works. In the visualized example, k is taken as 3. The white symbols (triangle, circle, square) seen in (a) on the left resemble the initial cluster centers, and they were chosen randomly. Afterward, the data points shown in the first iteration (b) are assigned to the nearest center and shown with the same symbols (triangle, circle, square). Next, the cluster centers are recalculated, considering the average of the elements in each cluster. Cluster centers (shown by looking at moving arrows) are redetermined. Steps b and c are performed again if there is a change in the cluster centers. In the example in the figure, the algorithm ends up dividing it into three clusters. (Böhm, 2001).

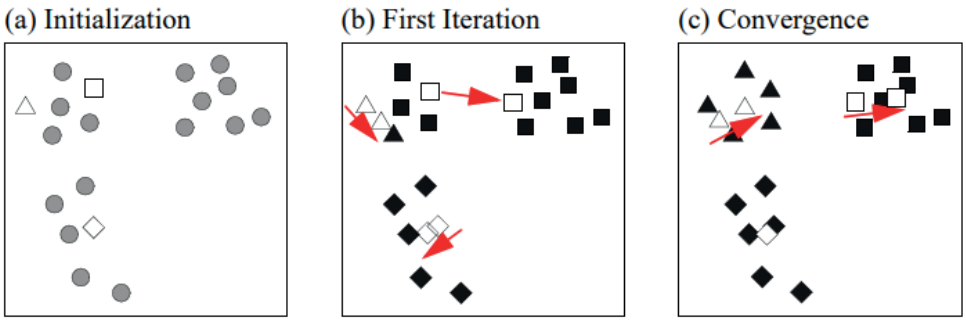


Figure 1. Representation of the K-Means algorithm (Böhm, 2001)

2.2 FUZZY C-MEANS ALGORITHM

The Fuzzy c-means algorithm was first introduced by Dunn in 1973 and implemented by Bezden in 1981. (Höppner et al., 1999) It is one of the most popular and most used fuzzy clustering algorithms in the literature. Different than regular clustering algorithms such as k-means, fuzzy c-means allows data elements to be assigned to multiple clusters. (Kruse et al., 1999) It uses 'goal functions' for cluster assignments and tries to minimize the goal score at every iteration. (Işık & Çamurcu, 2011). Like the k-means algorithm, it reevaluates the cluster centers and recalculates the goal function at every iteration. The algorithm iterates until the goal score is less than ε. (Silahtaroğlu, 2013). Like the k-means algorithm, the performance of fuzzy c-means is dependent on the initial cluster center assignments. (Hekim & Orhan, 2011). However, unlike the k-means algorithm, fuzzy c-means provide 'soft assignments' to the clusters, where each assignment contains a probability. Therefore, the fuzzy c-means algorithm can track an unlimited number of cluster assignments and pick the ones with the highest probabilities.

3. DATA ACQUISITION AND METHOD

The data set used in this study was collected from multiple data sources through TUIK. Data set contains data from 2015-2020 and includes 23 major categories as follows:

> *"Education", "Demography", "Population and Migration", "Health", "Environment", "Transportation", "Justice", "Foreign Trade", Energy", "Construction and Housing", "Culture", "Industry", "Agriculture and Livestock", "Tourism", "National Accounts", "Working Life", "Civil Participation", "Income and Wealth", "Security", "Access to Infrastructure Services", "Social Life", "Happiness" Level".*

Each category involves a number of variables; and thus, the data set contains 142 variables in total. Table 1 below shows the variables in each category.

Table 1
Demographic categories and variables used in this study

Category Name	Variables
General Information	Population Female Population Male Population Population Percentage Area (Km^2) Number Of Districts Number Of Municipalities Number Of Villages
Education	Literacy rate (%) Number of primary school students Number of secondary school students Number of high school students Number of primary school teachers Number of secondary school teachers Number of high school teachers Number of primary schools Number of secondary schools Number of high schools Primary and Secondary Schools / Number of Students Per Classroom High Schools / Number of Students Per Classroom Number of Students Per Teacher (Primary School) Number of Students Per Teacher (Secondary School) Number of Students Per Teacher (High School) Primary School Enrolment Rate (female) Primary School Enrolment Rate (overall) Primary + Secondary Schooling Rate (female) Primary + Secondary Schooling Rate (overall) High School Enrolment Rate (female) High School Enrolment Rate (overall) Net enrolment rate in pre-primary education (3-5 years) (%) The mean score for placing the TEOG system. YGS average score Percentage of faculty or college graduates (%) Satisfaction rate of public education services (%)

Demography	Infant Mortality Rate (per thousand) Number of Divorces Number of Births Expected Lifespan at Birth (years) Number of Marriages Divorce Rate (per thousand) Birth Rate (per thousand) Marriage Rate (per thousand) Death Rate (per thousand) Number of Deaths Total Fertility Rate (number of children) Under-5 Mortality Rate (per thousand)
Population And Migration	Child Dependency Rate (%) Net Migration Rate (per thousand) Population Density (number of people per square kilometer) Average Household Size Total Number of Households Total Age Dependency Rate (%) Immigration from Turkey to Abroad Elderly Dependency Rate (%) Annual Population Growth Rate (per thousand) Migration from Abroad to Turkey
Health	Total Number of Physicians Per Thousand People Number of Hospitals Number of Hospital Beds Total Number of Hospital Beds per Hundred Thousand People Number of Applications Per Physician Health satisfaction rate (%) Satisfaction rate of public health services (%)
Environment	Wastewater Collection Rate (%) Wastewater Recycle Rate (%) Access to Potable Water Rate (%) Potable Water Recycle Rate (%) (Air Pollution) Average of PM10 Station Values (μg/m³) Forest Area per km² (%) Rate of Street Noise Complaints (%) Cleaning Services Satisfaction Rate (%)
Transport	Number of Cars per Thousand People Number of Land Vehicles Number of Cars Number of Traffic Accidents
Justice	Number of Convicts Entering the Penitentiary Institution According to the Committed Crime
Foreign Trade	Total Exports (thousand $) Total Imports (thousand $)
Energy	Total Electricity Consumption per Person (kWh)
Construction And Housing	Number of Housing sales (first sale) Number of Housing Sales (total) Number of Buildings by Occupancy Permit Number of Flats by Occupancy Permit Area According to the Occupancy Permit (square meters) Number of Buildings by Building Permit Number of Flats by Building Permit Area According to Building Permit (square meters) Number of Rooms Per Person Availability of toilets in the residences (%) Housing services quality satisfaction rate (%)

Culture	Number of Public Library Users Number of Books in Public Libraries Number of Public Libraries Number of Museum Artifacts Affiliated with the Ministry of Culture and Tourism Number of Museums Affiliated to the Ministry of Culture and Tourism Number of Museum Visitors Number of Movie Theaters Number of Cinema Audiences Number of Theater Halls Number of Theater Audiences
Industry	Total Number of Initiatives
Agriculture And Livestock	Crop Production Value (thousand TL) Number of Cattle (head) Live Animals Value (thousand TL) Animal Products Value (thousand TL) Number of Ovine (head) Greenhouse Vegetable and Fruit Production (tons) Cereals and Other Herbal Products Production (tons) Total Agricultural Area (hectares)
Tourism	Total number of overnight stays Total Number of Arrivals (person) Number of Foreign Stays Number of Foreign Arrivals (person)
National Accounts	GDP (thousand TL) GDP per Capita ($) GDP per Capita (TL)
Work Life	Employment rate (%) Unemployment rate (%) Average daily earnings (TL) Job satisfaction rate (%)
Civil Participation	Local Administrations Election Rate (%) Political Parties Membership Rate (%) Rate of Union/Association Membership (%)
Income And Wealth	Savings Deposits Per Capita (TL) Percentage of Households in the Middle- and Upper-Income Group (%) Percentage of Households That Do Not Meet Basic Needs (%)
Security	Murder Rate (per one million people) Number of Fatal and Injured Traffic Accidents (per thousand) Rate of Feeling Safe When Outside at Night (%) Satisfaction Rate of Public Security Services (%)
Access To Infrastructure Services	Number of Internet Subscribers (per hundred) Access to Sewerage and Potable Water (%) Airport Access Rate (%) Public Transportation Services Satisfaction Rate (%)
Social Life	Shopping Center Area Per Thousand People (m^2) Social Relations Satisfaction Rate (%) Social Life Satisfaction Rate (%)
Happiness Level	Life Happiness Level (%)

This study uses the Fuzzy C-means clustering algorithm to analyze TUIK data from different perspectives. After an initial cluster analysis is done with all 142 variables, data is further analyzed from 7 different perspectives. The next section provides details about the cluster analysis. Figure 2 summarizes the approach taken in this study.

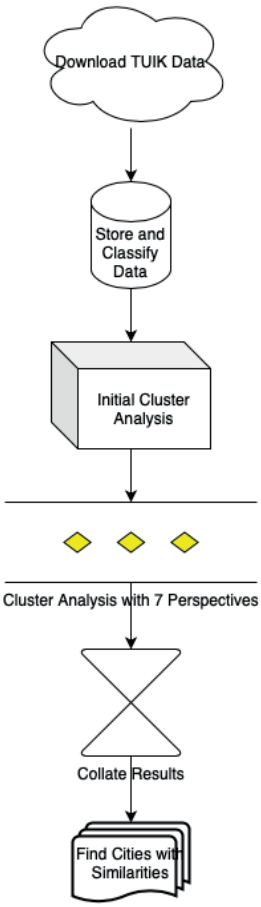


Figure 2. Workflow of the approach used in this study.

4. CLUSTER ANALYSIS

In the analysis section, the major categories in the data were examined in order to find similar ones among the cities. For data analysis, this study uses an enhanced version of the Fuzzy C-means clustering with a variable number of clusters. The study uses clustering indices such as Xie Beni and partition coefficient along with the Fuzzy C-means algorithm's scoring function, to find the most optimal number of clusters. Initially, all 142 variables were used in a general cluster analysis to identify overall similarities between the cities. Next, data is analyzed from 7 perspectives such as "Education", "Population and Settlement", "Livelihood, Income, and Purchasing Power", "Health", "Satisfaction", "Security", "Cultural Opportunities and Utilization" and at the same time, a clustering analysis was performed with the variables. Thus, by working on the data from 8 aspects, similarities between the cities across different categories were considered. Cities are colored based on the clusters they belong to and presented on the maps. The analysis results obtained when all variables are included in the analysis are shown in the table below. While determining the cluster numbers, Xie Beni, Partition Entropy, and Partition Coefficient values were used, different cluster numbers were tried and the number of clusters that gave the best coefficient was presented as the final cluster number in the study.

In the general clustering analysis performed with 142 different variables, 9 clusters were obtained. The provinces and the clusters they belong to are shown on the map in Figure 3. As can be seen, Istanbul, Ankara, Konya, and Izmir are separated from all provinces, and each is represented as a cluster.

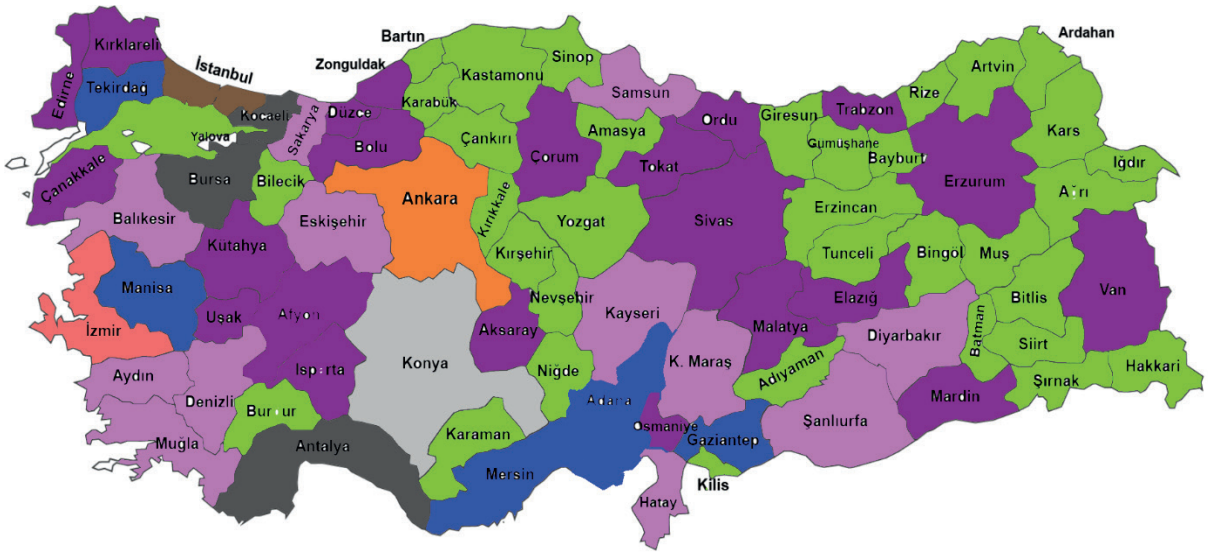


Figure 3. Representation of the clustering study based on all 142 variables on the map.

Cluster 0: Aydın, Balıkesir, Denizli, Diyarbakır, Eskişehir, Hatay, Kahramanmaraş, Kayseri, Muğla, Sakarya, Samsun, Şanlıurfa

Cluster 1: Antalya, Bursa, Kocaeli

Cluster 2: İstanbul

Cluster 3: Adıyaman, Amasya, Ardahan, Artvin, Ağrı, Bartın, Batman, Bayburt, Burdur, Bilecik, Bingöl, Bitlis, Erzincan, Gümüşhane, Giresun, Hakkâri, Iğdır, Karabük, Karaman, Kars, Kastamonu, Kırıkkale, Kırşehir, Kilis, Muş, Nevşehir, Niğde, Rize, Sinop, Siirt, Tunceli, Yalova, Yozgat, Çankırı, Şırnak

Cluster 4: Adana, Gaziantep, Manisa, Mersin, Tekirdağ

Cluster 5: Konya

Cluster 6: Ankara

Cluster 7: İzmir

Cluster 8: Afyonkarahisar, Aksaray, Bolu, Düzce, Edirne, Elâzığ, Erzurum, Isparta, Kırklareli, Kütahya, Malatya, Mardin, Ordu, Osmaniye, Sivas, Tokat, Trabzon, Uşak, Van, Zonguldak, Çanakkale, Çorum

It is observed from the cluster results that the biggest 3 cities in Turkey, Istanbul, Ankara, and İzmir are located in separate clusters. Cluster 0 contains the cities that are developing and have more than 1M residents, although they are geographically separate. Similarly, Cluster 3 contains underdeveloped cities. The initial clustering analysis with all variables revealed that the dataset can be used to distinguish developed and underdeveloped cities. So, as a next step data is analyzed focusing on different perspectives.

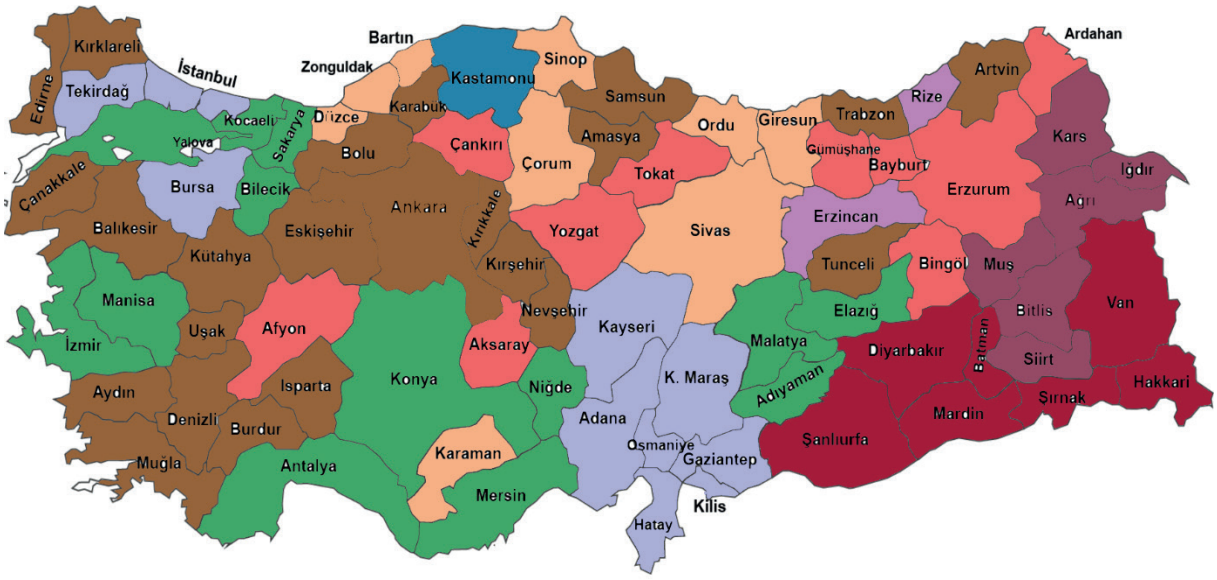


Figure 4. Representation of the cities based on clustering analysis for *education*.

Cluster 0: Amasya, Ankara, Artvin, Aydın, Balıkesir, Bolu, Burdur, Denizli, Edirne, Eskişehir, Isparta, Karabük, Kırıkkale, Kırklareli, Kırşehir, Kütahya, Muğla, Nevşehir, Samsun, Trabzon, Tunceli, Uşak, Çanakkale

Cluster 1: Erzincan, Rize

Cluster 2: Adıyaman, Antalya, Bilecik, Elâzığ, Kocaeli, Konya, Malatya, Manisa, Mersin, Niğde, Sakarya, Yalova, İzmir

Cluster 3: Kastamonu

Cluster 4: Ağrı, Bitlis, Iğdır, Kars, Muş, Siirt

Cluster 5: Bartın, Düzce, Giresun, Karaman, Ordu, Sinop, Sivas, Zonguldak, Çorum

Cluster 6: Batman, Diyarbakır, Hakkâri, Mardin, Van, Şanlıurfa, Şırnak

Cluster 7: Adana, Bursa, Gaziantep, Hatay, Kahramanmaraş, Kayseri, Kilis, Osmaniye, Tekirdağ, İstanbul

Cluster 8: Afyonkarahisar, Aksaray, Ardahan, Bayburt, Bingöl, Erzurum, Gümüşhane, Tokat, Yozgat, Çankırı

There is a resemblance between the clusters in Figure 4 with the clusters created with all 142 variables. Some of the developing cities such as Istanbul, Adana, and Bursa are in Cluster 7 along with their neighbor cities, although they are not developed (e.g., Osmaniye and Kilis). This suggests that the education level of a developed city can impact its neighbors. This can be explained by students traveling or relocating to big neighboring cities for advanced education.

On the contrary, clusters 1 – 5 contain mostly underdeveloped cities with no neighboring big cities. Residents of these small cities do not have easy access to higher education institutions that are generally located in big cities.

Figure 5. Representation of the cities based on clustering analysis for *security*.

Cluster 0: Ankara, Kocaeli, İstanbul, İzmir

Cluster 1: Ağrı, Batman, Hakkâri, Muş, Van, Şırnak

Cluster 2: Burdur, Karaman, Kastamonu, Kırıkkale, Kırşehir

Cluster 3: Afyonkarahisar, Balıkesir, Manisa, Sakarya, Uşak

Cluster 4: Artvin, Giresun, Kütahya, Niğde, Rize, Sinop

Cluster 5: Bursa, Elâzığ, Eskişehir, Samsun, Tekirdağ, Yalova

Cluster 6: Aksaray, Amasya, Karabük, Nevşehir, Tokat

Cluster 7: Aydın, Denizli, Kilis, Muğla, Osmaniye, Çorum

Cluster 8: Kayseri, Konya, Mersin

Cluster 9: Bayburt, Bolu, Bilecik, Erzincan, Gümüşhane, Isparta, Çankırı

Cluster 10: Adıyaman, Bingöl, Iğdır, Mardin, Trabzon, Zonguldak

Cluster 11: Diyarbakır

Cluster 12: Adana, Düzce, Gaziantep, Hatay, Tunceli

Cluster 13: Ardahan, Bitlis, Edirne, Erzurum, Kars, Kırklareli, Malatya, Ordu, Siirt, Şanlıurfa

Cluster 14: Bartın, Kahramanmaraş, Çanakkale

Cluster 15: Sivas, Yozgat

Cluster 16: Antalya

Although there are more clusters than the previous analysis results, the biggest 3 cities (Istanbul, Ankara, and Izmir) are placed in the same cluster. Also, the developed and underdeveloped cities show different security characteristics and are

therefore placed in different clusters. Cities that are in the western and eastern regions are put in separate clusters. Clusters 13 and 14 are the only exceptions to this where Ardahan-Edirne and Kahramanmaraş-Çanakkale are placed in the same cluster. It can be observed that political instability in the eastern regions over the years and the refugee migrations due to the Syrian internal conflict affected Security in the cities in eastern and Southeastern regions of Turkey.

Figure 6. Representation of the cities based on clustering analysis for *population*.

Cluster 0: Aydın, Balıkesir, Denizli, Kahramanmaraş, Kayseri, Manisa, Muğla, Sakarya, Samsun, Tekirdağ, Van

Cluster 1: Adıyaman, Ağrı, Batman, Elâzığ, Kütahya, Osmaniye, Sivas, Tokat, Zonguldak, Çanakkale, Çorum, Şırnak

Cluster 2: İzmir

Cluster 3: Adana, Gaziantep, Kocaeli, Konya, Şanlıurfa

Cluster 4: Afyonkarahisar, Erzurum, Eskişehir, Malatya, Mardin, Ordu, Trabzon

Cluster 5: Ardahan, Artvin, Bartın, Bayburt, Burdur, Bilecik, Bingöl, Erzincan, Gümüşhane, Hakkâri, Iğdır, Karabük, Karaman, Kars, Kırıkkale, Kırşehir, Kilis, Sinop, Tunceli, Yalova, Çankırı

Cluster 6: Bursa

Cluster 7: Ankara, İstanbul

Cluster 8: Aksaray, Amasya, Bolu, Bitlis, Düzce, Edirne, Giresun, Isparta, Kastamonu, Kırklareli, Muş, Nevşehir, Niğde, Rize, Siirt, Uşak, Yozgat

Cluster 9: Antalya

Cluster 10: Diyarbakır, Hatay, Mersin

Not surprisingly, the top 5 biggest cities such as İstanbul, Ankara, İzmir, Bursa, and Antalya are separated from the rest of the cities. Clusters 3 and 10 show the developing cities with more than 1 million residents, and the rest of the cities were placed in clusters based on their geographic regions. These results are on par with the previous clustering results with other perspectives where large and small cities are placed in separate clusters.

Figure 7. Representation of the cities based on clustering analysis for *health services*.

Cluster 0: Amasya, Artvin, Balıkesir, Bartın, Burdur, Bilecik, Giresun, Karabük, Karaman, Kastamonu, Kırklareli, Kütahya, Niğde, Ordu, Sakarya, Sinop, Uşak, Yalova, Çankırı

Cluster 1: Aksaray, Ardahan, Bayburt, Diyarbakır, Gümüşhane, Kahramanmaraş, Kırşehir, Kocaeli, Kilis, Mersin, Nevşehir, Osmaniye, Tekirdağ

Cluster 2: Bolu, Edirne, Elâzığ, Erzurum, Eskişehir, Isparta, Kırıkkale, Konya, Manisa, Samsun, Sivas, Tokat, Trabzon, Zonguldak, İzmir

Cluster 3: Afyonkarahisar, Aydın, Bursa, Denizli, Rize, Yozgat, Çanakkale, Çorum

Cluster 4: Adana, Ankara, Antalya, Düzce, Erzincan, Kars, Kayseri, Malatya, Muğla, Tunceli, İstanbul

Cluster 5: Adıyaman, Ağrı, Batman, Bingöl, Bitlis, Gaziantep, Hakkâri, Hatay, Iğdır, Mardin, Muş, Siirt, Van, Şanlıurfa, Şırnak

These results clearly show that geography has a direct impact on access to health services. Cluster 5 contains cities in the eastern regions only. Diyarbakır and Erzurum are the only 2 exceptions, as these cities are bigger than the others in the eastern regions. Smaller cities in the western regions were expected to be placed in the same cluster with their larger neighbors, as residents of the small cities can easily travel to a large city nearby to receive health services. However, this is not the case: Tekirdağ and Kocaeli are in a different cluster than İstanbul; Aydın and İzmir, Kırşehir and Ankara are also good examples of this case. These results suggest that the health standards in Western cities are different from the ones in Eastern cities. Also, in the Western regions small cities have similar health standards to the large cities. Large cities in the Western regions have better and bigger hospitals and large numbers of health professionals, so residents living in their neighboring cities can also benefit from these services. One can even argue that in the Western regions the 'number of doctors per capita' is higher in small cities.

Figure 8. Representation of the cities based on clustering analysis for *cultural and social activities*.

Cluster 0: Afyonkarahisar, Aksaray, Amasya, Bartın, Burdur, Erzincan, Giresun, Kahramanmaraş, Karaman, Kırşehir, Manisa, Rize, Sakarya, Çanakkale

Cluster 1: Antalya, Konya

Cluster 2: İstanbul

Cluster 3: Batman, Diyarbakır, Mardin, Muş, Osmaniye, Siirt, Van, Şanlıurfa

Cluster 4: Adıyaman, Artvin, Bolu, Bingöl, Gümüşhane, Hakkâri, Karabük, Kırıkkale, Kırklareli, Niğde, Yalova

Cluster 5: Balıkesir, Isparta, Kütahya, Sinop, Uşak

Cluster 6: Bilecik, Bitlis, Elâzığ, Eskişehir, Kars, Kastamonu, Kilis, Zonguldak, Çorum

Cluster 7: Aydın, Bursa, Mersin

Cluster 8: Muğla, Nevşehir

Cluster 9: Adana, Hatay, Kayseri, Kocaeli, Malatya, Samsun

Cluster 10: Ankara, İzmir

Cluster 11: Ardahan, Ağrı, Bayburt, Düzce, Iğdır, Ordu, Tunceli, Yozgat, Çankırı, Şırnak

Cluster 12: Denizli, Edirne, Erzurum, Gaziantep, Sivas, Tekirdağ, Tokat, Trabzon

The biggest 3 cities (İstanbul, Ankara, and İzmir) are isolated from the rest of the cities. This suggests that these cities attract the most cultural events and activities.

The rest of the cities are not very different from each other. Regardless of their sizes and locations, cities in Turkey (except for the top 3) have similar (and probably low) access to cultural activities.

The only exceptions to this are Antalya and Konya: Antalya is the tourism and conferencing center of Turkey, attracting millions of tourists every year. There are shows and cultural activities held throughout the year in the city, mostly for tourists. Similarly, Konya is the hometown of the world-famous philosopher and thinker Rumi. Millions of domestic and international tourists visit Konya to attend Rumi attractions throughout the year.

Figure 9. Representation of the cities based on clustering analysis for *purchasing power*.

Cluster 0: Adıyaman, Ardahan, Ağrı, Bingöl, Bitlis, Gaziantep, Hakkâri, Iğdır, Kars, Kilis, Muş, Van

Cluster 1: Artvin, Balıkesir, Bartın, Burdur, Denizli, Düzce, Edirne, Erzincan, Isparta, Karaman, Kayseri, Kırıkkale, Konya, Kütahya, Manisa, Rize, Uşak

Cluster 2: Adana, Afyonkarahisar, Aksaray, Amasya, Bayburt, Elâzığ, Erzurum, Gümüşhane, Giresun, Hatay, Kahramanmaraş, Kırşehir, Malatya, Mersin, Nevşehir, Niğde, Ordu, Samsun, Sinop, Sivas, Tokat, Trabzon, Tunceli, Yozgat, Çankırı, Çorum

Cluster 3: Aydın, Kastamonu

Cluster 4: Batman, Diyarbakır, Mardin, Osmaniye, Siirt, Şanlıurfa, Şırnak

Cluster 5: Ankara, Antalya, Bolu, Bursa, Bilecik, Eskişehir, Karabük, Kırklareli, Kocaeli, Muğla, Sakarya, Tekirdağ, Yalova, Zonguldak, Çanakkale, İstanbul, İzmir

Not surprisingly, the top 3 cities are located within the same cluster, along with several cities in the western regions. (Cluster 5) On the other hand, all eastern cities are placed in the same clusters (Cluster 0, 2, and 4) with no exceptions. This clustering analysis clearly shows how much the western and eastern regions of Turkey are separated from each other. From the economic perspective, the western and eastern regions are so divided that there are no exceptions in the clustering analysis results.

4.1 SIMILARITY ANALYSIS

As a next step in this study, a dual similarity analysis of the cities was performed. In the clustering analysis results detailed in the previous section, it can be observed that some of the cities are placed in the same cluster frequently by the algorithm. For example, Batman and Şırnak are two neighboring cities with similar population sizes, and they were placed in the same cluster in 7 out of 8 perspectives. While this result can be expected, Bartın and Karaman, two geographically distant cities

were placed in the same cluster 7 times as well. These cities have similar sizes, and both are in the western part of Turkey. These results coincide with the previous ones from the clustering analysis, as the population sizes of the cities and their location in the east-west direction play important role in showing similarities between them.

Table 2 shows the cities that are placed in the same cluster for 6 or more times. Except for the Balıkesir-Uşak couple, all other city groups have similar sizes and are also in the same East-West region of Turkey.

Table 2
List of cities that are placed in the same cluster 6 or more times.

City 1	City 2	Size Comparison	Geographic Location (East / West)
Bartın	Karaman	Small / Small	West / West
Batman	Şırnak	Small / Small	East / East
Aydın	Denizli	Small / Small	West / West
Amasya	Giresun	Small / Small	East / East
Ağrı	Muş	Small / Small	East / East
Ağrı	Şırnak	Small / Small	East / East
Bayburt	Gümüşhane	Small / Small	East / East
Bayburt	Çankırı	Small / Small	East / East
Burdur	Karaman	Small / Small	West / West
Bitlis	Kars	Small / Small	East / East
Karabük	Yalova	Small / Small	West / West
Isparta	Uşak	Small / Small	West / West
Kütahya	Uşak	Small / Small	West / West
Balıkesir	Uşak	Big / Small	West / West

Turkey is divided into 7 geographic regions, where each region consists of roughly 10-15 cities. The central government provides funding and investment to the cities based on their geographic regions (Arslan, 2014) However, clustering analysis performed in this study showed that most of the cities that are in the same geographic regions are not necessarily similar and thus placed in different clusters.

Using the results presented in this study, further analysis may reveal groups of cities that are actually very similar to each other. In future work, analysis should be done for finding city groups of 5 and more and define the similarities in each group.

5. DISCUSSION AND CON

In this study, demographic characteristics of the 81 cities in Turkey have been analyzed for similarities using variables from various categories. These categories are happiness, income and purchasing power, sporting and cultural activities, health and social protection, population and housing, security, and education. Demographic data for cities are obtained from TUIK, and data spans years between 2015 and 2020. Initially, the study performs a clustering analysis done all variables across all categories to illustrate the general demographic profile of the cities in Turkey. Next, cities were analyzed from seven different perspectives, as listed above. Finally, results from all 8-clustering analyses are compared to find city couples that are often placed in the same cluster. This study aims to help the local governments' budget and investment planning, by providing insights into city demographics and highlighting similarities between the cities that are geographically separate.

At each step, cities were segmented into different numbers of clusters, to find the most optimal cluster setup. The fuzzy c-means clustering algorithm, which evaluates cluster probabilities and decides the cluster assignments to maximize probability values is used for this purpose. The largest 3 cities in Turkey; Istanbul, Ankara, and İzmir are often placed in the same cluster and separated from the other cities. These cities not only have higher populations than others but also, they are located in the western regions of the country. Clustering results also show that cities in the western and eastern regions are often placed in separate clusters. While neighboring cities are expected to be placed in the same cluster, this study showed that cities in Turkey are categorized based on their east-west positions: Cities in the west are categorized together although they are not necessarily neighbors. The same outcome can be observed for the cities in the eastern region. These results clearly indicate

that urban planning for cities can be separated at the top level; eastern and western regions have different levels of urbanization and therefore different needs. This study highlights 14 city pairs that are very frequently placed in the same cluster under each category. As expected, these city-pairs are from the same east/west regions of Turkey, although they are not geographically neighbors.

Also, big cities in each region often have a positive impact on their smaller neighbors. If a small city has a large neighbor, it is often placed inside the same cluster as the big cities, instead of the other small ones. While large cities attract more investment from the central government than small ones, they also help develop the smaller cities around them. The reverse is also true – if a small city doesn't have a large neighbor, it will require direct assistance from the central government to improve. This study also highlights the cities and regions that need more attention from the central government for investment planning.

Peer-review: Externally peer-reviewed.
Author Contributions: Conception/Design of Study- Ö.D., K.Ş., G.S.; Data Acquisition- Ö.D., K.Ş., G.S.; Data Analysis/Interpretation- Ö.D., K.Ş., G.S.; Drafting Manuscript- Ö.D., K.Ş., G.S.; Critical Revision of Manuscript- Ö.D., K.Ş., G.S.; Final Approval and Accountability- Ö.D., K.Ş., G.S.; Material and Technical Support- Ö.D., K.Ş., G.S.; Supervision- Ö.D., K.Ş., G.S.
Conflict of Interest: The authors have no conflict of interest to declare.
Grant Support: The authors declared that this study has received no financial support.

REFERENCES

Bulut, H. (2019). Türkiye'deki illerin yaşam endekslerine göre kümelenmesi. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 23(1), 74-82.

Yılancı, A. G. V. (2010). Bulanık Kümeleme Analizi İle Türkiye'deki İllerin Sosyoekonomik Açıdan Sınıflandırılması. Süleyman Demirel Üniversitesi *İktisadi ve İdari Bilimler Fakültesi Dergisi*, 15(3), 453-470.

İncekırık, A., & Altın, E. (2021). Türkiye'deki İllerin Ulaştırma Göstergelerine Göre Clusterleme Analizi Yöntemleriyle Sınıflandırılması. *Manisa Celal Bayar Üniversitesi Sosyal Bilimler Dergisi*, 19(3), 186-206.

Tekin, B. (2015). Temel sağlık göstergeleri açısından Türkiye'deki illerin gruplandırılması: bir Clusterleme analizi uygulaması. *Çankırı Karatekin Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 5(2), 389-416.

Atalay, M. (2019). Clusterleme analizi ile Türkiye'deki illerin turizm verileri açısından incelenmesi. *Ekonomi Maliye İşletme Dergisi*, 2(2), 103-115.

Dinçer, Ş. E. (2006). *Veri madenciliğinde K-means algoritması ve tıp alanında uygulanması* (Master's thesis, Fen Bilimleri Enstitüsü).

Silahtaroğlu, G. (2004). *Veri Madenciliğinde Clusterleme Analizi ve Öğretim Başarısının Değerlendirilmesine İlişkin Bir Uygulama* (Doctoral dissertation, Doktora Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İşletme Anabilim Dalı, İstanbul).

Mining, H. P. D. Powerful Database Support for High Performance Data Mining.

Höppner, F., Klawonn, F., Kruse, R., & Runkler, T. (1999). *Fuzzy Cluster analysis: methods for classification, data analysis and image recognition*. John Wiley & Sons.

Moertini, V. (2002). Introduction to Five DataClustering Algorithms Clustering Algorithm. *Integral*, 7(2).

Salem, S. A., & Nandi, A. K. (2005, September). New assessment criteria for Clustering algorithms. In 2005 IEEE Workshop on Machine Learning for Signal Processing (pp. 285-290). IEEE.

Kruse, R., Borgelt, C., & Nauck, D. (1999, August). Fuzzy data analysis: challenges and perspectives. In FUZZ-IEEE'99. 1999 IEEE International Fuzzy Systems. Conference Proceedings (Cat. No. 99CH36315) (Vol. 3, pp. 1211-1216). IEEE.

Meltem, I. Ş. I. K., & Çamurcu, A. Y. (2011). K-Means Ve Aşırı Küresel C-Means Algorıtmaları İle Belge Madenciliği. *Marmara Fen Bilimleri Dergisi*, 22(1), 1-18.

Ross, T. J. (2005). *Fuzzy logic with engineering applications*. John Wiley & Sons.

Hekim, M., & Orhan, U. (2011). Bulanık C-Means Clusterleme Yöntemine Çıkarımlı Yaklaşım. *İtüdergisi*/D, 10(1).

Atalay, A., & Tortum, A. (2010). Türkiye'deki illerin 1997-2006 yılları arası trafik kazalarına göre Clusterleme analizi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 16(3), 1997-2006.

Kandemir, A. Ş. (2018). Bulanık Clusterleme Analizi İle Türkiye'deki İllerin Konaklama İstatistiklerine Göre Sınıflandırılması. *Seyahat ve Otel İşletmeciliği Dergisi*, 15(3), 657-668.

Atalay, M., & Öztürk, Ş. (2016, September). Türkiye'deki İllerin göç ve işsizlik istatistiklerine göre Clusterlenmesi. In 2nd International Congress on Applied Sciences:"Migration, Poverty and Employment", 23-25 Eylül, Konya/Turkey, Bildiriler Kitabı.

Servi, T., & Erişoğlu, Ü. (2020). Türkiye'deki Şehirlerin Sosyo-Ekonomik Gelişmişlik Düzeylerinin İstatistiksel Analizi. *Al Farabi Uluslararası Sosyal Bilimler Dergisi*, 5(2), 174-186.

Silahtaroğlu, G. (2008). *Data mining concepts and algorithms*. İstanbul: Papatya Publishing.

Silahtaroğlu, G. (2013). Veri madenciliği: Kavram ve algoritmaları. İstanbul: Papatya Publishing.

Arslan, İ. (2014). Türkiye'de Bölgesel Alanda Uygulanan İktisadi Politikalar (Yatırım Teşvikleri-İstihdam Analizi 1980-2006). *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi* , 4 (8) .

ACTA INFOLOGICA 2023;7(1):125-138
dergipark.org.tr/acin

DOI: 10.26650/acin.1222890

RESEARCH ARTICLE

An Intrusion Detection Approach based on the Combination of Oversampling and Undersampling Algorithms

Örneklem Arttırma ve Örneklem Azaltma Algoritmalarının Kombinasyonuna Dayalı Bir Saldırı Tespit Yaklaşımı

Ahmet Okan Arık[1], **Gülsüm Çiğdem Çavdaroğlu**[2]

ABSTRACT

The threat of network intrusion has become much more severe due to the increasing network flow. Therefore, network intrusion detection is one of the most concerned areas of network security. As demand for cybersecurity assurance increases, the requirement for intrusion detection systems to meet current threats is also growing. However, network-based intrusion detection systems have several shortcomings due to the structure of the systems, the nature of the network data, and uncertainty related to future data. The imbalanced class problem is also crucial since it significantly negatively affects classification performance. Although high performance has been achieved in deep learning-based methodologies in recent years, machine learning techniques may also provide high performance in network intrusion detection. This study suggests a new intrusion detection system called ROGONG-IDS (Robust Gradient Boosting - Intrusion Detection System) which has a unique two-stage resampling model to solve the imbalanced class problem that produces high accuracy on the UNSW-NB15 dataset using machine learning techniques. ROGONG-IDS is based on gradient boosting. The system uses Synthetic Minority Over-Sampling Technique (SMOTE) and NearMiss-1 methods to handle the imbalanced class problem. The proposed model's performance on multi-class classification was tested with the UNSW-NB15, and then its robust structure was validated with the NSL-KDD dataset. ROGONG-IDS reached the highest attack detection rate and F1 score in the literature, with a 97.30% detection rate and 97.65% F1 score using the UNSW-NB15 dataset. ROGONG-IDS provides a robust, efficient intrusion detection system for the UNSW-NB15 dataset, which suffered from imbalanced class distribution. The proposed methodology outperforms state-of-the-art and intrusion detection methods.

Keywords: Machine learning, cyber security, intrusion detection system, imbalanced data, gradient boosting

ÖZ

Artan ağ akışı nedeniyle ağa izinsiz giriş tehdidi çok daha şiddetli hale gelmiştir. Bu nedenle, ağ güvenliğinde en çok endişe duyulan alanlardan biri ağ saldırı tespitidir. Siber güvenlik güvencesine olan talep arttıkça mevcut tehditleri karşılamak için saldırı tespit sistemlerine olan gereksinim de artmaktadır. Bununla birlikte, ağ tabanlı saldırı tespit sistemlerinin, sistemlerin yapısı, ağ verilerinin doğası ve gelecekteki verilerle ilgili belirsizlik nedeniyle bazı eksiklikleri vardır. Dengesiz veri problemi de sınıflandırma performansını kötü etkilediği için çok önemlidir. Son yıllarda derin öğrenme tabanlı metodolojilerde yüksek performans elde edilmesine rağmen, makine öğrenme teknikleri de ağ saldırı tespitinde yüksek performans sağlayabilir. Bu çalışma, makine öğrenme tekniklerini kullanarak UNSW-NB15 veri setinde yüksek doğruluk üreten dengesiz sınıf problemini çözmek için benzersiz bir iki aşamalı yeniden örnekleme modeline sahip olan ROGONG-IDS (Robust Gradient Boosting - Saldırı Tespit sistemi) adlı yeni bir saldırı tespit sistemi önermektedir. ROGONG-IDS, gradyan artırmaya dayalıdır. Sistem, dengesiz sınıf problemini çözmek için Sentetik Azınlık Aşırı Örnekleme Tekniği (SMOTE) ve NearMiss-1 yöntemlerini kullanır. Önerilen modelin çok sınıflı sınıflandırma performansı UNSW-NB15 ile test edilmiş, güçlü yapısı NSL-KDD veri seti ile doğrulanmıştır. ROGONG-IDS, UNSW-NB15 veri setini kullanarak %97,30 tespit oranı ve %97,65 F1 skoru ile literatürdeki en yüksek saldırı tespit oranı ve F1 skoruna ulaşmıştır. ROGONG-IDS, dengesiz sınıf dağılımından muzdarip UNSW-NB15 veri kümesi için sağlam, verimli bir saldırı tespit sistemi sağlamaktadır. Önerilen metodoloji, literatürdeki en gelişmiş saldırı tespit metotlarından daha iyi performans göstermektedir.

Anahtar Kelimeler: Makine öğrenmesi, siber güvenlik, saldırı tespit sistemi, dengesiz veri, gradyan artırma

[1](PhD), Istanbul University, Istanbul, Turkiye
[2](Assist. Prof. Dr.), Isik University, Faculty of Economics, Administrative and Social Sciences, Department of Information Technologies

ORCID: A.O.A. 0000-0002-6572-1605; G.Ç.Ç. 0000-0002-4875-4800

Corresponding author:
Ahmet Okan ARIK
Istanbul University, Istanbul, Turkiye
E-mail address: aokanarik@gmail.com

Submitted: 23.12.2022
Revision Requested: 17.02.2023
Last Revision Received: 14.02.2023
Accepted: 17.04.2023
Published Online: 14.06.2023

Citation: Arik, A.O., Cavdaroglu, G.C. (2023). An intrusion detection approach based on the combination of oversampling and undersampling algorithms. *Acta Infologica, 7*(1), 125-138. https://doi.org/10.26650/acin.1222890

1. INTRODUCTION

According to the Mobility Report of Ericsson, mobile network data traffic grew 42 percent between Q3 2020 and Q3 2021. Total monthly mobile network data traffic in Q3 2021 reached around 78EB (Ericsson, 2021). Due to stay-at-home activities, the Covid-19 pandemic resulted in a spike in network traffic from 2019-2020. This increase also varied regionally. The most rapid growth of international internet bandwidth was experienced in Africa, growing at a compound annual rate of 45% between 2017 and 2021. A 38% compound annual rate during the same period was experienced in Oceania (Mauldin, 2021). Hence, the threat of network intrusion has become much more severe. Consequently, network intrusion detection is considered one of the significant concerns in the network domain. As the demand for cybersecurity assurance increases, the requirement for intrusion detection systems (IDS) to meet current threats is also growing. IDS can be divided into three groups according to the collection mechanisms: (1) Network-based IDS (NIDS), (2) Host-based IDS (HIDS), and (3) Hybrid IDS. The main goal of a HIDS is to monitor network traffics in a particular host and analyze the file system, login activities, and currently running processes. On the other hand, NIDS detects any attacks on the hosts of that network. Hybrid IDS models use both of them.

NIDS has several shortcomings due to the structure of the systems, the nature of the network data, and uncertainty related to future data. First, NIDS schemes are occasionally insufficient since they can detect normal/abnormal attacks, but not the exact attack type. Secondly, the up-to-datedness of data sets in which the schemes are tested is crucial because detecting emerging attacks is necessary to develop a scheme that will work with high performance in modern networks. However, most NIDS schemes were tested on outdated data sets. Finally, the imbalanced class problem is also crucial since it has a significant effect on classification performance (Zhang, Huang, Wu, & Li, 2020). Imbalanced network intrusion data makes it hard to detect minority attack classes accurately.

According to the detection techniques used in these systems, IDS can be categorized as (1) Misuse-based IDS and (2) Anomaly-based IDS. Anomaly-based IDS have increasingly attracted attention in recent years since the other types of IDS can identify only known attacks and suffer from the incompetence to detect new attacks. This study proposes a unique two-stage resampling method within the scope of ROGONG-IDS, developed to detect minority class attacks that anomaly-based IDS types have difficulty detecting.

The techniques used in anomaly-based IDS can be divided into (1) Machine Learning methods and (2) Deep Learning methods. *Fig. a* shows these methods and their sub-methods.

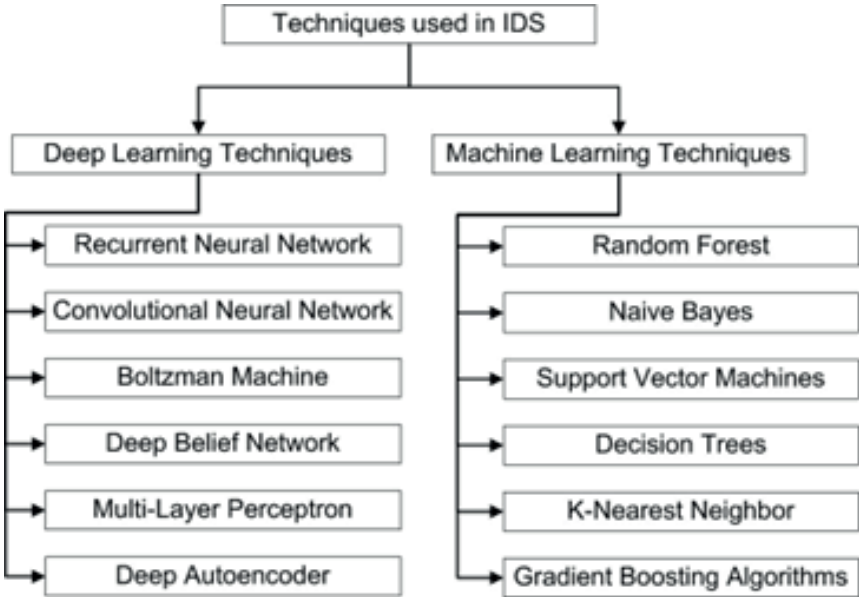


Figure a. Techniques used in anomaly-based IDS

Although high performance has been achieved in methods developed using deep learning techniques in recent years, models that produce results with high accuracy (ACC) in wide-ranging data sets can be developed using machine learning techniques. It was indicated that machine learning techniques, such as support vector machine (SVM), random forest (RF), and decision tree (DT), are insufficient to distinguish normal and abnormal network activities due to the diversification of attack categories and the surge of network traffic (Zhang *et al*., 2020). However, it was observed that the gradient boosting technique used in the presented study produced results with high performance in the UNSW-NB15 dataset, which has a dramatically imbalanced class problem. Consequently, this study presents a new method that produces results with high ACC on imbalanced data sets using machine learning techniques. The presented method currently has the highest performance in the literature.

The contributions of the ROGONG-IDS (Robust Gradient Boosting - Intrusion Detection System) method are as follows:

1. Shortening of testing times: Since ROGONG-IDS provides a highly effective and low-complexity model for the feature selection and classifier method, testing times are shortened. Due to its short testing time, ROGONG-IDS is suitable for real-time network environments.

2. A two-step solution method for the imbalanced class problem: ROGONGIDS offers a unique, robust, and effective two-step solution for the imbalanced class problem.

3. The ROGONG-IDS method has the highest ACC (97.30%) and F_1 score (97.65%) in the literature.

4. The ROGONG-IDS method can be helpful in frequently incorporating gradient boosting methods and resampling studies into IDS development, which are rarely used in IDS models. Therefore, the method triggers the development of IDS in this direction.

The rest of the paper is organized as follows. Section 2 discusses the related work of machine learning and deep learning techniques used in Anomaly-based IDS. Section 3 provides a brief description of the ROGONG-IDS method. Experiment results are presented in Section 4 and discussed in Section 5. Finally, Section 6 concludes the work.

2. RELATED WORK

The techniques used in IDS can be examined under two headings according to the methods used. First-generation techniques were created using Machine Learning methods, and second-generation techniques were created using Deep Learning methods. This section will discuss the techniques available in the literature, and performance measures will be examined.

2.1. Machine learning techniques

Chkirbene, Eltanbouly, Bashendy, AlNaimi, & Erbad (2020) proposed a hybrid approach that combines two machine learning algorithms. The proposed methodology detects possible attacks by performing effective feature selection and classification. They used the RF algorithm to find the essential features, classification, and regression trees (CART) to classify the different attack classes. They tested the approach using the UNSW-NB15 dataset. The ACC is 95.73% for the UNSW-NB15 dataset and 97.03% for the KDD99 dataset.

Injadat, Moubayed, Nassif, & Shami (2021) proposed a novel multi-stage optimized machine learning-based framework to reduce computational complexity and maintain detection performance. They evaluated the framework's performance using the CICIDS 2017 and the UNSW-NB15 datasets. The detection accuracies are over 99% for both datasets.

Bhavani, Rao, & Reddy (2020) proposed a mix of the DT and RF algorithms. The ACC of the DT algorithm is 81.86%, and the F_1 score is 82.00%. On the other hand, the ACC of the RF algorithm is 95.32%, and the F_1 score is 95.00%. According to the results, the RF algorithm is a better method to overcome the over-fitting problem

Kaja, Shaout, & Ma (2019) proposed a two-stage architecture. The architecture based on machine learning algorithms uses K-Means to detect attacks in the first stage; it uses supervised learning to classify such attacks and eliminate the number of false positives. The ACC of their approach is 99.95%, and the F_1 score is 99.99%.

Belouch, El Hadaj, & Idhammad (2018) presented a study that evaluates the performances of several machine learning methods, such as SVM, Naive Bayes (NB), DT, and RF. The evaluation criteria used in this study are ACC, building time, and prediction time. They used the UNSW-NB15 dataset to evaluate the mentioned methods. The detection rate (DR), also known as recall, and precision of the RF method considered the best method according to this study, are 97.49% and 93.53%, respectively.

2.2. Deep learning techniques

Sumaiya Thaseen, Saira Banu, Lavanya, Rukunuddin Ghalib, & Abhishek (2021) proposed a new methodology based on deep learning techniques. The methodology includes a correlation-based feature selection phase integrated with neural network for identifying anomalies. They tested the approach on the UNSW-NB15 and KDD99 datasets. The performance results show that their approach is superior in ACC, sensitivity, and specificity compared to some state-of-the-art techniques. The overall ACC is 96.44% for the UNSW-NB15 dataset.

The model proposed by Liu, Gao, & Hu (2021) addresses the imbalanced data problem. An ensemble model is used to solve the imbalanced data problem in the presented study. This model uses ADASYN for oversampling and LightGBM for classification. After normalization, the authors evaluated the model using the KDD, UNSW-NB15, and CICIDS2017 datasets. The model, which offers a more prosperous and shorter training time than other IDS models, reached 85.89% ACC on the UNSW-NB15 dataset.

Mulyanto, Faisal, Prakosa, & Leu (2021) proposed a cost-sensitive neural network based on focal loss (FL-NIDS) to overcome the imbalanced data problem. To evaluate the UNSW-NB15, NSL-KDD, and Bot-IoT intrusion detection datasets, they applied this system using a convolutional neural network (CNN). The ACC score does not reflect the DR of the minority classes. They evaluated the approach using the F_1 score. For the UNSW-NB15 dataset, the CNN-SMOTE model reached a 36% F_1 score, while FL-NIDS reached a 39% F_1 score.

Zhang *et al.* (2020) proposed a flow-based IDS model. They developed a new class imbalance processing technology for large-scale data and combined it with CNN. Their methodology's DR, precision, and F_1 score on the UNSW-NB15 dataset were 96.54%, 98.30%, and 97.26%, respectively. This study is currently the one that produces results with the highest performance in the literature.

Andresini, Appice, Mauro, Loglisci, & Malerba (2020) proposed a multi-channel deep learning method called MINDFUL. It combines an unsupervised approach with a supervised one. The unsupervised one is for multi-channel feature construction. This phase is based on two encoder neural networks. The supervised approach is for exploiting cross-channel feature correlations. They have tested the method on the KDDCUP99Test, UNSW-NB15Test, and CICIDS2017Test datasets. The performance criterion for the UNSW-NB15 dataset is 93.40% ACC and 95.29% F_1 score.

Khan, Gumaei, Derhab, & Hussain (2019) proposed a two-stage IDS model. The model first detects whether the network packets are normal or abnormal based on the probability score generated by the stacked auto-encoder (AE). The attacks are then classified using the Softmax classifier. Therefore, the model can also classify unlabeled data. The model, which was evaluated with different algorithms, reached 89.13% ACC, 0.74 false alarm rate (FAR) on the UNSW-NB15 dataset.

Yang, Zheng, Wu, & Yang (2019) combined an improved conditional variational AE with a DNN. The trained encoder was used to automatically reduce data dimension and initialize the weight of DNN hidden layers. This way, the DNN can quickly achieve global optimization through backpropagation and fine-tuning. They used the NSL-KDD and UNSW-NB15 datasets to evaluate the performance of their model. According to the results, the proposed method shows better performance metrics than the nine intrusion detection methods. The ACC, DR, precision, F_1 score, and FPR metrics are 85.97%, 77.43%, 97.39%, 86.27%, and 2.74, respectively.

Zhang *et al.* (2018) proposed a new network intrusion detection scheme based on deep learning techniques. Reducing the feature dimensionality is crucial for network intrusion systems. Therefore, the proposed scheme used a denoising auto-encoder (DAE) with a weighted loss function for feature selection. The selected data is classified using a compact multi-layer perceptron

(MLP) to identify intrusions. The proposed scheme was tested on the UNSW-NB15 dataset. The DR, precision, and F_1 score of the proposed scheme on the UNSW-NB15 dataset were 98.80%, 95.98%, and 95.2%, respectively.

Mulyanto *et al.* (2021) utilized deep neural networks (DNNs) to predict the attacks on network IDS. They compared the approach with Ada Boost, DT, K-Nearest neighborhood (K-NN), linear regression, NB, RF, and SVM methods. As a result, one-layer DNN architecture achieved 92.9% ACC and 95.4% F_1 score and outperformed traditional machine learning techniques.

Naseer *et al.* (2018) developed anomaly detection models based on different DNN structures, including CNN, AEs, and recurrent neural networks (RNNs). They trained the models on the NSL-KDD training dataset and evaluated them using the NSL-KDDTest+ and NSL-KDDTest21 datasets. According to the study results, CNN reached 85% ACC, and LSTM reached 89% ACC.

Yin, Zhu, Fei, & He (2017) proposed a deep learning approach for intrusion detection using RNNs (RNN-IDS). They studied the model's performance in binary and multi-class classification and compared it with J48, artificial neural network (ANN), RF, SVM, and other machine learning methods. They tested the model on the NSL-KDD dataset. Their model reached 81.29% ACC with RNN and 78.10% ACC with MLP.

Table I
Briefly provides some of the advanced anomaly-based IDS models' performances on the UNSW-NB15 dataset in terms of development and testing

Author	Method	Classification Type	Accuracy (%)	Detection Rate (%)	F_1 Score (%)	Precision (%)
Chkirbene *et al*. (2020)	ML	Multiclass	95.73	-	-	-
Injadat *et al.* (2021)	ML	Binary	99	-	-	-
Belouch *et al.* (2018)	ML	Binary	97.49	93.53	-	-
Sumaiya Thaseen *et al.* (2021)	DL	Multiclass	96.44	-	-	-
Zhang *et al.* (2020)	DL	Multiclass	96.54	96.54	98.30	97.26
Andresini et al. (2020)	DL	Binary	93.40	-	95.29	-
Khan *et al.* (2019)	DL	Multiclass	89.13	-	-	-
Yang *et al.* (2019)	DL	Multiclass	85.97	77.43	86.27	97.39
Zhang *et al.* (2018)	DL	Binary	-	98.80	95.2	95.98
Mulyanto *et al.* (2021)	DL	Multiclass	92.9	95.4	-	-

Table i. Advanced anomaly-based IDS models use the UNSW-NB15 dataset in the literature.
(ML: machine learning, DL: deep learning)

3. METHOD

The ROGONG-IDS method consists of three modules, as shown in *Fig. b*. Data has been made suitable for modeling with the operations performed in the data preprocessing module. The processing of categorical data is provided with one-hot encoding and label encoding applied in this module. Feature selection was made to improve ACC by reducing the training time of the model and removing redundant features. Data standardization has been applied to examine different types of measurable features in a common standard. The module that handles the imbalanced class problem includes resampling operations to ensure class balance. The resampling method has two stages: the NearMiss-1 method for undersampling and

the SMOTE method for oversampling. The resampling method is the most critical method that increases the model's ACC. Finally, several gradient boosting method tests to decide the classification decision method. Hyperparameters of the selected method are optimized using Bayesian optimization.

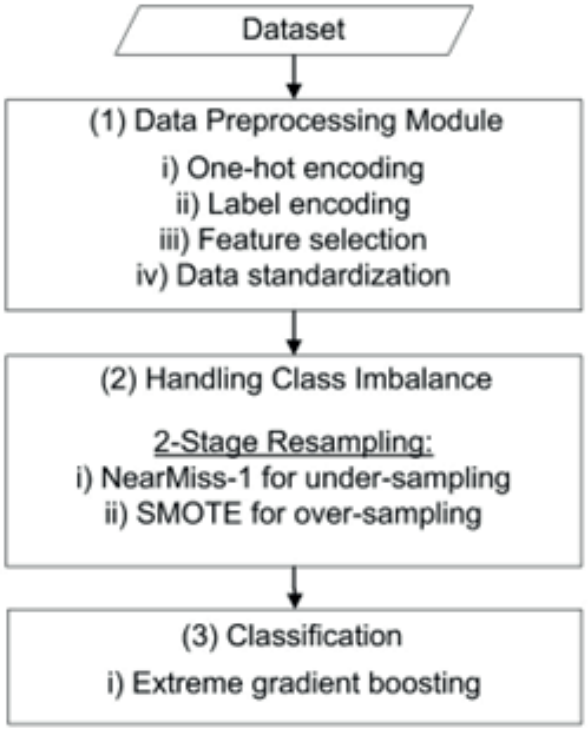


Figure b. Flow of the proposed method

3.1.1. Data preprocessing

The data preprocessing module was carried out feature selection, one-hot encoding, label encoding, and data standardization processes. The DAE developed from Zhang *et al.* (2018), which reduces the feature size by limiting the number of critical features, was used for feature selection. This process selected twelve properties: Dtcpb, Stcpb, Service, Dload, Dmeansz, Service dns, Smeansz, Sload, Trans depth, Sttl, Service ftp-data, and Ct ftp. The UNSW-NB15 dataset has three nominal properties:" proto"," state", and" service". These attributes have 135, 16, and 14 different values, respectively. In order to process these features with machine learning algorithms, the one-hot encoding technique is used while maintaining the irregular relationship. With this process, the data set's features increased from 47 to 208. In addition, label encoding was applied to the target feature attack class. Finally, features were standardized to ensure that the estimators equally weighted numerical features of different units and scales. Z Score Scaling is used for standardization in the study. This method sets the mean value to 0 and the standard deviation equal to 1 for each feature, making the data have comparable scales.

3.1.2. Handling imbalanced class problem

The classes that account for a significant part of the data set are named majority classes, and the classes that account for the minor are named minority classes. The data set of skewed class balances is called the imbalanced data set. This is because classifiers are highly sensitive to the majority and less sensitive to the minority classes. Imbalanced classes need to be balanced to develop high-accuracy intrusion, detection models. The most typical strategy for balancing class distributions is using different resampling methods (Haibo He & Yunqian Ma, 2013). These are undersampling and oversampling methods. Undersampling aims to reduce the number of majority class observations at a specified rate or number. On the other hand, oversampling aims to increase the number of observations of minority classes at a determined rate or number (Chawla, Bowyer, Hall, & Kegelmeyer, 2002; Haibo He & Yunqian Ma, 2013).

The UNSW-NB15 dataset contains highly imbalanced class distribution. Among the 2,000,054 million samples, there are two classes with less than 2,000 instances. Undersampling methods cause information loss by reducing the number of observations in the majority class, thus reducing the representativeness of classes (Demidova & Klyueva, 2017). Studies indicate that oversampling methods work better than undersampling methods in handling the problem of imbalanced data because they do not cause data loss. However, there are also disadvantages to applying only oversampling methods while achieving class balance. It can significantly increase the data, thus increasing the computational cost. Another disadvantage is that it may lead to an overfitting problem. Accordingly, oversampling for these minority classes or undersampling for majority classes is insufficient to solve the imbalanced data problem. Therefore, the ROGOND-IDS model uses a method that recommends using oversampling and undersampling methods together to overcome this challenge. As shown in *Table II*, 10 different undersampling methods were tried to find the proper one that produced the most successful result. The NearMiss-1 undersampling method produced the best results with oversampling method SMOTE. NearMiss-1 (Yen & Lee, 2006) selects majority class observations close to some minority class observations. Class balance is achieved by calculating the minimum distance between the majority class observation and the three nearest minority class instances in this undersampling method. On the other hand, SMOTE (Chawla *et al*., 2002) is an oversampling method used in generating minority class samples. It generates a new observation based on the similarity of a minority class instance and its nearest neighbor. Depending on the amount of oversampling sample required, neighbors are selected from the k nearest neighbors using the k-NN algorithm. This means that observations similar to existing minority class examples are generated (Demidova & Klyueva, 2017). These samples generated with SMOTE prevent overfitting and improve the classifier's performance.

Using this 2-stage resampling method, the data set is resampled with all classes containing an equal number of samples $\boldsymbol{Iresample = int}\left(\frac{\boldsymbol{number\ of\ samples}}{\boldsymbol{number\ of\ classes}}\right)$. In this way, imbalanced class distribution is prevented.

Table ii
Studied undersampling methods (ACC: accuracy, DR: detection rate).

Undersampling Method	Accuracy (%)	F_1 Score (%)	Detection Rate (%)	Algorithm	Oversampling Method
AIIKNN	96.05	96.77	96.05	XGBoost	SMOTE
Edited Nearest Neighbours	96.26	96.89	96.26	XGBoost	SMOTE
Repeated Edited Nearest Neighbours	96.08	96.78	96.08	XGBoost	SMOTE
Instance Hardness Threshold	94.20	95.00	94.20	XGBoost	SMOTE
NearMiss (v1)	96.49	97.10	96.49	XGBoost	SMOTE
NearMiss (v3)	95.45	96.41	95.45	XGBoost	SMOTE
Neighbourhood Cleaning Rule	96.03	96.64	96.03	XGBoost	SMOTE
Random Undersampling	96.22	96.97	96.22	XGBoost	SMOTE
Tomek Link	96.24	96.89	96.24	XGBoost	SMOTE
One Sided Selection	95.44	96.41	95.44	XGBoost	SMOTE

Algorithm 1 shows the pseudocode of ROGONG-IDS, which handles the imbalanced class problem.

Input:
Training set $D = \{D_i,\ i=1,2,...,C\}$;
C=the total number of classes;
$|D| = N$;*#the total number of samples*
Output:
A balanced training set D';
$I_{resample} = int(N/C)$
for $i \leftarrow 1$ to C **do**
  if $|D_i| < I_{resample}$ **then**
    $D_i' = SMOTE(D_i, I_{resample})$ *#Generating new samples to minority class*
  end if
  if $|D_i| > I_{resample}$ **then**
    $D_i' = NearMiss(D_i, I_{resample})$ *#Removing values from majority class*
  end if
end for
return D'

Algorithm 1. Two-stage resampling method of ROGONG-IDS

3.1.3. Classification decision: Extreme Gradient Boosting (XGBoost)

The XGBoost algorithm is a supervised learning method in machine learning and aims to turn weak learners into strong learners with ensemble learning (Chen & Guestrin, 2016). It is an improved version of the gradient boosting method for DTs. This algorithm aims to provide scalability in tree boosting systems, ensure efficient use of computational resources, and improve the model's performance in classification-regression problems. In the implementation phase, the initial leaf is created, then new trees are created based on the prediction errors. This continues until the number of decision trees that can be provided as hyperparameters or until the development in the model stops. Its difference from other gradient boosting methods tested for the classifier model within the scope of this study, such as GBM or LightGBM, is that it is suitable for parallel processing and is tolerant of datasets with missing data. Since it consists of many hyperparameters, the optimization phase is essential in its application for ideal hyperparameter values.

3.2. Dataset

The UNSW-NB15 dataset was used in this study. The Intelligent Security Group collected this dataset at the Australian Centre for Cyber Security (Moustafa & Slay, 2015). The research group combined the current standard network and synthetic attack data to generate this dataset. The network flow samples were stored as vectors of 49 attributes, of which two are binary, three are categorical, 37 are numerical input attributes, and 1 class attribute. It is a comprehensive dataset representing a modern network with these features.

On the other hand, there is a high level of imbalanced class problem in the dataset. 87.35% of the dataset is regular traffic, and only 12.65% is attack traffic. In the presented study, the dataset was divided into training and testing at a ratio of 7:3. *Table III* shows the attack class distributions in the dataset in detail.

Table iii
Attack class distributions

Class	Training set size	Test set size	Total
Analysis	1,874	803	2,677
Backdoor	1,630	699	2,329
DoS	11,447	4,906	16,353
Exploits	31,167	13,358	44,525
Fuzzers	16,972	7,274	24,246
Generic	150,837	64,644	215,481
Normal	1,553,134	665,630	2,218,764
Reconnaissance	9,791	4,196	13,987
Shellcode	1,058	453	1,511
Worms	122	52	174
Total (10 classes)	1,778,032	762,015	2,540,047

4. EXPERIMENTAL ANALYSIS

We used the UNSW-NB15 dataset to measure the overall performance of the ROGONG-IDS method. *Table IV* represents the system environment parameters used in this study.

Table iv
Test environment

Parameter	Environment / Version
Operating System	macOS Monterey
CPU	1,4 GHz Quad-Core Intel Core i5
GPU	Intel Iris Plus Graphics 645
Memory	16GB

4.1. Evaluation metrics

We use ACC, DR, FAR, F_1 score, and precision indicators, which are commonly used in class imbalance systems. Samples corresponding to the attack are considered positive; other samples are considered negative. The meanings of the metrics are listed below:

ACC (Accuracy): the percentage of correctly classified samples among all samples.

DR (Detection Rate): the rate of correctly predicted positive samples.

FAR (False Alarm Rate): the proportion of negative samples incorrectly evaluated as positive.

Precision: how many samples are predicted to be positive are positivesamples.

F_1 Score: the harmonic average of precision and DR parameters.

When applying multi-class classification, each class must be calculated using a weighted average method based on the number of samples in the category to understand the detection performance of the model on unbalanced data. Equations 1 - 5 represents the formula for the metrics used (TP: true positive, TN: true negative, FP: false positive, FN: false negative).

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$DR = \frac{TP}{TP + FN} \quad (2)$$

$$FAR = \frac{FP}{FP + TN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$F_1\ Score = \frac{2 * DR * Precision}{DR + Precision} \quad (5)$$

TP/FP and TN/FN are the numbers of the correctly and incorrectly predicted samples, respectively.

4.2. Hyper-parameter optimization

Hyperparameters can improve the performance of the model's learning process. It may be possible to reach the best performance of the model in the shortest time by adjusting the hyperparameters. Therefore, choosing the optimization method that will improve model ACC with the least time and power cost is essential. Grid search, one of the frequently used optimization methods, is a brute-force technique (Putatunda & Rama, 2018). It uses manually created subsets to optimize hyperparameters (Schaer, Müller, & Depeursinge, 2016). Although it is a simple method, increasing the number of hyperparameters increases the computational cost exponentially. It is reliable for low-dimensional spaces (Bergstra & Bengio, 2012). On the other hand, random search aims to tune hyperparameters by selecting random points in the search space (Bergstra, Bardenet, Bengio, & Kégl, 2011). It is not suitable for models with many hyperparameters like Grid search. Therefore, these two techniques, which are frequently used, are pricey for models with many hyperparameters. In this context, studies show that Hyperout outperforms Random search and Grid search in terms of ACC and time in optimizing the hyperparameters of the Extreme gradient boosting model and different machine learning models (Bergstra, Komer, Eliasmith, Yamins, & Cox, 2015; Putatunda & Rama, 2018). ROGONG-IDS uses Distributed Asynchronous Hyperparameter Optimization (Hyperopt) for hyperparameter tuning. Hyperopt, identified as a black box optimization technique (Klein, Falkner, Bartels, Hennig, & Hutter, 2017), was developed to automate hyperparameter optimization based on Bayesian optimization. Hyperopt uses Bayesian optimization to define and narrow the search space and maximize the probability function. ROGONG-IDS model ACC increased from 96.49% to 97.30% after using Hyperopt within a reasonable time. *Table V* shows the default XGBoost hyperparameters and the final hyperparameter values used after optimization.

Table v
XGBoost hyperparameters before and after from Bayesian optimization

Parameter	Value Before Optimization	Value After Optimization
Learning Rate	0.3	0.5
Number of Estimators	100	5,000
Max Depth	6	36
Colsample Bytree	1	0.61
Min Child Weight	1	4
Subsample	1	0,9

4.3. Multi-class classification

Table VI represents the DR metrics for each class. Although the ROGONG-IDS essentially uses the XGBoost algorithm, other gradient boost-based algorithms have also been tried in processing the method with the UNSW-NB15 dataset. According to the metrics shown in *Table VI*, ROGONG-IDS with XGBoost achieves the best overall performance in terms of DR, ACC, and F_1 score. These metrics are 97.30%, 98.16%, and 97.65%, respectively.

Table vi
Multi-class classification perfomance comparison between LightGBM, GBM, and XGBoost

Class	LightGBM	GBM	RXGBoost
Analysis	0.84	0.67	0.31
Backdoor	0.23	0.11	0.26
DoS	0.06	0.13	0.47
Exploits	0.46	0.48	0.54

Fuzzers	0.66	0.73	0.70
Generic	0.97	0.97	0.98
Normal	0.99	0.99	0.99
Reconnaissance	0.81	0.81	0.77
Shellcode	0.88	0.55	0.53
Worms	0.83	0	0.83
DR (%)	96.55	96.26	97.30
Accuracy (%)	96.55	96.26	97.30
Precision (%)	98.30	97.91	98.16
F_1 Score (%)	97.18	96.91	97.65
Train-Time (s)	15.08	4,336.71	205.27
Test-Time (s)	2.2	0.73	0.82

Table VII compares advanced IDS methods in the literature and the ROGONG-IDS method. With the ROGONG-IDS method, the DR metric has been improved for many classes. However, the DR value for the three classes remained below 50%. These classes are ”Analysis”, “Backdoor”, and “DoS” classes. When we look at the test times, it is seen that ROGONG-IDS shows the best performance in the literature. The test time is 8 seconds in the SGM method, while only 0.81 seconds in the ROGONG-IDS method.

Table vii
Comparison multi-class classification results with advanced methods on the UNSW-NB15 dataset

Class	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**
Analysis	0.27	0.01	0	0.15	-	0.31
Backdoor	0.51	0	0.6	0.21	-	0.26
DoS	0.39	0	0.18	0.80	-	0.47
Exploits	0.45	0.57	0.86	0.71	-	0.54
Fuzzers	0.67	0.40	0.53	0.35	-	0.70
Generic	0.97	0.61	0.97	0.96	-	0.98
Normal	0.98	0.82	0.80	0.81	-	0.99
Reconnaissance	0.82	0.24	0.79	0.80	-	0.77
Shellcode	0.88	0,00	0.51	0.92	-	0.53
Worms	0.83	0,00	0.59	0.79	-	0.83
DR (%)	96.54	63.27	78.65	95.68	-	97.30
Accuracy (%)	96.54	89.13	78.65	89.08	85.89	97.30
Precision (%)	98.30	89.13	78.65	86.05	-	98.16
F_1 Score (%)	97.26	90.85	78.65	90.61	-	97.65
FAR	-	-	0.11	-	0.6	0.51
Train-Time (s)	47.22	-	-	-	-	205.27
Test-Time (s)	8.26	-	-	-	-	0.81

(M1: SGM-CNN (Zhang *et al.,* 2020), M2: Two stage – DL (Khan *et al*., 2019), M3: Hybrid Machine Learning (Chkirbene *et al.,* 2020), M4: ICVAE-DNN (Yang *et al*., 2019), M5: ADASYN and LightGBM (Liu *et al*., 2021), M6: ROGONG-IDS)

The validity of the robust structure of the ROGONG-IDS model, the performance of which was tested with UNSW-NB15, was evaluated with the NSL-KDD dataset used to assess many attack detection models in the literature (Tavallaee, 2009). The categorical features of the NSL-KDD training dataset, which includes different types of cyber attacks, were digitized with one-hot encoding, and the attack types were mapped in the data preprocessing stage. The classification results obtained using the resampling approach proposed within the scope of ROGONG-IDS confirm the robust structure of the model. Classification results are presented in *Table VIII*.

Table viii
Multi-class classification results on the NSL-KDD dataset

Class	ROGONG-IDS
Normal	0.95
DoS	0.95
Probe	0.99
R2L	0.39
U2R	0.10
DR (%)	94.31
Accuracy (%)	94.31
Precision (%)	96.67
F_1 Score (%)	95.23
FAR	0.0002
Train-Time (s)	48.41
Test-Time (s)	0.19

5. DISCUSSION

Experimental results show that the ROGONG-IDS method significantly improves the DR metric. The source of this improvement is using a two-method imbalance data module with XGBoost in the ROGONG-IDS method. According to the experimental analysis results, the XGBoost algorithm produces more successful results than other methods (GBM, LightGBM). XGBoost provided a higher DR value than the other two classifiers in attack types "Backdoor", "DoS", "Exploits", "Generic", "Normal", and "Worms". When examined in general, it produces more successful results than the other two algorithms based on DR, ACC, F_1 score, and test time metrics. When comparing the ROGONG-IDS method with other state-of-the-art methods, it is seen that ROGONG-IDS is the most successful IDS model in the literature in terms of DR, ACC, F_1 score, and test time. Therefore, it outperforms state-of-the-art intrusion detection methods.

The ROGONG-IDS method has been tested on two different data sets in the literature and has produced successful results in both data sets. Therefore, the method is considered to be robust.

The FAR value of the ROGONG-IDS method was observed to be lower when compared to the FAR values of similar techniques in the literature. However, ROGONG-IDS has a higher classification accuracy than these methods, which makes the approach valuable. In future studies, it is planned to improve the FAR value by keeping the classification accuracy high.

6. CONCLUSION

There are many problems in IDS that are difficult to solve. One of these problems is the datasets used in the evaluation phase. In the evaluation phase of the presented method, the UNSW-NB15 dataset, which includes the most up-to-date attack types and offers many different network parameters, was used to develop a method suitable for modern network environments. However, due to the dynamic nature of the field, it is crucial to keep the datasets up-to-date. Another problem is the case of imbalanced class. In network intrusion systems datasets, attack data items are less frequent than normal data items. This leads to an imbalance between the classes in the dataset, known as the imbalanced class problem in the literature. Increasing the data size to overcome this problem also causes an increase in the computing power and time required for data processing.

According to the evaluation findings of the proposed model, the XGBoost algorithm is more successful than other methods (GBM, LightGBM). The ROGONG-IDS model was compared with five advanced IDS models in the literature during the evaluation phase. The model's DR, ACC, and F_1 score metrics were obtained as 97.30%, 97.30%, and 97.65%, respectively. These results prove that the ROGONG-IDS model outperforms the state-of-the-art methods. On the other hand, the ROGONG-IDS model has a fast testing time (0.81s). As a result, ROGONG-IDS is an efficient solution for real-time intrusion detection applications, delivering high success quickly. The ROGONG-IDS model could therefore be applied to areas where streaming data is imbalanced.

Appendix A. Source Code: The source codes written in Python are provided in datastd-dev/Github (2021).

Peer-review: Externally peer-reviewed.
Author Contributions: Conception/Design of Study- A.O.A., G.Ç.Ç.; Data Acquisition- A.O.A.; Data Analysis/Interpretation- A.O.A.; Drafting Manuscript- A.O.A.; Critical Revision of Manuscript- A.O.A., G.Ç.Ç.; Final Approval and Accountability- A.O.A., G.Ç.Ç.
Conflict of Interest: The authors have no conflict of interest to declare.
Grant Support: The authors declared that this study has received no financial support.

REFERENCES

Andresini, G., Appice, A., Mauro, N. D., Loglisci, C., & Malerba, D. (2020). Multi-Channel Deep Feature Learning for Intrusion Detection. *IEEE Access*, *8*, 53346–53359. https://doi.org/10.1109/ACCESS.2020.2980937

Belouch, M., El Hadaj, S., & Idhammad, M. (2018). Performance evaluation of intrusion detection based on machine learning using Apache Spark. *Procedia Computer Science*, *127*, 1–6. https://doi.org/10.1016/j.procs.2018.01.091

Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for Hyper-Parameter Optimization. *Advances in Neural Information Processing Systems*, *24*. Curran Associates, Inc. Retrieved from https://papers.nips.cc/paper/2011/hash/86e8f7ab32cfd12577bc2619bc635690-Abstract.html

Bergstra, J., & Bengio, Y. (2012). *Random Search for Hyper-Parameter Optimization*.

Bergstra, J., Komer, B., Eliasmith, C., Yamins, D., & Cox, D. D. (2015). Hyperopt: A Python library for model selection and hyperparameter optimization. *Computational Science & Discovery*, *8*(1), 014008. https://doi.org/10.1088/1749-4699/8/1/014008

Bhavani, T. T., Rao, M. K., & Reddy, A. M. (2020). *Network Intrusion Detection System Using Random Forest and Decision Tree Machine Learning Techniques*. *1045*, 637–643. https://doi.org/10.1007/978-981-15-0029-9_50

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Chkirbene, Z., Eltanbouly, S., Bashendy, M., AlNaimi, N., & Erbad, A. (2020). Hybrid Machine Learning for Network Anomaly Intrusion Detection. *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, 163–170. https://doi.org/10.1109/ICIoT48696.2020.9089575

datastd-dev/Github. (2021). ROGONG-IDS/GitHub. Retrieved December 21, 2022, from https://github.com/datastd-dev/ROGONG-IDS (Original work published December 15, 2021)

Demidova, L., & Klyueva, I. (2017). SVM classification: Optimization with the SMOTE algorithm for the class imbalance problem. *2017 6th Mediterranean Conference on Embedded Computing (MECO)*, 1–4. https://doi.org/10.1109/MECO.2017.7977136

Ericsson. (2021). *Ericsson Mobility Report November 2021.*

Haibo He & Yunqian Ma. (2013). Imbalanced Learning: Foundations, Algorithms, and Applications | Wiley. Retrieved December 21, 2022, from Wiley.comwebsite: https://www.wiley.com/en-us/Imbalanced+Learning%3A+Foundations%2C+Algorithms%2C+and+Applications-p-9781118074626

Injadat, M., Moubayed, A., Nassif, A. B., & Shami, A. (2021). Multi-Stage Optimized Machine Learning Framework for Network Intrusion Detection. *IEEE Transactions on Network and Service Management*, *18*(2), 1803–1816. https://doi.org/10.1109/TNSM.2020.3014929

Kaja, N., Shaout, A., & Ma, D. (2019). An intelligent intrusion detection system. *Applied Intelligence*, *49*, 3235–3247. https://doi.org/10.1007/s10489-019 01436-1

Khan, F. A., Gumaei, A., Derhab, A., & Hussain, A. (2019). A Novel Two-Stage Deep Learning Model for Efficient Network Intrusion Detection. *IEEE Access*, *7*, 30373–30385. https://doi.org/10.1109/ACCESS.2019.2899721

Klein, A., Falkner, S., Bartels, S., Hennig, P., & Hutter, F. (2017, March 7). *Fast Bayesian Optimization of Machine Learning Hyperparameters on Large Datasets*. arXiv. https://doi.org/10.48550/arXiv.1605.07079

Liu, J., Gao, Y., & Hu, F. (2021). A fast network intrusion detection system using adaptive synthetic oversampling and LightGBM. *Computers & Security*, *106*, 102289. https://doi.org/10.1016/j.cose.2021.102289

Mauldin, A. (2021). *Global Internet Traffic and Capacity Return to Regularly Scheduled Programming*. Retrieved from https://blog.telegeography.com/internet-traffic-and-capacity-return-to-their-regularly-scheduled-programming

Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). *2015 Military Communications and Information Systems Conference (MilCIS)*, 1–6. https://doi.org/10.1109/MilCIS.2015.7348942

Mulyanto, M., Faisal, M., Prakosa, S. W., & Leu, J.-S. (2021). Effectiveness of Focal Loss for Minority Classification in Network Intrusion Detection Systems. *Symmetry*, *13*(1), 4. https://doi.org/10.3390/sym13010004

Naseer, S., Saleem, Y., Khalid, S., Bashir, M. K., Han, J., Iqbal, M. M., & Han, K. (2018). Enhanced Network Anomaly Detection Based on Deep Neural Networks. *IEEE Access*, *6*, 48231–48246. https://doi.org/10.1109/ACCESS.2018.2863036

Putatunda, S., & Rama, K. (2018). A Comparative Analysis of Hyperopt as Against Other Approaches for Hyper-Parameter Optimization of XGBoost. *Proceedings of the 2018 International Conference on Signal Processing and Machine Learning*, 6–10. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3297067.3297080

Schaer, R., Müller, H., & Depeursinge, A. (2016). Optimized Distributed Hyperparameter Search and Simulation for Lung Texture Classification in CT

Using Hadoop. *Journal of Imaging*, *2*(2), 19. https://doi.org/10.3390/jimaging2020019

Sumaiya Thaseen, I., Saira Banu, J., Lavanya, K., Rukunuddin Ghalib, M., & Abhishek, K. (2021). An integrated intrusion detection system usin correlation-based attribute selection and artificial neural network. *Transactions on Emerging Telecommunications Technologies*, *32*(2), e4014. https://doi.org/10.1002/ett.4014

Tavallaee, M., Bagheri, E., Lu, W.,& Ghorbani, A. A.(2009). A detailed analysis of the KDD CUP 99 data set. *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, Ottawa, ON, Canada, 2009, pp. 1-6, doi: 10.1109/CISDA.2009.5356528

Yang, Y., Zheng, K., Wu, C., & Yang, Y. (2019). Improving the Classification Effectiveness of Intrusion Detection by Using Improved Conditional Variational AutoEncoder and Deep Neural Network. *Sensors*, *19*(11), 2528. https://doi.org/10.3390/s19112528

Yen, S.-J., & Lee, Y.-S. (2006). Under-Sampling Approaches for Improving Prediction of the Minority Class in an Imbalanced Dataset. In D.-S. Huang, K. Li, & G. W. Irwin (Eds.), *Intelligent Control and Automation: International Conference on Intelligent Computing, ICIC 2006 Kunming, China, August 16–19, 2006* (pp. 731–740). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-37256-1_89

Yin, C., Zhu, Y., Fei, J., & He, X. (2017). A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks. *IEEE Access*, *5*, 21954 21961. https://doi.org/10.1109/ACCESS.2017.2762418

Zhang, H., Huang, L., Wu, C. Q., & Li, Z. (2020). An effective convolutional neural network based on SMOTE and Gaussian mixture model for intrusion detection in imbalanced dataset. *Computer Networks*, *177*, 107315. https://doi.org/10.1016/j.comnet.2020.107315

Zhang, H., Wu, C. Q., Gao, S., Wang, Z., Xu, Y., & Liu, Y. (2018). An Effective Deep Learning Based Scheme for Network Intrusion Detection. *2018 24th International Conference on Pattern Recognition (ICPR)*, 682–687. https://doi.org/10.1109/ICPR.2018.8546162

ACTA INFOLOGICA 2023;7(1):139-151
dergipark.org.tr/acin

DOI: 10.26650/acin.1180583

RESEARCH ARTICLE

Machine Learning Models on MOBA Gaming: League of Legends Winner Prediction

MOBA Oyunlarında Makine Öğrenimi Teknikleri: League of Legends Kazanan Tahmini

Kaan Arık[1]

ABSTRACT
The entertainment industry includes companies engaged in telecommunications services, television, music streaming, video games, and live events. Gaming has gained momentum in revenue growth in the entertainment industry over the past decade. This momentum has made the gaming industry one of the most popular areas of the entertainment industry. Official leagues have been teamed up with professional players, and the concept of e-sports has become widespread. MOBA (Multiplayer Online Battle Arena), which is a derivative of MMO (massively multiplayer online) games, is the name given to the games played on the Internet in which players destroy the opponent's base by dominating specific objectives on a map, usually with two teams of five players each. LoL (League of Legends) is one of the most popular MOBA games. Predicting winners in online games has become an essential application for machine learning models. This research aims to predict classification with machine learning methods of match winner with LoL player metrics. Key performance metrics and their impact on each game model were analyzed. The results show that winner prediction is possible in League of Legends, also, LightGBM (0.97), Logistic Regression (0.96), SVM and GBC (Gradient Boosting Classifier) (0.95) are outperformed with a high accuracy ratio. This paper will contribute to the classification research on topic of gaming with machine learning.
Keywords: Gaming, classification, machine learning, league of legends, MOBA

ÖZ
Eğlence endüstrisi, telekomünikasyon hizmetleri, televizyon, müzik, video oyunları ve canlı konserler gibi işlerle uğraşan alışılmadık derecede geniş bir şirket yelpazesini içerir. Oyun, son on yılda eğlence sektöründe gelir artışı ivmesi elde etmiştir. Bu ivme oyun sektörünü eğlence endüstrisinin en popular alanlarından biri haline getirmiştir. Profesyonel oyuncularla resmi ligler kurulmuş ve e-spor kavramı yaygın hale gelmeye başlamıştır. Çevrimiçi oyun türlerinden olan MMO (Massive Multiplayer Online) oyunların bir türevi olarak karşımıza çıkan MOBA (Multiplayer Online Battle Arena) internet üzerinde genellikle 5 kişilik 2 takımla bir harita üzerinde belirli yapıları domine ederek rakibin üssünü yok etme hedefiyle oynanan oyunlara verilen isimdir. LoL (League of Legends) bir MOBA oyunudur. Çevrimiçi video oyunlarında kazananların tahmini, makine öğrenmesi tabanlı tahmin modelleri için önemli bir uygulama haline gelmiştir. Araştırmanın hedefi LoL oyuncu metrikleriyle maç kazanma tahminin makine öğrenmesi yöntemleriyle sınıflandırma tahminidir. Önemli performans ölçütleri ve bunların her bir oyun modeli üzerindeki etkisi de analiz edildi. Sonuçlar, League of Legends oyununda kazanan tahmininin mümkün olduğunu göstermektedir. LightGBM (0.97), Lojistik regresyon (0,96), SVM ve GBC (0.95) başarım oranı ile öne çıkan algoritmalardır. Çalışmanın oyun alanında makine öğrenmesiyle sınıflandırma çalışmalarına katkı sağlayacağı düşünülmektedir.
Anahtar Kelimeler: Makine öğrenmesi, sınıflandırma, oyun, league of legends, MOBA

[1](Lect.) Sakarya Applied Sciences University, Department of Multidimensional Modeling and Animation, Sakarya, Turkiye

ORCID: K.A. 0000-0002-0930-8955

Corresponding author:
Kaan ARIK
Sakarya Applied Sciences University, Department of Multidimensional Modeling and Animation, Sakarya, Turkiye
E-mail address: kaanarik@subu.edu.tr

Submitted: 26.09.2022
Revision Requested: 23.03.2023
Last Revision Received: 02.05.2023
Accepted: 09.05.2023
Published Online: 15.06.2023

Citation: Arik, K. (2023). Machine learning models on moba gaming: league of legends winner prediction. *Acta Infologica, 7*(1), 139-151. https://doi.org/10.26650/acin.1180583

1. INTRODUCTION

In recent years, video games have been one of the popular fields of the entertainment industry for adults and children. Games started to be played in the 1960s and prime arcade games became widespread in the mid-1970s by introducing Pac-Man and Space Invaders by Atari (Donovan, 2010). During the 1980s and 1990s, Nintendo released handheld game consoles and gained popularity among a colossal gamer audience (Kent, 2001). With the development of specialized hardware, in the 2000s, Xbox and Sony launched their first-generation consoles that are still widely played today (Schreier, 2017). Since the introduction of smartphones and tablets between 2000 and 2010, games are becoming a product that can be reached instantaneously. And now, thanks to hardware performance and computer abilities, people can play games with realistic graphics anytime. People often use smart devices while online shopping, having fun, communicating with others, and spending more time with those devices than older generations. In 2001, Marc Prensky introduced the "Digital Natives" term and defined them as "ones at a higher level in terms of technical competence compared to their predecessors". Undoubtedly, most of the time these devices are used is for games and their community.

The entertainment industry includes companies engaged in telecommunications services, television, music streaming, video games, and live events (Nathan Reiff, 2022). Gaming has gained momentum in revenue growth in the entertainment industry over the past decade. Games are interactive and have succeeded thanks to ensuring more satisfaction than visual and written content(Robson & Meskin, 2016). For instance, no matter how much people like a movie in their life, they never watch that movie all the time, because movies open up an opportunity for linear expression. However, if one likes the game, they want to play it multiple times. Games are non-linear interactive structures that allow players to make in-game decisions, thus keeping their interest. Eventually, it's worth stating that games take the experience and storytelling in movies(Beverly Peders, 2018).

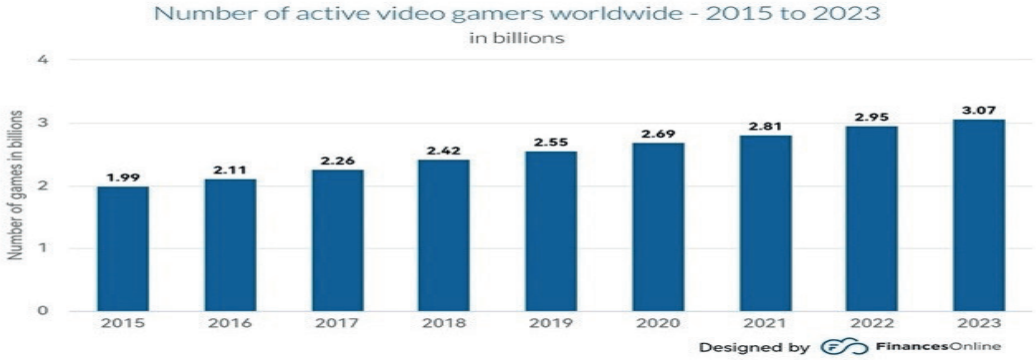


Figure 1. Number Of Active Players Between 2015-2023 in Billions

Recently, the interest in online games has increased with the ease of Internet access worldwide. At the beginning of 2022, there were 4.95 billion internet users worldwide, making up 62.5 percent of the world's population. According to an estimate for yearly growth of 5.6%, the number of video game players would rise from 2.69 billion in 2020 to 3.07 billion in 2023. In 2020, the global gaming industry brought in $159.3 billion, with the Asia Pacific region accounting for nearly half of that total. The number of active players and the global increase between 2015-2023 are given in Figure 1(Simon Kemp, 2022).

In addition, games are popular and yield high revenues; it's also paved the way for forming its own culture(Nestor Gilbert, 2022). Magazines, websites and internet content for digital games have been launched, and this situation has gone to the next level, becoming a structure where a virtual world exists over the Internet. Moreover, League of Legends, DOTA2, Counter-Strike Global Offense, Valorant and others are included in the MOBA games that are referred to as online games today.

PC games are one of the platforms in which video games are best experience. This platform stands out with the advantage of supporting high-quality resolution and offering more gameplay. Mostly, PC games outperform when compared to consoles,

because PC games provide more customization through various accessories and options that can enhance the experience. For example, while you can use your keyboard and mouse to play games on a PC, you can also use it by integrating a game controller. It also offers a broader range of games, thanks to game-selling media such as Steam and Epic Store.

2. MOBAs AND MACHINE LEARNING

This section presents the definition and features of MOBAs and research on machine learning applications in the game industry.

2.1. MOBA (Multiplayer Online Battle Arena)

MMO (Massive Multiplayer Online) is a game many people can play simultaneously (Donovan, 2014). The reason for including the definition of MMO games is MOBA (Multiplayer Online Battle Arena) games partially use MMORPG mechanics (Galehantomo, 2015). An online multiplayer battle arena, or MOBA, is a real-time strategy video game genre. League of Legends, DOTA II, Heroes of the Storm, etc. are included in this category. Games are classified differently based on gameplay and shooting angles. The MOBA (Multiplayer Online Battle Arena) genre is included in the list of online games. MOBAs, an exciting subgenre of MMO games, have become an essential part of video games and a pop culture phenomenon in recent years. MOBAs typically comprise five players on each team, each controlling a single character. In some other MOBAs, player number changes regarding game mechanics and map. Unlike MMORPG (Massively Multiplayer Online Role-Playing Game) games, a MOBA game does not have a unified structure or many simultaneous players on a map; much of the strategy is developed around cooperative team play and individual characters.

MOBAs tend to have a real-time strategy element and revolve around a simple goal: to compete against opponents in a team and defeat them in combat. Players make instant progress and must have a constant Internet connection; otherwise, they can be accepted as AFK (Away from Keyboard) players. Game sessions usually are between 30 minutes and 1 hour, although they can be longer or shorter based on player progress. In MOBAs, champions with a specific skill kit are usually offered to players by game designers. Also, players focus on winning the game in different lanes by choosing these champions. LoL (League of Legends) is played on the 'Summoner's Rift' map. In addition, there are different forms (drakes, Baron Nashor, jungle camps, towers etc.) that can kill players, which is important for customizing their champion and purchasing different items. Often the task of these forms is to provide teams with extra skill points and buffs. Every player has league points in their profiles, and they were categorized into different leagues based on win/lose rates. Players are often located in different lanes, where they fight against rivals. It mainly consists of four corridors: top-lane, mid-lane, bot-lane, and jungle. Only bot-lane players can play as a duo. The primary source of money for players while playing is to have in-game resources by hitting the last hit to minions that are produced from the bases. Matches are divided into two groups, ranked and normal. While players get points only in a ranked match, they only gain experience in a normal match. LoL had 117 million monthly players as of 2022. Also, it has played a crucial role in the popularity of the new sports branch, called e-sports, in recent years (League of Legends Live Player Count and Statistics, 2022). League of Legends has official leagues with professional teams in countries such as America, Europe, China, South Korea, and Türkiye.

2.2. Machine Learning and Gaming

Customer loyalty is the persistent emotional bond between a firm and a customer. This relationship is evident when a consumer consistently purchases from you rather than one of your rivals (Oracle, 2020). When a customer has a good experience with you, loyalty develops naturally and aids in developing trust. In recent years, the gaming industry has placed more emphasis on the idea of loyalty, because businesses want users to play their games frequently and to pay them money. Companies examine the participants and apply the right policies in this procedure using statistical and artificial intelligence techniques (Alpaydin, 2020). The video game industry benefits from artificial intelligence in numerous ways, including player league classification, customer churn analysis, winner prediction, and the ability for players of the same level to compete in a match.

Machine learning is a branch of artificial intelligence that uses statistical models and algorithms to manage mechanical motion (Nilsson, 2010). Contrast this with traditional AI methods like search trees and expert systems. In recent years,

machine learning has been one of the most popular areas in academia and practice. It has evolved into one of the frameworks researchers use to assess and categorize participants, much like games. Video games have utilized various artificial intelligence techniques, from PCG (procedural content generation) to NPC (Non-player Character) control. It's also actively used to make enemy characters playable against rivals. Using deep learning agents to compete against skilled human players in sophisticated strategy games is the most well-known example of how machine learning is used in video games. Machine learning has found significant use in games like Atari/ALE, Doom, StarCraft, and Minecraft. Machine learning also impacts games like Chess and Go that weren't designed to be played on screens (Justesen et al., 2019).

Machine learning algorithms perform quite well in game AI problems. Ensemble learning and tree-based models have been successful on various datasets. These are primarily solutions to classification problems. Issues such as player defection, winner/loser prediction, and CLV estimation are well mentioned in the literature. This section provides solutions to problems such as business analysis and classification. One of the most common problems in game AI is data access. High-budget game studios buy low-budget game studios and consolidate these companies into their ecosystems into a single hub. For this reason, we are very cautious about sharing data with researchers. However, in-game metrics data via API may be used for research within the guidelines set by the company.

3. DATASET AND METHOD

This section mentions the data used during research and the theoretical structures of implemented algorithms. There is information about the scraping dates of the data, data information, and how data was scraped. In addition, training and testing division and the pre-processing steps are discussed in this section.

3.1. Dataset

The data set consists of approximately 138,000 players and between 10.01.2022 – 12.09.2022 with 96 columns. The data was scraped via Riot API, keeping with Riot Games data privacy and framework policy(Tim Sevenhuysen, n.d.). The IDs of the players have been anonymized, and no-account information has been included in the research. Each game has ten players that target column winner (1)/loser (0). Attributes are divided into two groups player and in-game metrics. The attributes and explanations in the data are given in Appendix B.

3.2. Method

The research aims to predict the winner/loser classification using machine learning techniques based on player and in-game metrics. Data mining steps were chosen throughout the research based on the CRISP-DM model. CRISP-DM (Cross-Industry Standard Process for Data Mining) is a standardized methodology used to describe how business problems are solved with data-based solutions and to increase the efficiency of business applications. Due to CRISP-DM being a high-level methodology, the steps outlined in the model can be implemented in many different ways, sequences, and technologies to meet business needs. It consists of 6 steps: business understanding, data understanding, data preparation, modelling, evaluation, and deployment (IBM, 2021).

Table 1
Data partition for train and testing

Split Ratio	Train Set	Test Set	Total
%70 - %30	92979	39849	132828

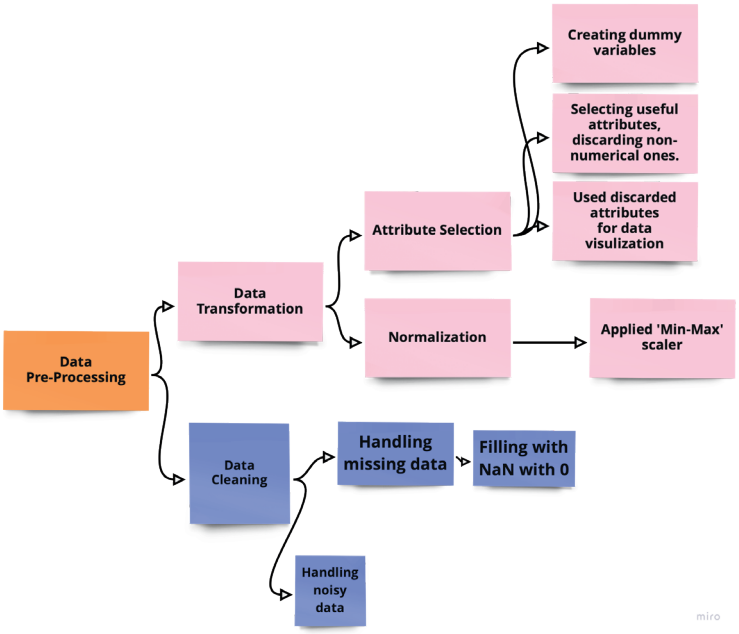


Figure 2. Diagram Of Data Preprocessing

Here are the data pre-processing steps:

· Non-numeric (date, group, id, status etc.) attributes were eliminated from the data as feature engineering.

· Attributes with correlations above 0.90 were excluded because they cause over-fitting in tree-based algorithms.

· In total, data has been reduced from 123 to 96 features.

· NaN values are marked as 0, not because the player does not have the opportunity to perform the tasks but because they didn't choose to do it, described as sparse data.

· When an AFK (Away from Keyboard) player is in-game, other players can start a new match by voting in the first 10 minutes without losing league points. In this respect, rows with a game session of less than 600 seconds were removed from the data.

· Data was normalized between a value of 0 and 1.

3.3. Machine Learning Algorithms

The ability to utilize the best method to turn a dataset into a model is known as machine learning. Machine learning is a collection of techniques for building models out of data. The most effective method (supervised, unsupervised, reinforcement learning, etc.) will vary depending on the data, the resources at hand, the nature of the data, and the predicted outcome of all processes (Han & Kamber, 2012). Ordinary programming algorithms simply tell the computer what to do. Machine learning algorithms contain simpler methods than nonlinear regression, giving more accurate results and outputs compatible with mathematical functions. It is often possible to talk about two methods: regression and classification. While making inferences on quantitative data (income level, plant height, etc.) with regression, classification also includes non-numerical variables (credit approval status, gender, number of rooms, etc.). Among algorithms implemented throughout the research, the best-performing ones were mentioned. LightGBM, Logistic Regression, Support Vector Machines, and Gradient Boosted Classifiers are algorithms with the best accuracy ratio for prediction. Performance metrics were explained in the findings section and discussed in the conclusion. Also, the confusion matrix and ROC curves for implemented algorithms have been added to Appendix A.

3.3.1. LightGBM (Light Gradient Boosting Machine)

As a component of the Microsoft DMTK (Distributed Machine Learning Toolkit) project, the boosting method LightGBM was created in 2017. Its benefits over other boosting methods include fast processing, handling large amounts of data, using fewer resources (RAM), a high prediction rate, parallel learning, and GPU learning support. A histogram-based method is used in LightGBM(Microsoft, 2022). Making the variables with continuous values lowers the computing cost. The computation and, thus, the number of branching directly relate to the training duration of the decision trees. This approach results in a decrease in both resource utilization and training time. Learning decision trees can be done in one of three ways: level-wise, depth-wise, or leaf-wise. Two strategies can be used when learning decision trees: level-by-level, depth-by-depth, or leaf-by-leaf. A level-oriented strategy keeps the tree in balance as it grows.

A hand-focused strategy, however, will continue to split hands and reduce losses. This feature sets LightGBM apart from other boosting algorithms. A leaf-oriented strategy results in a lower model error rate and faster learning. However, the leaf-oriented growth strategy makes the model prone to overfitting when the number of data is small. Therefore, this algorithm is more suitable for use with big data. Additionally, parameters such as tree depth and number of leaves can be optimized to prevent overfitting(Ke et al., n.d.).

3.3.2. Logistic Regression

Logistic regression is like a regression problem in which the dependent variable is categorical, and it is frequently applied to linear classification issues. A classification occurs here even though it is labelled regression(Wright, 1995).

$$P = \frac{e^{a+bX}}{1 + e^{a+bX}} \qquad \text{Formula (1)}$$

An analysis of a dataset with one or more independent variables predicting an outcome is done statistically using logistic regression. The predicted class is measured with a binary variable (only two possible outcomes)(Kleinbaum & Klein, 2010). Although logistic regression has the word regression in its name, it is a classification algorithm. A few iterations of artificial neural networks and a logit function are applied, as seen in Formula 1(Alpaydin, 2020).

3.3.3. Support Vector Machines

Support Vector Machine is a supervised learning method generally used in classification problems. It draws a line to separate points placed on a plane, aiming to have this line at the maximum distance for the points of both classes. It suits complex but small to medium datasets (William Noble, 2006).

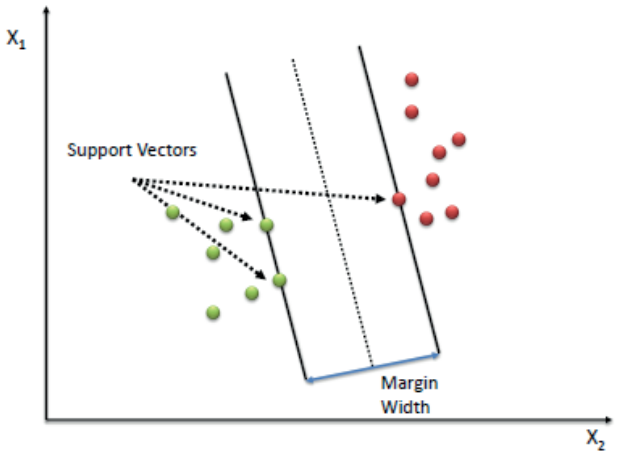


Figure 3. SVM Algorithm with Margins

The main purpose of classification problems is to decide in which class the future data will take place. A line separating the two classes is drawn to make this classification, and the region between ±1 of this line is called the margin. The wider the

margin, the better the separation of two or more classes. At the same time, apart from the linear model, it performs data separation in different kernels(Géron, 2019).

3.3.4. Gradient Boosting Classifier

Gradient boosting classifiers update the classifiers and weighted inputs using the AdaBoosting algorithm in conjunction with weighted minimization. Reducing the loss, or the discrepancy between the actual class value of the training example and the predicted class value is the goal of Gradient Boosting classifiers. Although comprehension of the method for decreasing the classifier's loss is unnecessary, it works similarly to gradient descent in neural networks(Lemarechal, 2012).

In the case of Gradient Boosting Machines, the weights of the preceding learners are frozen or cemented in place, remaining intact when the new layers are added, and this is done each time a new weak learner is added to the model. The methods employed in AdaBoosting, where the values are modified when additional learners are added, differ from this. Gradient boosting algorithms' strength stems from the fact that they can be applied to situations involving more than just binary classification; they can also be used to solve regression and multi-class classification problems(Mason et al., 1999).

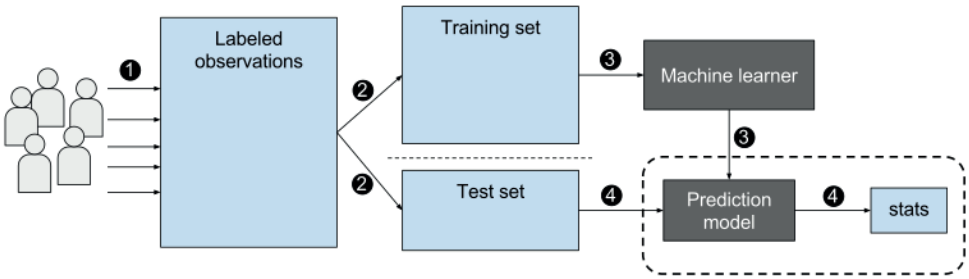


Figure 4. GBC (Gradient Boosting Classifier) Algorithm Diagram

A loss function is necessary for the Gradient Boosting Classifier to operate. Gradient boosting classifiers can handle a variety of standardized loss functions in addition to bespoke loss functions. However, the loss function must be differentiable. Regression techniques employ squared errors, whereas classification algorithms often use logarithmic loss. Any differentiable loss function may be used in gradient boosting systems instead of having to be derived specifically for each additional boosting procedure. A gradient-boosting model's additive component results from adding new trees over time without changing the values of the model's already-existing trees. The error between the specified parameters is minimized using a method akin to gradient descent.

3.4. Performance Metrics and Libraries

Accuracy is the percentage of samples classified as correct. Recall is a metric that shows how many of the transactions of positively predict. Precision shows how many of the values predicted as positive are positive. The F1 score measures a test's accuracy—the harmonic mean of precision and sensitivity (Géron, 2019)-(Stehman, 1997).

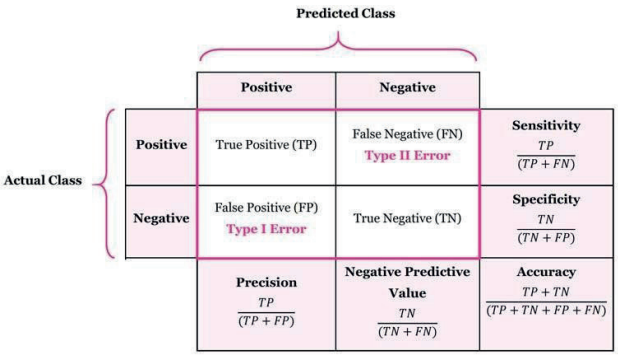


Figure 5. Performance Evaluation Metrics in the Classification Problem (Shajihan, 2020)

The research used 'Pandas and NumPy' for data processing and the 'Scikit-Learn' library in Python to applying machine learning algorithms. The applications were carried out on Google Collab, and 'pyCaret' library was used to evaluate the performance of the algorithms.

4. FINDINGS

In this section, classification algorithms are implemented on the dataset. Also, a parameter tuning was made for all the methods and parameters that could obtain the best results. Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample. The procedure has a single parameter called k, which refers to the number of groups a given data sample will be split into. For ML algorithms, 'k-fold cross-validation' method has been deployed with k=10. LightGBM, Logistic Regression, Support Vector Machines, Gradient Boosting Classifier, Random Forest, Linear Discriminant Analysis, Ridge Classifier, Extra Tree, AdaBoost, and Decision Tree performed over 90%, as seen in Table 2. The best classifier in terms of accuracy is LightGBM. Therefore, when compared to runtime, it is evident that LightGBM outperforms results based on performance. It performs not only in terms of accuracy ratio, but also recall, precision and F1 score. LightGBM algorithm is a histogram-based algorithm that selects the best split in the sorted histogram. It speeds up training and reduces memory usage focusing on the accuracy of results. It provides faster and higher efficiency. It performs better with less memory usage and can handle large data sets. As a result, ensemble methods perform with high accuracy rates in classification in gaming (Arık et al., 2022).

Table 2
Performance metrics of algorithms in classification

Model	Accuracy	AUC	Recall	Prec.	F1	TT (Sec)
LGBM	0.968	0.996	0.971	0.966	0.969	4.8870
LR	0.955	0.988	0.956	0.954	0.955	8.6470
SVM	0.950	0.000	0.954	0.946	0.950	0.6480
GBC	0.949	0.990	0.953	0.946	0.949	89.9920
RF	0.9481	0.990	0.955	0.941	0.948	32.7530
LDA	0.9449	0.982	0.948	0.941	0.945	3.1410
RC	0.9416	0.000	0.946	0.937	0.941	0.3960
ETC	0.9393	0.987	0.949	0.930	0.939	19.2340
ADAB	0.9353	0.983	0.936	0.934	0.935	17.3250
DT	0.9263	0.926	0.926	0.925	0.926	4.6720
KNN	0.8879	0.948	0.898	0.880	0.889	59.9160
QDA	0.7951	0.952	0.625	0.947	0.753	1.6390
NB	0.6481	0.857	0.341	0.881	0.492	0.4530

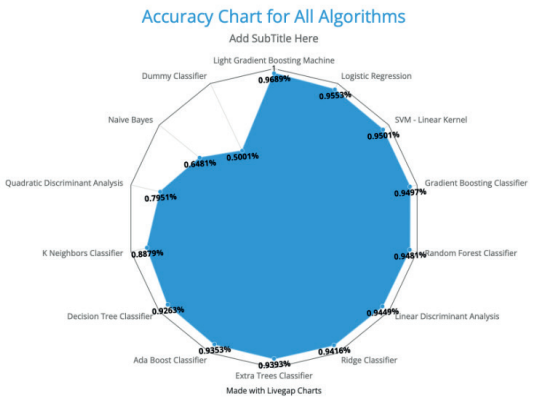


Figure 6. Radar Chart of Accuracy Ratio for Models

Accuracy ratio, precision, recall, and F1 scores of LR, SVM, and GBC algorithms also perform well compared to LightGBM. All implemented algorithms are successful except Quadratic Discriminant Analysis and Naive Bayes. In addition, compared to the running time of other algorithms, KNN, GBC, and RF are slower than others.

Table 3
LightGBM tune parameters and description

Parameter	Description	Value
boosting_type	Boosting type	gbdt
importance_type	Type of feature importance	split
learning_rate	Boosting learning rate	0.1
max_depth	maximum distance between the root node and leaf node	-1
n_estimators	Number of estimators	100
num_leaves	Control parameter for the complexity of the tree model	31
subsample	Sub-samples ratio	1.0

The majority of high accuracy ratios are ensemble learning models. As a result, ensemble learning algorithms outperform results in classification problems in gaming. Specifying the learning rate as small and the number of iterations as high in the LightGBM algorithm and having large training data are essential parameters for developing successful models. For this reason, the use of 'learning_rate' = 0.1, 'n_estimators' = 100 and "gbdt" as the enhancement type in our model has been effective in the performance success of the model.

5. CONCLUSION AND DISCUSSION

Table 4 lists research published on MOBA games over the past few years. It includes the year, author, game genre, and best accuracy. Since 2017, classification and clustering algorithms, mainly classification, have been implemented in MOBA games every year. According to the literature, Random Forest and Logistic Regression algorithms are high-performance in gaming. Research has been conducted with LoL (League of Legends) data with high accuracy scores. Among the reasons for this, as mentioned in the previous parts of the article, it is thought that there is a problem accessing data. In [30][33] Naive Bayes and Logistic Regression achieved a good level of success with 0.77 in previous studies but not as well as others. In addition to classification, clustering models were implemented on such gaming datasets [31].

However, among those implemented in Table 4, it has a lower accuracy ratio than others. In another study, statistics-based models do not perform much accuracy in AI applications in gaming. Moreover, ML algorithms that implemented ensemble learning have become life-saving models in data competitions and researchers. Here again, the best results are performed through tree-based ensemble learning algorithms.

Table 4
Related works

Year	Authors	Genre	Best Accuracy
2017	Almeida and et al. (Almeida et al., 2017)	MOBA	Naive Bayes (0.77)
2018	Mora-Cantallops and Sicilia (Mora-Cantallops & Sicilia, 2018)	MOBA	k-means
2019	Ani and et al. (Ani et al., 2019)	MOBA	Random Forest (0.99)
2019	Porokhnenko, etc.(Porokhnenko et al., 2019)	MOBA	Logistic Regression (0.70)
2021	Costa and et al.(Costa et al., 2021)	MOBA	TSSTN
2021	Yang and et al. (Yang et al., 2021)	MOBA	Logistic Regression and Random Forest (0.97)

Machine learning algorithms are generally implemented for classification problems in gaming, and deep learning is not needed much. Because deep learning works better on datasets with image, audio, video, and text content, it is also difficult to predict how well it will work, as it's a black-box model. For future works, there will be models with much larger data sets to work with sets with a large variety of attributes and missing data, which will increase the applicability of such models on a company basis.

ROC and Precision-Recall curves, features importance plots, number of predicted classes, confusion matrix, and algorithms abbreviation and title are attached in Appendix A. AUC plots explain how well the model can predict classes, as seen as ROC curves for the LightGBM classifier in Appendix A. The higher the AUC, the better the model is at predicting false as false and true as true. Accuracy increases as the curve gets closer to the top-left (Krzanowski & Hand, 2009). For example, the higher the AUC in our data dataset of winner and loser, the better the model performs in predicting between winners and losers. When the ROC plot of LightGBM algorithm is analyzed, the curve has an almost 100% accuracy ratio. In other words, the model performed very well in both positive and negative classes for problems. Looking at the confusion matrix, samples are labelled as a winner in data but classified as a loser in prediction (n=591) and labelled as a loser but classified as a winner (n=704). It is almost impossible to achieve 100% accuracy rate in gaming datasets. Because LoL is a game played with 5v5 and teammates' well performance can make the same team's bad players win the game. For this reason, the misclassification of 1295 samples is acceptable. When examining players, it can be interpreted both groups were equally classified correctly.

The ROC curve is a very crucial performance measure for classification problems. ROC is a probability curve and area under it, and it represents the degree or measure of separability. The ROC curve has FPR (False Positive Rate) on the X axis and TPR (True Positive Rate) on the Y axis. The higher the level under the curve, the higher the class discrimination performance. In research, successful results were achieved on negatives by examining ROC curves. As seen in the Features Importance plot, 'earned_gpm' (Earned Game Per Minutes) is the most important attribute. Then the list goes on to include 'earned_gold_shared', 'gold_spent', 'assist' and 'death' attributes. Checking the chart and interpreting that gold earned in-game is important for a player who has never played LoL. The game's main aim is not to kill rivals throughout the session but to develop quickly individually by killing the rivals with the resources in the session. In this respect, the importance of predicted features makes this research meaningful.

Peer-review: Externally peer-reviewed.
Conflict of Interest: The author has no conflict of interest to declare.
Grant Support: The author declared that this study has received no financial support.

REFERENCES

Almeida, C. E. M., Correia, R. C. M., Eler, D. M., Olivete-Jr, C., Garci, R. E., Scabora, L. C., & Spadon, G. (2017). Prediction of winners in MOBA games. *2017 12th Iberian Conference on Information Systems and Technologies (CISTI)*, 1–6. https://doi.org/10.23919/CISTI.2017.7975774

Alpaydin, E. (2020). *Introduction to machine learning* (Fourth edition). The MIT Press.

Ani, R., Harikumar, V., Devan, A. K., & Deepa, O. S. (2019). Victory prediction in League of Legends using Feature Selection and Ensemble methods. *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, 74–77. https://doi.org/10.1109/ICCS45141.2019.9065758

Arık, K., Gezer, M., & Tayali, S. T. (2022). Bibliometric Analysis of Scientific Studies Published on Game Customer Churn Analysis Between 2008 and 2022. *Journal of Politics*, *5*(1), 21.

Beverly Peders. (2018). *Film vs. Video Games From A Screenwriter's Perspective*. https://www.wescreenplay.com/blog/film-vs-video-games-from-a-screenwriters-perspective/

Costa, L. M., Mantovani, R. G., Monteiro Souza, F. C., & Xexéo, G. (2021). Feature Analysis to League of Legends Victory Prediction on the Picks and Bans Phase. *2021 IEEE Conference on Games (CoG)*, 01–05. https://doi.org/10.1109/CoG52621.2021.9619019

Donovan, T. (2010). *Replay: The history of video games*. Yellow Ant.

Galehantomo, G. (2015). Platform Comparison Between Games Console, Mobile Games And PC Games. *SISFORMA*, *2*, 23. https://doi.org/10.24167/sisforma.v2i1.407

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (Second edition). O'Reilly Media, Inc.

Han, J., & Kamber, M. (2012). *Data mining: Concepts and techniques* (3rd ed). Elsevier.

IBM. (2021, August 17). *IBM Documentation*. https://prod.ibmdocs-production-dal-6099123ce774e592a519d7c33db8265e-0000.us-south.containers.appdomain.cloud/docs/en/spss-modeler/saas?topic=dm-crisp-help-overview

Justesen, N., Bontrager, P., Togelius, J., & Risi, S. (2019). *Deep Learning for Video Game Playing* (arXiv:1708.07902). arXiv. https://doi.org/10.48550/arXiv.1708.07902

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (n.d.). *LightGBM: A Highly Efficient Gradient Boosting Decision Tree*. 9.

Kent, S. L. (2001). *The ultimate history of video games: From Pong to Pokémon and beyond: the story behind the craze that touched our lives and changed the world* (1st ed). Prima Pub.

Kleinbaum, D. G., & Klein, M. (2010). Introduction to Logistic Regression. In *Logistic Regression: A Self-Learning Text* (pp. 1–39). Springer. https://doi.org/10.1007/978-1-4419-1742-3_1

Krzanowski, W. J., & Hand, D. J. (2009). *ROC curves for continuous data*. Chapman & Hall/CRC.

League of Legends Live Player Count and Statistics. (2022, July 11). https://activeplayer.io/league-of-legends/

Lemarechal, C. (2012). Cauchy and the Gradient Method. *Documenta Mathematica*, 4.

Mason, L., Baxter, J., Bartlett, P. L., & Frean, M. R. (1999). *Boosting Algorithms as Gradient Descent*. *3*, 7.

Michael Donovan. (2014, February 3). *The MMO vs. The MOBA*. https://www.gamedeveloper.com/disciplines/the-mmo-vs-the-moba

Microsoft. (2022). *LightGBM Documentation*. https://lightgbm.readthedocs.io/en/v3.3.2/

Mora-Cantallops, M., & Sicilia, M.-Á. (2018). Player-centric networks in League of Legends. *Social Networks*, *55*, 149–159. https://doi.org/10.1016/j.socnet.2018.06.002

Nathan Reiff. (2022). *10 Biggest Entertainment Companies*. https://www.investopedia.com/articles/investing/020316/worlds-top-10-entertainment-companies-cmcsa-cbs.asp

Nestor Gilbert. (2022). *Number of Gamers Worldwide 2022/2023: Demographics, Statistics, and Predictions*. https://financesonline.com/number-of-gamers-worldwide/

Nilsson, N. J. (2010). *The quest for artificial intelligence: A history of ideas and achievements*. Cambridge University Press.

Oracle. (2020). *What is Customer Loyalty?* https://www.oracle.com/tr/cx/marketing/customer-loyalty/what-is-customer-loyalty/

Porokhnenko, I., Polezhaev, P., & Shukhman, A. (2019). Machine Learning Approaches to Choose Heroes in Dota 2. *2019 24th Conference of Open Innovations Association (FRUCT)*, 345–350. https://doi.org/10.23919/FRUCT.2019.8711985

Robson, J., & Meskin, A. (2016). Video Games as Self-Involving Interactive Fictions. *The Journal of Aesthetics and Art Criticism*, *74*(2), 165–177.

Schreier, J. (2017). *Blood, sweat, and pixels: The triumphant, turbulent stories behind how video games are made* (First edition). Harper Paperbacks.

Shajihan, N. (2020). *Classification of stages of Diabetic Retinopathy using Deep Learning*. https://doi.org/10.13140/RG.2.2.10503.62883

Simon Kemp. (2022). *Digital 2022: Global Overview Report* [Analysis]. https://datareportal.com/reports/digital-2022-global-overview-report

Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, *62*(1), 77–89. https://doi.org/10.1016/S0034-4257(97)00083-7

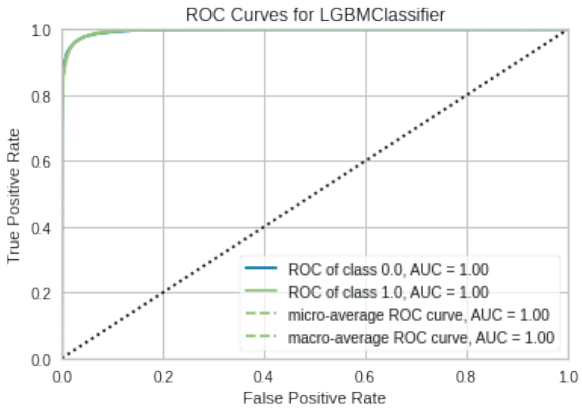
Tim Sevenhuysen. (n.d.). *Oracle's Elixir—LoL Esports Stats*. Retrieved September 15, 2022, from https://oracleselixir.com/tools/downloads

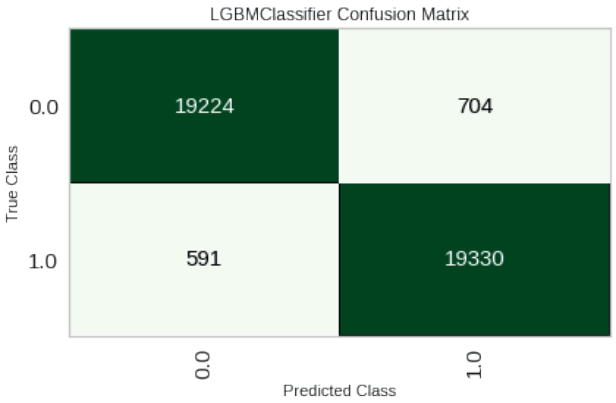
William Noble. (2006). *What is a support vector machine? | Nature Biotechnology*. https://www.nature.com/articles/nbt1206-1565

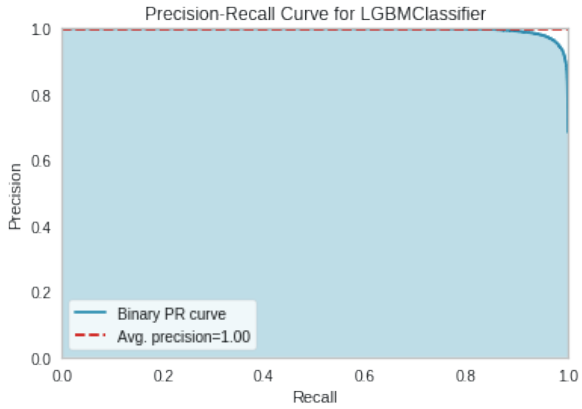
Wright, R. E. (1995). Logistic regression. In *Reading and understanding multivariate statistics* (pp. 217–244). American Psychological Association.

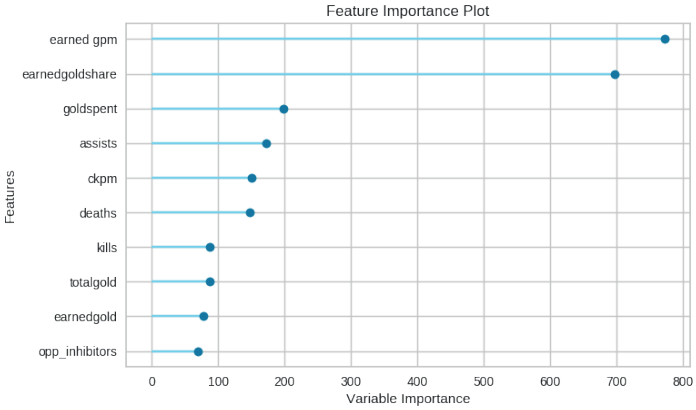
Yang, Z., Pan, Z., Wang, Y., Cai, D., Liu, X., Shi, S., & Huang, S.-L. (2021). *Interpretable Real-Time Win Prediction for Honor of Kings, a Popular Mobile MOBA Esport* (arXiv:2008.06313). arXiv. https://doi.org/10.48550/arXiv.2008.06313

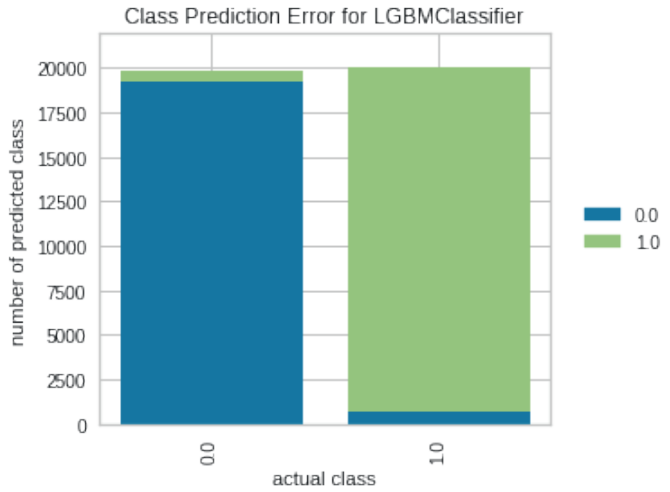
APPENDIX A











Abbreviation	Full Algorithm Title
LGBM	LightGBM
LR	Logistic Regression
SVM	Support Vector Machines
GBC	Gradient Boosting Classifier
RF	Random Forest
LDA	Linear Discriminant Analysis
RC	Ridge Classifier
ETC	Extra Trees Classifier
ADAB	Adaptive Boosting / AdaBoost
DT	Decision Tree
KNN	K-Nearest Neighbor
QDA	Quadratic Discriminant Analysis
NB	Naive Bayes

APPENDIX B

Term	Description
A	Total assists
AGT	Average game time/duration, in minutes
APG	Assists per game
B%	Percentage of games in which the champion was banned (not tied to a specific role)
BLND%	Blind-pick rate: percentage of games in which this player/champion was picked before their lane opponent (not always available)
BN%	Baron control rate
CCPM	Crowd control dealt to champions per minute
Champion	Champion name
CKPM	Average combined kills per minute (team kills + opponent kills)
CS%P15	Average share of team's total CS post-15-minutes
CSD10	Average creep score difference at 10 minutes
CSD15	Average creep score difference at 15 minutes
CSD20	Average creep score difference at 20 minutes
CSPM	Average monsters + minions killed per minute
CTR%	Counter-pick rate: percentage of games in which this player/champion was picked after their lane opponent (not always available)
CWPM	Control wards purchased per minute
D	Total deaths
D%P15	Average share of team's damage to champions post-15-minutes
DMG%	Damage Share: average share of team's total damage to champions
DMG%P15	Average share of team's damage to champions post-15-minutes
DPG	Deaths per game
DPM	Average damage to champions per minute
DRG%	Dragon control rate: percent of all Dragons killed that were taken by the team, reflecting only elemental drakes if ELD% is present
DTH%	Average share of team's deaths
EGPM	Average earned gold per minute (excludes starting gold and inherent gold generation)
EGR	Early-Game Rating
ELD%	Elder dragon control rate
Event	Event name
F3T%	First-to-three-towers rate (percentage of games in which team was the first to 3 tower kills
FB%	First Blood rate -- for players/champions, percent of games earning a First Blood participation (kill or assist)
FBN%	First Baron rate
FBV%	First Blood Victim rate -- percent of games player/champion was killed for First Blood
FD%	First dragon rate
FT%	First tower rate
GD10	Average gold difference at 10 minutes
GD15	Average gold difference at 15 minutes
GD20	Average gold difference at 20 minutes
GOLD%	Gold Share: average share of team's total gold earned (excludes starting gold and inherent gold generation)
GP	Games Played
GPM	Average gold per minute
GPR	Gold percent rating (average amount of game's total gold held, relative to 50%)
GSPD	Average gold spent percentage difference
GXD10	Average gold+experience difference at 10 minutes
GXD15	Average gold+experience difference at 15 minutes
GXD20	Average gold+experience difference at 20 minutes
HLD%	Rift Herald control rate
IWC%	Average percentage of opponent's invisible wards cleared
JNG%	Jungle Control: average share of game's total jungle CS
K	Total kills
KD	Kill-to-Death Ratio
KDA	Total Kill/Death/Assist ratio
KP	Kill participation: percentage of team's kills in which player earned a Kill or Assist
KPG	Kills per game
KS%	Kill share: player's percentage of their team's total kills
L	Losses
LNE%	Lane Control: average share of game's total lane CS
Losses	Total Losses
LP	Ladder Points
MLR	Mid/Late Rating
OE Rating	Oracle's Elixir Performance Rating
OE Rtg	Oracle's Elixir Performance Rating
P%	Percentage of games champion was picked in this role.
P+B%	Percentage of games in which the champion was either banned or picked in any role
Player	Player's in-game name
Pos	Position
PPG	Turret plates destroyed per game
Rank	Official Leaderboard Rank
STL	Neutral objectives stolen
STLPG	Neutral objectives stolen per game
StPG	Neutral objectives stolen per game
Team	Team name
VSPM	Vision score per minute
VWC%	Average percentage of opponent's visible wards cleared
W	Wins
W%	Win percentage
WC%	Average percentage of opponent wards cleared
WCPM	Average wards cleared per minute
Wins	Total Wins
WPM	Average wards placed per minute
XPD10	Average experience difference at 10 minutes
XPD15	Average experience difference at 15 minutes
XPD20	Average experience difference at 20 minutes

ACTA INFOLOGICA 2023;7(1):153-163
dergipark.org.tr/acin

DOI: 10.26650/acin.1231944

RESEARCH ARTICLE

Melting of Privacy with Machine Learning, Big Data, and Social Media

Pelin Canbay[1], Zübeyde Demircioğlu[2]

ABSTRACT

Every individual has the right to keep their information private. However, there is a big question: is this possible in the digital era? While social media attracts people to share personal data, most advanced technologies are continually developing in the area of how to exploit information from this personal data. Is it possible to talk about keeping personal data private? This study aims to investigate whether it is possible both to connect to the cyber-world and remain private in the digital era, where intensive studies have been conducted to protect privacy. This study discusses: (1) the social perception of privacy, (2) the contradiction between privacy expectations and behaviors, and (3) the current state of both disclosure and protection efforts of privacy with machine learning and big data techniques. As a result of our research, it was concluded that it is almost impossible to exist in the cyber/digital world and remain private, that most users are not uncomfortable with the current situation, and that institutions and technology developers should take more responsibility in this regard.

Keywords: Privacy, social media, big data, machine Learning

[1](Assist. Prof. Dr.), Kahramanmaraş Sutcu Imam University, Department of Computer Engineering, Kahramanmaras, Turkiye
[2](Assist. Prof. Dr.), Istanbul Medeniyet University, Faculty of Arts and Humanities, Department of Sociology, Istanbul, Turkiye

ORCID: P.C. 0000-0002-8067-3365; Z.D. 0000-0002-8749-006X

Corresponding author:
Pelin CANBAY
Kahramanmaraş Sutcuimam University, Department of Computer Engineering, Kahramanmaras, Turkiye
E-mail address: pelincanbay@ksu.edu.tr

Submitted: 10.01.2023
Revision Requested: 12.04.2023
Last Revision Received: 25.04.2023
Accepted: 17.05.2023
Published Online: 14.06.2023

Citation: Canbay, P., Demircioglu, Z. (2023). Melting of privacy with machine learning, big data, and social media. *Acta Infologica, 7*(1), 153-163.
https://doi.org/10.26650/acin.1231944

1. Introduction

Although privacy norms vary from culture to culture, society to society, and even from person to person, it is seen as a universal need as old as the history of humanity. Privacy is a right that should be protected and this means that an individual can determine what information or data are collected and analyzed about him/her in any given context. Privacy is one of the most critical concerns of the digital age where almost all personal information and data are stored electronically. Nowadays, people have a digital existence beyond their physical existence, and this digital existence tends to expand constantly. Therefore, people vacillate between the concern of protecting their personal information and the desire to be discovered and recognized (Bauman & Lyon, 2013). The rapid rise of social media allows users to share their personal data unquestioningly.

Social media is a computer-based application that uses digital entities such as texts, voice messages, images, or videos for interaction between individuals. Social media, also called Web 2.0, offers three specific services to users: ease of use, sociability, and uploading/sharing content the way they want. The distribution of data has gained a new form with social media (Fuchs, 2014). In the early days of social media, people who were enthusiastic about interacting, sharing, and collaborating through this application preferred these environments (Correa et al., 2010), and today it has become inevitable that personal data is stored in almost every service such as health and education. The volume and variety of data stored in digital media are increasing day by day. The necessity of managing this massive amount of data increases the use of big data technologies and the need for these technologies. Big data technologies with modern data science techniques, especially with machine learning methods, have the power to integrate data from multiple sources and reveal hidden patterns in them. The data collected from social media have led to many practical applications such as crime detection, recommendation systems, anomaly detection, behavioral analysis, bioinformatics, event detection, business intelligence, relationships, epidemics and opinion, sentiment, and emotions analysis, etc. (T.k. et al., 2021). It is possible to develop tools or applications that provide high benefits to humanity with the analysis of the personal data collected from digital media. Spoken dialog systems, for example, allow machines to help patients easily (Rosenthal et al., 2010; Balcı, 2019) or support therapy for less-abled persons (Matarić et al., 2007). On the other hand, the power of machine learning and big data to integrate data and reveal hidden patterns can have unpredictable negative consequences on data subjects (Kelleher & Tierney, 2018).

Many social and technological studies are carried out to protect the data owner. General Data Protection Regulation (GDPR) (Voigt & von dem Bussche, 2017) is one of the most important measures. The GDPR, established by the EU, legally enforces that the data subject's privacy must be respected. However, only a few governments have legislated the most basic measures. Although many privacy protection methods have been developed, such as anonymity, encryption, and distributed system privacy practices, new approaches are necessary. Since many privacy disclosure techniques have also been developed and used, it is obvious that more than current measures are needed to protect personal information while there are so many obligations to share data.

In this context, this study aims to comprehend how to protect personal privacy with regard to users, governments, and technology developers. According to our research;

- In many digital media, users are obliged to share their data to benefit from the service, while in many applications such as social media they share their personal data voluntarily.
- By using big data technologies and machine learning methods, unexpected inferences about individuals can be found with high success from the data that seem unrelated to each other. On the other hand, technologies developed to protect the privacy of shared and processed data have lagged behind emerging technologies in information extraction.
- It is imperative that the sharing of personal information should not be left to the user merely through the information and permission procedures, that Governments take measures to protect the privacy of users with legislation and follow these regulations, and that technology developers are aware of their responsibility about personal privacy before unexpected violations lead to more significant problems.

The rest of this paper includes the explanation of individuals' digital existence in Section 2. Section 3 discusses the current power of machine learning and big data technologies in regard to privacy. While Section 4 gives the current situation regarding

technology related to privacy, Section 5 provides the users' privacy attitudes and behaviors. In Section 6, the conclusion of the study is provided.

2. Privacy and Individual As a Social Being

One of the most fundamental problems of the cyber/digital age that we live in is privacy. In the digital age, people have a digital presence beyond their physical existence, and the area of this digital asset tends to grow constantly. It has become almost impossible to carry out daily work without sharing personal data to use digital services (Zhu, 2011). However, people's use of e-services is not always a necessity. People also use such services when they think it offers them convenience, cheapness, speed, or some other benefit (Bennett, 2009). In this section, the preferences and obligations between individuals' commitment to digital environments and their privacy are discussed, and suggestions from some studies are presented.

2.1. Socialization As a Human Need in the Digital World

In the age of modern technology, the individual cannot live without leaving a data trail behind. Not being in a social network in the digital age can lead to consequences such as being unable to maintain communication in social and business life, not being able to develop new relationships, missing opportunities, and sometimes even being excluded (Fuchs, 2014). However, humans are social creatures, and information sharing is a central feature of human connection. Making oneself public, sharing, and disclosing provide numerous benefits, including psychological and physical health. The desire for interaction, socialization, disclosure, recognition or fame, and fear of insignificance are basic human motives like privacy. Social media has provided a unique opportunity to meet these basic human motives (Acquisti et al., 2015).

Both the internet and social media deliver messages to large audiences and make messages much more visible. The thought of "I am seen (watched, noted, recorded) therefore I am" (Bauman & Lyon, 2013) is becoming widespread as a view that dominates social media. According to Niedzviecki, who deals with the concept of "peeping culture," an individual becomes aware of being an individual when she/he makes herself/himself watched and when others comment on himself/herself (Niedzviecki, 2009).

2.2. Personal Data as a Valuable Asset

Personal data is any data that can be used to distinguish one person from another. These personal data, some of which are presented in Figure 1, are used to group people in digital environments as target audiences, especially for commercial and political purposes (see the Cambridge Analytica scandal (Confessore, 2018)).

Figure 1. Some personal information given to or obtained from digital media

Personal data can be easily extracted through continuous individual traces left behind when browsing social networks and websites, or using devices such as smartphones. These data may be collected and used for marketing, provocation, or experimental research. With the commercial value of data, every user has become critical; applications and platforms aim to obtain more data from more users to gain maximum profit. Precisely for this reason, these commercial enterprises are designed to meet the basic needs of users, such as communicating, socializing, and having fun, in order to encourage data disclosure (Acquisti et al., 2015).

The use of personal data represents a power that can affect society and individuals in an unprecedented way, not only in economic terms. Thus, most experts agree that the complex balance between individual rights and collective knowledge should not be entrusted solely to market dynamics (Politou et al., 2021).

2.3. Current Social Measures

Users have hardly any control over their personal information stored and analyzed by the relevant data controllers (Mittelstadt & Floridi, 2016; S. Yu, 2016). There are two contexts of digital data that pose a problem in terms of privacy. In some cases, even if the individual has chosen to share the data by her/himself, she/he may need to gain the knowledge and control that this data can be used and shared by third parties. The concepts of data shadow and data traces are used to explain this distinction. While the data shadow consists of the data collected without the individual's knowledge, consent, and awareness, the data traces comprise data that a person knowingly makes public (Koops, 2013). With up-to-date machine learning and big data methods and tools, sensitive information can be disclosed from seemingly unrelated data, and many parties can be involved in the distribution of this data. Therefore, the consent and knowledge of the individual are insufficient to solve privacy problems. Based on the seemingly unconnected data that the individual voluntarily shares on social media, extremely personal information about the individual that is undesirable to be disclosed can be accessed (Kelleher & Tierney, 2018).

As expressed in the Science and Technology Board (PCAST) Report 2014[1], notification and consent is the most widely used method to protect privacy. However, this means putting the entire responsibility of protecting privacy on the individual. It is optimistic to think that the user reads all the notifications and understands the legal implications, but the situation is different in reality. For this reason, there is a need for administrative rules to be in force in data collection and processing activities.

The vast amount of people's information on digital media offers severe vulnerabilities to the right to privacy. Governments have had to make more specific regulations to ensure privacy is protected in any transaction involving sensitive information. To mitigate the privacy risk, the European Parliament and Council enforced the General Data Protection Regulations (GDPR) (Voigt & von dem Bussche, 2017). Although privacy has to be protected by such kinds of regulations, there is more need for frameworks or practices to guide the implementation of these regulations (Jones & Kaminski, 2021). Many countries still do not have such regulations; so there is a greater need to establish country-specific and legal regulations with their framework and practices (Carey & Acquisti, 2018).

3. Technological Privacy Protections and Disclosures

3.1. Machine Learning and Privacy

Machine Learning is the most significant part of Artificial Intelligence that has various algorithms for making the system learn itself. In artificial intelligence, learning means establishing a relation between the system's inputs and output(s). The system refers to the computational and algorithmic-based software. Researchers can use various software techniques to extract, identify, determine, predict or explore valuable information in the data of any domain. Artificial intelligence means to create a machine that has the ability to make one or more human-like behavior; machine learning means a computer program that can learn to produce a human-like behavior. This behavior is learned based on data, metrics, and feedback mechanisms, but the most crucial part is data (Joshi, 2020).

1 https://obamawhitehouse.archives.gov/blog/2014/05/01/pcast-releases-report-big-data-and-privacy

Machine learning techniques construct a mathematical model of data samples to make decisions, predictions, or information extraction, called a "train set." It is one of the essential parts of a machine learning system. In social media, there are comprehensive and streaming data that have to be stored and processed with high-level devices and mechanisms. Machine learning techniques are the possible ways to gain knowledge from these data by processing them. Machine learning has various interdisciplinary and intra-disciplinary domain applications such as opinion mining, medical sciences, textual forensics, etc. to obtain meaningful information from the data (Joshi, 2020; T.k. et al., 2021).

Machine learning and privacy are intertwined concepts. In the analysis or publication of personal data, machine learning techniques are primarily used to protect privacy, increase personal data privacy, and detect disclosure of personal data privacy. On the other hand, machine learning models, such as prediction models for insurance rates or stock prices, are also cases where the privacy of both models and their parameters should be protected (Liu et al., 2022).

A large part of communication in digital environments is done through digital texts. According to a survey published in 2021 (T.k. et al., 2021), more than 18.2 million text messages are transmitted in a minute. Natural Language Processing (NLP) is one of the fundamentals of Artificial Intelligence applications that use computational techniques in order to understand, learn and produce human language content (Hirschberg & Manning, 2015). NLP, also known as Computational Linguistics, focuses on generating technologies to discover the knowledge/wisdom of digital texts like a human does (Chowdhary, 2020). There have been significant improvements in NLP studies over the past 20 years. In the early times of NLP studies, scientists were interested in the automated analysis of linguistic structures like language translation. Today, NLP studies are in a position to make social media analyses that can reveal the user's depression level (Patidar & Umre, 2021), mental health status (Hao et al., 2013), demographic characteristics (Wang et al., 2017), personality (Tay et al., 2020), etc. With the increase in the computing power of computers, the amount of data in digital environments, and advanced models in artificial intelligence techniques, the diversity and success of NLP studies are also increasing. Figure 2 shows the process flow that some of the current text analysis processes. On the one hand, high-performance tools such as Stanford Core NLP (Manning et al., 2015) that can extract syntactic and semantic information in a text are being developed; on the other hand, some resources are still insufficient, especially for low-level languages.

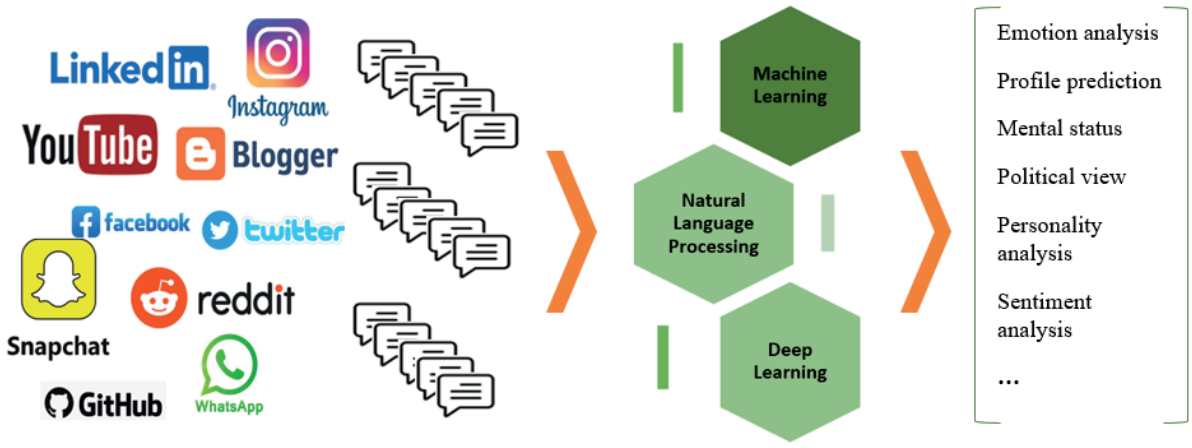


Figure 2. Process flow of some of the current applications through social media and text

NLP research has evolved with the increasing popularity of social media. Social media provide an incredible amount of information about users. Sources from these media, such as Facebook, Twitter, Instagram, LinkedIn, GitHub, Blogs, YouTube, forum sites, and more, are precious inputs for NLP studies. With these inputs, researchers are able to extract relations between social interactions, personal information, and language usage (Ali, 2015). NLP studies are frequently used in studies aimed at protecting or improving the privacy of the owner of the data used, as well as making valuable inferences from the data. Using NLP tools and methods, for example, domain-specific sensitive information can be identified from medical and legal documents and excluded from analyzes to enhance data privacy (Martinelli et al., 2020).

In addition to NLP studies, image processing and voice processing techniques in privacy protection are also among the current study topics. For privacy protection, the detection of sensitive objects in images with deep learning techniques (J. Yu et al., 2017), and the evaluation of voice conversion-based privacy protection against attacks (Lal Srivastava et al., 2020) are some of the studies in this area.

3.2. Big Data and Privacy

Machine learning systems work better as data sets are more extensive. The emergence of big data has contributed significantly to machine learning and data mining techniques, especially those for profiling. According to the UK ICO[2] (Information Commissioner's Office) report, big data analytics, mainly characterized by using machine learning algorithms with new types of data, are frequently reused and can benefit businesses, society, and consumers as citizens. Big data analytics can also help to deliver the public more effective and efficient services and produce positive outcomes that improve people's quality of life.

It is not possible to talk about social analysis without talking about big data. Today, data comes from everywhere: social media sites, sensors, cell phones, sharing sites, e-commerce sites, etc. A daily increase in the vast amount of data is the first indicator of the requirement for big data. Big data is not just a means to evaluate the vast amount of data; it also means to evaluate different properties of data that increase as data increases. In 2001, Laney defined the concept of big data with 3V: high Volume, high Velocity, and high Variety (Laney, 2001). This concept has then widened to 4V (adding high Value) and 5V (adding high Veracity) (Hashem et al., 2015). Big data is an important affair that aims to generate an effective alternative to traditional solutions regarding databases and data analysis (Bello-Orgaz et al., 2016).

Big data has become a significant issue in the field of privacy and machine learning with the emergence of cloud computing. Advances in various big data framework such as Spark (Zaharia et al., 2010) and Hadoop (White, 2012) have increased the use of machine learning application in different fields. Also, the increasing of machine learning libraries of big data such as Mahout (Owen et al., 2011) and SparkMLib (Deshai et al., 2019) has taken machine learning studies one step further. In order to gain more benefits from big data, there are also developed privacy-preserving big data publishing techniques by researchers (Canbay et al., 2019; Zakerzadeh et al., 2015). The vast amount of data with the big data facilities has been used to discover significant knowledge to improve decision-making processes. However, there are still some open problems and challenges, such as the determination of how much data is enough for high-quality data (quantity versus quality) or ensuring enough privacy (security and ownership) (Bello-Orgaz et al., 2016).

Big data analytics generate new knowledge by locating unexpected and previously unknown structures, correlations, and patterns by combining algorithms and information from large and various datasets (Hildebrandt, 2009). Thus, a person's online and offline activities are converted into profiling scores, while predictive algorithms extract personal information to make predictions about individuals' likely actions and behaviors. In summary, activities such as extensive profiling and scoring people based on their profiles with big data analytics and machine learning algorithms are now more suitable than ever to be used by private companies or public authorities (Politou et al., 2021).

4. Technology with Privacy

We live in a digital world, and many applications on the Internet, especially social media applications, are waiting to receive our personal information. These applications, which offer us small rewards or conveniences in return for our personal data, can make millions by processing, selling, or using this data. The general opinion is that people should protect their privacy, but when the practices and observations are examined , it has been shown that users can ignore permanent privacy concerns over their personal information against short-term benefits or some small rewards (Acquisti & Grossklags, 2005). By human nature, our abilities are limited by our bounded rationality (Simon, 2019). We have insufficient memory and processing power to calculate possible implications for the protection and publication of complex, branched data such as our personal information (Acquisti & Grossklags, 2005).

2 https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

Technically, it is not possible to walk around in digital environments without leaving a trace. Even with the most powerful defensive techniques used by the most privacy-conscious users, individuals seem to face great difficulties in avoiding tracking techniques (Acar et al., 2014). Therefore, at this point, there is a need for intelligence applications, responsible technology developers, and governmental regulations that will protect personal privacy. Providing a comprehensive data protection schema considering an application, organization, or government alone is challenging. There needs to be combined countermeasures to enhance privacy protection. While governments make regulations, organizations should employ experts to obey the regulations on data processing or transferring/transactions. Besides, there is a need for applications to help experts to set up and enhance privacy protection.

Many statistical, theoretical, and cryptographic methods have been developed so that personal data can be shared and processed in a way that does not disclose the privacy of the data owners (Churi & Pawar, 2019; Kreso et al., 2021; Majeed & Lee, 2021). The Differential Privacy method is used as one of the most up-to-date and reliable privacy protection methods today. Differential privacy (Dwork, 2008) provides a method that tries to keep the accuracy of the query requested from a database high while minimizing the chance of identifying a query's records. Differential privacy offers solutions to protect the privacy of unstructured data content and to share it with untrusted parties (Zhao & Chen, 2022).

With the proliferation of distributed systems, which is a requirement today because of the amounts of stored and processed data, collaborative learning methods have been generated with respect to privacy protection. One of the popular collaborative learning frameworks is Federated learning (Zhang et al., 2021). Federated learning is a distributed machine learning approach that gives efficiency while providing parallel machine learning model training across many clients. Another popular framework for collaborative learning is Split Learning (Vepakomma et al., 2018). Split learning, another popular distributed machine learning approach, ensures a more private model than Federated learning because the machine learning model architecture is split between server and clients (Thapa et al., 2022).

As mentioned in the previous sections, the privacy risks of the data increase as the data grows. Thus, especially in big data analysis, there is a need for more applications to enhance privacy. In healthcare, the collected data from cyber-physical systems play an essential role in making decisions about the health of humans. The data of these heterogeneous systems are stored in private or public clouds and used for analysis. Many of the existing privacy-preserving data mining and privacy-preserving data publishing techniques do not seem proper to evaluate the unstructured, huge stream of data alone. Using differential privacy, homomorphic encryption, and key-based and clustering-based anonymization techniques alone is not found sufficient to ensure privacy on big data streams. Instead, a hybrid approach with both anonymization and encryption is a good choice (Maleh et al., 2019). Although studies in this field are increasing, data continues to grow, vary, and increase the privacy risk they carry.

5. Changed Privacy Perception (Post-Privacy) and Privacy Paradox

Discussions about the death of privacy have been going on since the mid-1960s. The widespread use of central hosts and monitoring technologies such as cameras and listening devices, which were gradually developing at that time, was the first source of discussion (Schulte, 2018). In the first book declaring the end of privacy, Rosenberg talked about a national computer system in which all citizens' information would be stored and argued that urgent and radical measures should be taken (Rosenberg, 1969). The second wave of privacy concerns escalated with the emergence of the Internet in 1983, the World Wide Web in 1993, and the simultaneous spread of personal computers. With the rapid rise of social media use at the beginning of the 21st century, concerns about "exhibitionist" privacy and sharing personal data with companies have been added to the privacy issue (Schulte, 2018).

Niedzviecki, on the other hand, accepts 2008 as the turning point of a fundamental cultural change and declares that the "age of peeping culture" has begun (Niedzviecki, 2009). As Mark Zuckerberg, the founder of Facebook which is one of the social media platforms, stated in his speech[3] in 2010, privacy is no longer a social norm, and people tend to share more and various information with more people openly. Thus, social media emerged and developed as an area of disclosure and peeping. Figure 3 gives the chronological order of socially melting privacy by the effect of technological and perceptual changes.

3 https://www.theguardian.com/technology/2010/jan/11/facebook-privacy

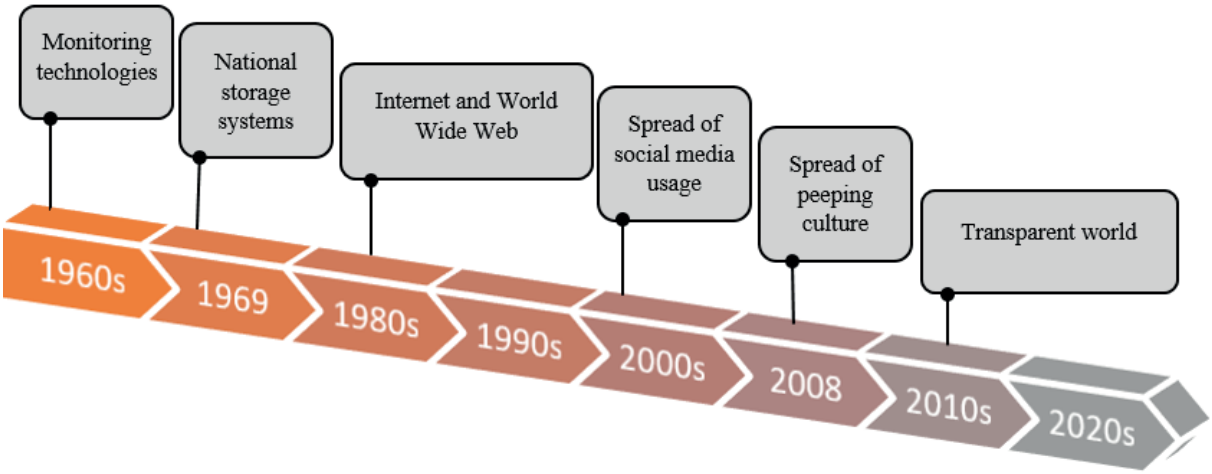


Figure 3. Chronological order of melting privacy by the effect of technological and perceptual changes

Han adapted the definition of Miller's "transparent world" (Miller, 1972) to the 21st century (Han, 2020). He stated that the transparency created by social networks penetrates society irreversibly, and people have to adapt to this new situation. Claiming that we now live in a post-privacy society, Han argues that paying attention to concerns about privacy does not match the realities of the age (Han, 2020). Acquisti et al., on the other hand, argue that sharing more personal data does not always mean development, productivity, and equality. They state that the erosion of privacy threatens the autonomy of individuals (Acquisti et al., 2015).

The complete absence of data sources is neither practical nor desirable. For example, having access to the health records of an unconscious patient brought to the emergency room can save the patient's life. Therefore, keeping data recorded is necessary and essential in some cases. However, it is not cognitively possible for the individual to actively decide on his privacy in his daily life. This cognitive difficulty creates the privacy paradox: people worry about privacy, but in practice, they often do little to protect it (Stalder, 2002). This inconsistency between privacy attitudes and privacy behavior is called the privacy paradox (Norberg et al., 2007).

Privacy calculation theory assumes that individuals make decisions by calculating the potential gain of disclosure and the loss of privacy. For example, Debatin et al. argue that participation in social networks is associated with three needs: the need for entertainment, social connection, and identity construction (Debatin et al., 2009). Therefore, the individual gives up his privacy at the expense of meeting these needs. Also, in many cases, people may lack the cognitive ability or knowledge necessary to make an informed privacy decision. Limited rationality and incomplete information are the main factors determining the privacy decision (Acquisti & Grossklags, 2005). Incomplete and asymmetric information reveals the uncertainty of privacy. It is usual for individuals to be hesitant about sharing information, as they often do not clearly understand what information other people, firms, and governments have about them and how they use that information.

6. Conclusions

Via online activities, people leave easy-to-follow digital trails that reveal who we are, what we eat, where we go, whom we talk to, why we are happy, what we buy, and much more. Those trails are collected and stored somewhere. Once data is collected and stored, we have almost no control over who uses it or how it is used.

While technological developments that benefit humanity are realized in many areas with personal data, private information that even data owners cannot predict can be obtained through analyzing these data. The anonymity of digital environments, the vulnerability of users to attacks, the fact that users do not have enough information about the dangers of these environments, and perhaps most importantly, the inevitable sharing of information in these environments have increased privacy concerns.

Although informing the data owner and obtaining consent is presented as a solution in many studies, it is clear that this cannot prevent the risks of privacy violations in reality. Almost every step in the digital world is followed by applications

such as cookies. Even with the strongest protective measures taken by the most privacy-conscious people, it is almost impossible to stay away from these tracings. Especially without loss of content or functionality in digital media, it is challenging to prevent monitoring. Once tracing has occurred, it is almost impossible to start from a truly clean profile.

In the analysis or publication of personal data, big data, and machine learning techniques are mostly used to protect or enhance personal data privacy. However, the current technological countermeasures are not complete enough for unstructured data such as texts, images, videos, or voices. On the other hand, the privacy risk of data increases as the volume and variety of data continue to grow. Measures for privacy protections are not evolving as quickly as threats to violate privacy.

In the cyber/digital world, countries and institutions have to take measures to protect data privacy. While there is a need for additional resources to explain the necessary practices and frameworks in countries that have enacted measures such as GDPR, many countries still have not announced any legal regulations.

When an evaluation is made together with (1) today's technological developments, (2) individuals' enthusiasm and obligation to share information, (3) the economic and social value of data, and (4) the measures taken by governments against privacy violations, the theory of the death of privacy is increasingly valid from the 2000s. This theory, which was first discussed in the mid-1960s, is in a position that cannot be ignored today. In this digital age, transparency is at the forefront, and individuals are willing to accept temporary benefits in return for their personal data. For these reasons, there is an urgent need for lawmakers and technology developers to resort to advanced measures before privacy violations lead to more significant problems.

Peer-review: Externally peer-reviewed.
Author Contributions: Conception/Design of Study- P.C., Z.D.; Data Acquisition- P.C.; Data Analysis/Interpretation- P.C., Z.D.; Drafting Manuscript- P.C.; Critical Revision of Manuscript- Z.D.; Final Approval and Accountability- P.C., Z.D.
Conflict of Interest: The authors have no conflict of interest to declare.
Grant Support: This work was supported by a grant from the Kahramanmaras Sutcu Imam University Scientific Research Projects Unit, Project Number: 2021/2-38 M.

REFERENCES

Acar, G., Eubank, C., Englehardt, S., Juarez, M., Narayanan, A., & Diaz, C. (2014). The web never forgets: Persistent tracking mechanisms in the wild. *Proceedings of the ACM Conference on Computer and Communications Security*. https://doi.org/10.1145/2660267.2660347

Acquisti, A., & Grossklags, J. (2005). Privacy and rationality in individual decision making. In *IEEE Security and Privacy* (Vol. 3, Issue 1). https://doi.org/10.1109/MSP.2005.22

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. In *Science* (Vol. 347, Issue 6221). https://doi.org/10.1126/science.aaa1465

Ali, D. (2015). Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, Github, and More, by Matthew A. Russell. *Journal of Information Privacy and Security*, *11*(2). https://doi.org/10.1080/15536548.2015.1046287

Balcı, E. (2019). Overview of Intelligent Personal Assistants. *Acta INFOLOGICA*, *3*(1). https://doi.org/10.26650/acin.571303

Bauman, Z., & Lyon, D. (2013). Liquid Surveillance. In *Statewide Agricultural Land Use Baseline 2015* (Vol. 1).

Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. *Information Fusion*, *28*. https://doi.org/10.1016/j.inffus.2015.08.005

Bennett, L. (2009). Reflections on privacy, identity and consent in on-line services. *Information Security Technical Report*, *14*(3). https://doi.org/10.1016/j.istr.2009.10.003

Canbay, Y., Vural, Y., & Sagiroglu, S. (2019). Privacy Preserving Big Data Publishing. *International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism, IBIGDELFT 2018 - Proceedings*. https://doi.org/10.1109/IBIGDELFT.2018.8625358

Carey, P., & Acquisti, A. (2018). Data protection: a practical guide to UK and EU law. In *Economics of Information Security*.

Chowdhary, K. R. (2020). Fundamentals of artificial intelligence. In *Fundamentals of Artificial Intelligence*. https://doi.org/10.1007/978-81-322-3972-7

Churi, P. P., & Pawar, A. v. (2019). A systematic review on privacy preserving data publishing techniques. In *Journal of Engineering Science and Technology Review* (Vol. 12, Issue 6). https://doi.org/10.25103/jestr.126.03

Confessore, N. (2018). Cambridge Analytica and Facebook: The Scandal and the Fallout So Far. *The New York Times*.

Correa, T., Hinsley, A. W., & de Zúñiga, H. G. (2010). Who interacts on the Web?: The intersection of users' personality and social media use. *Computers in Human Behavior*, *26*(2). https://doi.org/10.1016/j.chb.2009.09.003

Debatin, B., Lovejoy, J. P., Horn, A. K., & Hughes, B. N. (2009). Facebook and online privacy: Attitudes, behaviors, and unintended consequences.

Journal of Computer-Mediated Communication, *15*(1). https://doi.org/10.1111/j.1083-6101.2009.01494.x

Deshai, N., Sekhar, B. V. D. S., & Venkataramana, S. (2019). Mllib: machine learning in apache spark. *International Journal of Recent Technology and Engineering*, *8*(1).

Dwork, C. (2008). Differential privacy: A survey of results. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *4978 LNCS*. https://doi.org/10.1007/978-3-540-79228-4_1

Fuchs, C. (2014). Social Media: A Critical Introduction. In *Social Media: A Critical Introduction*. https://doi.org/10.4135/9781446270066

Han, B.-C. (2020). The Transparency Society. In *The Transparency Society*. https://doi.org/10.1515/9780804797511

Hao, B., Li, L., Li, A., & Zhu, T. (2013). Predicting mental health status on social media a preliminary study on microblog. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8024 LNCS*(PART 2). https://doi.org/10.1007/978-3-642-39137-8_12

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Ullah Khan, S. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems*, *47*, 98–115. https://doi.org/10.1016/j.is.2014.07.006

Hildebrandt, M. (2009). Who is Profiling Who? Invisible Visibility. In *Reinventing Data Protection?* https://doi.org/10.1007/978-1-4020-9498-9_14

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. In *Science* (Vol. 349, Issue 6245). https://doi.org/10.1126/science.aaa8685

Jones, M. L., & Kaminski, M. E. (2021). AN AMERICAN'S GUIDE TO THE GDPR. *Denver Law Review*, *98*(1).

Joshi, A. v. (2020). *Machine Learning and Artificial Intelligence*. Springer International Publishing. https://doi.org/10.1007/978-3-030-26622-6

Kelleher, J. D., & Tierney, B. (2018). Data Science. In *Data Science*. https://doi.org/10.7551/mitpress/11140.001.0001

Koops, B.-J. (2013). Forgetting Footprints, Shunning Shadows: A Critical Analysis of the "Right to Be Forgotten" in Big Data Practice. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1986719

Kreso, I., Kapo, A., & Turulja, L. (2021). Data mining privacy preserving: Research agenda. In *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* (Vol. 11, Issue 1). https://doi.org/10.1002/widm.1392

Lal Srivastava, B. M., Vauquier, N., Sahidullah, M., Bellet, A., Tommasi, M., & Vincent, E. (2020). Evaluating Voice Conversion-Based Privacy Protection against Informed Attackers. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *2020-May*. https://doi.org/10.1109/ICASSP40776.2020.9053868

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*, *6*(70).

Liu, B., Ding, M., Shaham, S., Rahayu, W., Farokhi, F., & Lin, Z. (2022). When Machine Learning Meets Privacy. *ACM Computing Surveys*, *54*(2), 1–36. https://doi.org/10.1145/3436755

Majeed, A., & Lee, S. (2021). Anonymization Techniques for Privacy Preserving Data Publishing: A Comprehensive Survey. *IEEE Access*, *9*. https://doi.org/10.1109/ACCESS.2020.3045700

Maleh, Y., Shojafar, M., Darwish, A., & Haqiq, A. (2019). Cybersecurity and Privacy in Cyber-Physical Systems. In *Cybersecurity and Privacy in Cyber-Physical Systems*. https://doi.org/10.1201/9780429263897

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2015). *The Stanford CoreNLP Natural Language Processing Toolkit*. https://doi.org/10.3115/v1/p14-5010

Martinelli, F., Marulli, F., Mercaldo, F., Marrone, S., & Santone, A. (2020). Enhanced Privacy and Data Protection using Natural Language Processing and Artificial Intelligence. *Proceedings of the International Joint Conference on Neural Networks*. https://doi.org/10.1109/IJCNN48605.2020.9206801

Marwick, A. E. (2012). The public domain: Social surveillance in everyday life. *Surveillance and Society*, *9*(4). https://doi.org/10.24908/ss.v9i4.4342

Matarić, M. J., Eriksson, J., Feil-Seifer, D. J., & Winstein, C. J. (2007). Socially assistive robotics for post-stroke rehabilitation. *Journal of NeuroEngineering and Rehabilitation*, *4*. https://doi.org/10.1186/1743-0003-4-5

Miller, A. R. (1972). *Computers, Data Banks and Individual Privacy: An Overview*. Colum. Hum. Rts. L. Rev. 4.

Mittelstadt, B. D., & Floridi, L. (2016). The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. In *Science and Engineering Ethics* (Vol. 22, Issue 2). https://doi.org/10.1007/s11948-015-9652-2

Niedzviecki, H. (2009). *The peep diaries: How we're learning to love watching ourselves and our neighbors*. City Lights Publishers.

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs*, *41*(1). https://doi.org/10.1111/j.1745-6606.2006.00070.x

Owen, S., Anil, R., Dunning, T., & Friedman, E. (2011). Mahout in Action. In *Online*.

Patidar, H., & Umre, J. (2021). Predicting depression level using social media posts. *International Journal of Research -GRANTHAALAYAH*, *8*(12). https://doi.org/10.29121/granthaalayah.v8.i12.2020.1972

Politou, E., Alepis, E., Virvou, M., & Patsakis, C. (2021). Privacy and Data Protection Challenges in the Distributed Era. In *Learning and Analytics in Intelligent Systems*.

Rosenberg, J. M. (1969). *The Death of Privacy*. Random House (NY).

Rosenthal, S., Biswas, J., & Veloso, M. (2010). An effective personal mobile robot agent through symbiotic human-robot interaction. *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, *2*.

Schulte, P. (2018). David Vincent, Privacy. A Short History. Cambridge, Polity Press 2016. *Historische Zeitschrift*, *307*(2). https://doi.org/10.1515/hzhz-2018-1399

Simon, H. A. (2019). Models of Bounded Rationality. In *Models of Bounded Rationality*. https://doi.org/10.7551/mitpress/4711.001.0001

Stalder, F. (2002). Opinion. Privacy is not the antidote to surveillance. In *Surveillance and Society* (Vol. 1, Issue 1). https://doi.org/10.24908/ss.v1i1.3397

Tay, L., Woo, S. E., Hickman, L., & Saef, R. M. (2020). Psychometric and Validity Issues in Machine Learning Approaches to Personality Assessment: A Focus on Social Media Text Mining. *European Journal of Personality*, *34*(5). https://doi.org/10.1002/per.2290

Thapa, C., Mahawaga Arachchige, P. C., Camtepe, S., & Sun, L. (2022). SplitFed: When Federated Learning Meets Split Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, *36*(8), 8485–8493. https://doi.org/10.1609/aaai.v36i8.20825

T.k., B., Annavarapu, C. S. R., & Bablani, A. (2021). Machine learning algorithms for social media analysis: A survey. In *Computer Science Review* (Vol. 40). https://doi.org/10.1016/j.cosrev.2021.100395

Vepakomma, P., Gupta, O., Swedish, T., & Raskar, R. (2018). Split learning for health: Distributed deep learning without sharing raw patient data. *ArXiv Preprint ArXiv:1812.00564*.

Voigt, P., & von dem Bussche, A. (2017). The EU General Data Protection Regulation (GDPR) A Practical Guide. In *The EU General Data Protection Regulation (GDPR)*.

Wang, Q., Ma, S., & Zhang, C. (2017). Predicting users' demographic characteristics in a Chinese social media network. *Electronic Library*, *35*(4). https://doi.org/10.1108/EL-09-2016-0203

White, T. (2012). Hadoop: The definitive guide 4th Edition. *Online*, *54*. https://doi.org/citeulike-article-id:4882841

Yu, J., Zhang, B., Kuang, Z., Lin, D., & Fan, J. (2017). IPrivacy: Image Privacy Protection by Identifying Sensitive Objects via Deep Multi-Task Learning. *IEEE Transactions on Information Forensics and Security*, *12*(5). https://doi.org/10.1109/TIFS.2016.2636090

Yu, S. (2016). Big Privacy: Challenges and Opportunities of Privacy Study in the Age of Big Data. *IEEE Access*, *4*. https://doi.org/10.1109/ACCESS.2016.2577036

Zaharia, M., Chowdhury, M., Franklin, M. J., Shenker, S., & Stoica, I. (2010). Spark: Cluster computing with working sets. *2nd USENIX Workshop on Hot Topics in Cloud Computing, HotCloud 2010*.

Zakerzadeh, H., Aggarwal, C. C., & Barker, K. (2015). Privacy-preserving big data publishing. *Proceedings of the 27th International Conference on Scientific and Statistical Database Management*, 1–11. https://doi.org/10.1145/2791347.2791380

Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., & Gao, Y. (2021). A survey on federated learning. *Knowledge-Based Systems*, *216*. https://doi.org/10.1016/j.knosys.2021.106775

Zhao, Y., & Chen, J. (2022). A Survey on Differential Privacy for Unstructured Data Content. *ACM Computing Surveys*. https://doi.org/10.1145/3490237

Zhu, L. (2011). Privacy in Context: Technology, Policy, and the Integrity of Social Life. *Journal of Information Privacy and Security*, *7*(3). https://doi.org/10.1080/15536548.2011.10855919

ACTA INFOLOGICA 2023;7(1):165-172
dergipark.org.tr/acin

DOI: 10.26650/acin.1282824

RESEARCH ARTICLE

Malware Detection in Forensic Memory Dumps: The Use of Deep Meta-Learning Models

Adli Bellek Dökümlerinde Kötü Amaçlı Yazılım Tespiti: Derin Meta Öğrenme Modellerinin Kullanılması

Yalçın Özkan[1]

[1](Assist. Prof. Dr.) Istinye University, Faculty of Economics, Administrative and Social Sciences, Department of Management Information Systems Istanbul, Turkiye

ORCID: Y.Ö. 0000-0002-3551-7021

Corresponding author:
Yalçın ÖZKAN
Istinye University, Faculty of Economics, Administrative and Social Sciences, Department of Management Information Systems Istanbul, Turkiye
E-mail address: yalcin.ozkan@istinye.edu.tr

Submitted: 13.04.2023
Revision Requested: 28.04.2023
Last Revision Received: 29.04.2023
Accepted: 05.05.2023
Published Online: 20.06.2023

Citation: Ozkan, Y. (2023). Malware detection in forensic memory dumps: The use of deep meta-learning models. *Acta Infologica, 7*(1), 165-172. https://doi.org/10.26650/acin.1282824

ABSTRACT

The present study aimed to design a high-performance deep meta-learning model that could be utilized in classification predictions using forensic memory datasets and propose a framework that would ensure the generalization and consistency of the predictions with the help of this model. To achieve this aim, a dataset containing malware and obtained from forensic memory dumps was addressed. First, it was subjected to the classification process with a deep learning algorithm, and a predictive model was acquired. The predictive model was found to have an accuracy metric of 98.25%. In addition to this finding, a meta-learning model consisting of five different models with the same hyperparameters was created. The accuracy of the obtained meta-model was computed as 97.69%. With the thought that this model would reduce the prediction variance and thus the predictive model could be generalized, it was ensured to be run 5 times in a row. As a result of this process, the prediction variance, indicating a very small change, was calculated as 0.000012. Accordingly, considering the acquired performance value, it can be determined that high performance is achieved in malware detection, and thus what hyperparameters ensure success can be revealed. If deep learning methods are used as a single model, the problem is that the variance between the predictions is large due to its stochastic structure. To avoid such drawbacks, a deep meta-learning model using the same parameters was designed instead of a deep learning model comprising a single model, and considerably smaller variance values were achieved, thus providing generalized and consistent predictions.
Keywords: Forensic memory, cyber security, deep learning, meta-learning

ÖZ

Bu çalışmada adli bellek veri kümelerinden yararlanılarak, sınıflandırma öngörülerinde kullanılabilecek yüksek performanslı bir derin meta öğrenme modelinin tasarlanması ve bu model yardımıyla öngörülerin genelleştirme ve tutarlılığını sağlayacak bir çerçevenin önerilmesi amaçlanmaktadır. Bu amaca ulaşabilmek için, kötü amaçlı yazılımları içeren ve adli bellek dökümlerinden elde edilen bir veri kümesi ele alınarak önce derin öğrenme algoritması ile sınıflandırma sürecine tabi tutuldu ve bir öngörü modeli elde edildi. Öngörü modelinin %98,25lik bir doğruluk metriğine sahip olduğu görülmüştür. Bu bulgunun yanı sıra, aynı hiper parametrelere sahip 5 ayrı modelden oluşan bir meta öğrenme modeli oluşturulmuştur. Elde edilen meta modelin doğruluğu %97,69 olarak hesaplandı. Bu modelin öngörü varyansını azaltacağı ve böylece öngörü modelini genelleştirilebileceği düşüncesiyle ardı ardına 5 kez çalıştırılması sağlandı. Bu işlem sonucunda çok küçük bir değişime işaret eden öngörü varyansı 0,000012 olarak hesaplandı. Sonuç olarak, elde edilen performans değeri göz önüne alındığında, kötü amaçlı yazılım tespitinde yüksek bir performansın elde edildiği ve böylece başarıyı sağlayan hiper parametrelerin neler olduğu belirlenebilmektedir. Derin öğrenme yöntemlerinin tek model olarak kullanılması durumunda, stokastik bir yapıya sahip olması nedeniyle öngörüler arasındaki varyansın büyük olması sorunuyla karşılaşılmaktadır. Bu tür sakıncaları önlemek üzere, tek modelden oluşan derin öğrenme modeli yerine, aynı parametreleri kullanan bir derin meta öğrenme modeli tasarlanarak çok daha küçük varyans değerlerine ulaşılmış, böylece genelleştirilmiş ve tutarlı öngörüler üretilmesi sağlanmıştır.
Anahtar Kelimeler: Adli bellek, siber güvenlik, derin öğrenme, meta öğrenme

1. INTRODUCTION

Installing malware in a computer's memory significantly damages computer systems and information security. Hence, it is understood that detecting malware and eliminating threats are important to prevent the hacking of computer user data, credentials, and other important information. To this end, memory analysis can enable the analysis of volatile data in a computer's memory. As with hard drive data, these data can be accessed by forensic memory analysis to investigate and identify attacks or malicious behaviors that do not leave easily detectable traces (Sihwail & Omar & Zainol, 2021). Viruses can be listed among examples of malware. Furthermore, trojans, worms, and spyware can be added to this category. Among them, computer viruses can cause significant destruction by destroying files on infected computers. If the attacker's goal is to collect data from computers, he may aim to reach financial information, such as bank and credit cards, by infecting the target computers with malware (Christensson, 2006).

Studies that address threats and propose solutions in this regard demonstrate that machine learning techniques can be used to detect malware. Various studies stress that machine learning techniques provide numerous advantages and these techniques can produce faster, more accurate, and more effective results than conventional attack prevention methods. However, it can be indicated that these techniques pose some other difficulties. For example, it has been stated that difficulties such as data quality, data size, data integrity, and other ethical and legal issues can hinder the effective use of machine learning techniques (Qadir & Noor).

Sihwail et al. (2019) used datasets acquired on the basis of memory analyses to detect malware, and classification was made through machine learning. It was found that a performance level of 98.5% was achieved by applying the support vector machine algorithm, considered among the conventional classification methods. Additionally, a false positive rate, revealing false alarms, was obtained as 1.7%. However, due to the limitations of conventional classification methods, the tendency to turn to the advantages of deep learning algorithms has begun. Deep learning algorithms in particular, such as convolutional neural networks (CNN) and long short-term memory (LSTM), can be preferred to classify memory dumps and extract important information (Yang et al., 2021; Karamitsoz et al., 2020).

Both traditional and deep learning models can be preferred in the analysis of memory dumps. In a study conducted in 2022, Dener et al. (2022) made a classification using the Random Forest, Decision Tree, Gradient Boosted Tree, Logistic Regression, Naive Bayes, Linear Vector Support Machine, Multilayer Perceptron, Deep Feed Forward Neural Network, and Long Short-Term Memory (LSTM) algorithms and compared their performances. As a result of this classification study, it was seen that the performance of machine learning algorithms approached 99.97%.

In machine learning, meta-learning refers to learning algorithms that learn from other learning algorithms. Meta-learning means the use of machine learning algorithms that learn how best to combine predictions from other machine learning algorithms in the ensemble learning field (Brownlee, 2021). Another purpose of meta-learning is to train a model on various learning tasks so that it can solve new learning tasks using a small number of training examples (Finn et al., 2017). In the present study, a framework on how to acquire both high-performance and generalized and consistent predictions with deep learning and deep meta-learning methods is proposed by considering memory dump data.

2. MATERIALS AND METHODS

2.1. Deep learning networks

Artificial neural networks are among the machine learning methods inspired by biological neural networks and used in solving numerous problems. Artificial neural networks are trained on datasets and perform operations such as classification and prediction development by identifying patterns among data. Artificial neural networks are preferred in machine learning, particularly to obtain high-performance predictions. The artificial neural network model consists of three interconnected layers, called input, hidden, and output. Such networks contain interconnected neural chains forming the neural architecture. If the number of hidden layers exceeds one, the artificial neural network is called a "deep learning network." As seen in a very small size example in Figure 1, a deep neural network can consist of an input, two outputs, and three hidden layers.

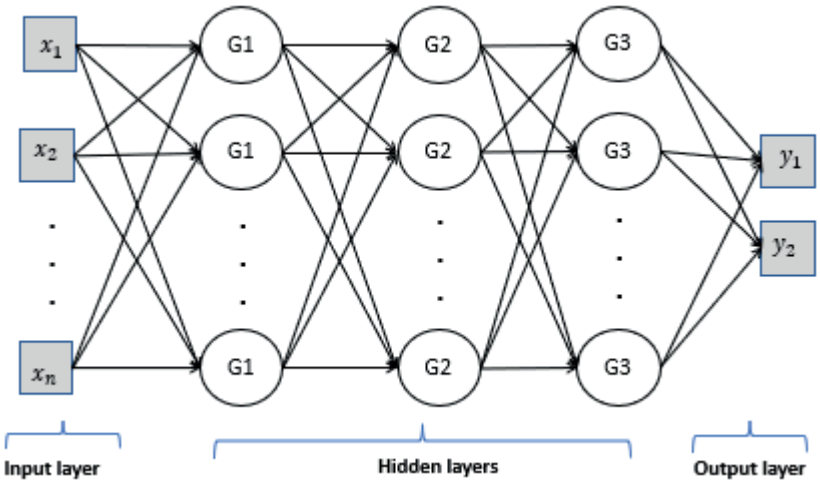


Figure 1. Deep learning classification network

The neurons in deep learning networks are formed as a result of the interconnection of the nerves indicated in Figure 2. A nerve is in contact with all neurons in the next layer. Within the framework of this connection, there is data transfer between the nerves. Every relationship between neurons is provided by numerical values called "weights." Each connection has a numerical weight. Weights are parameters that the neurons in neural networks use when performing a process. The weights of a neuron multiplied by the input signal are an important factor determining the neuron's output signal. During the training of the network, it is essential to set the weights correctly in order to improve the network's performance. To generalize, the input value of a nerve is obtained with the help of equation (1). While calculating the inputs, the used $w_1, w_2, \dots, w_n$ values denote the weights, and $x_1, x_2, \dots, x_n$ values represent the input data. The expression b_1 in the equation is considered the "threshold value."

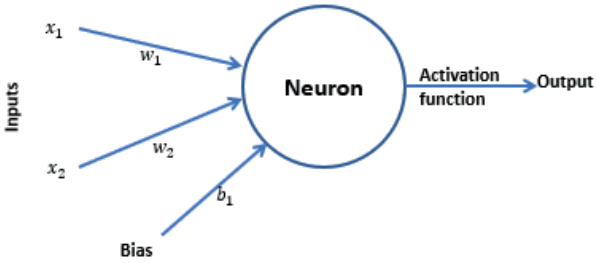


Figure 2. The behavior of a neuron in forward propagation

$$Input = w_1x_1 + w_2x_2 + \cdots + w_nx_n + b_1 \quad (1)$$

The output value is calculated after the total input values for a neuron are computed. The outputs can be expressed as a sigmoid activation function using the inputs, as specified in equation (2), and, thus, a conversion operation is carried out.

$$Output = \frac{1}{1+e^{-Input}} \quad (2)$$

The specified operations are calculated one by one by following the connections for each neuron. Thus, "forward propagation" on the network is performed. Forward propagation is the calculation of outputs by processing the input data of the network over the connections between the neurons in the layers. "Backpropagation" ensures that the weights are updated to minimize the error function of the network.

Backpropagation is performed based on an error function. The error function, also known as the loss function, is expressed as L_{total} and measures the difference between the outputs produced by the network and the actual outputs. The backpropagation method is employed to minimize this error function by changing the weights and threshold values. The error function

measures how close the network is to its intended output. The said function is among the most important factors that determine the targeted performance in network training. The selection of the error function significantly impacts the success of the network in the training process. Mean squared error, cross-entropy, and log loss are among the most common loss functions. After the error function value is computed, the total loss amount is distributed to all weights in the network. To this end, in order to determine the effect of the change to be made in each weight w_i on the total error, its derivatives are calculated according to the mentioned weight and multiplied by η, the learning rate, and thus the amount of change is calculated. This amount is updated by subtracting it from the previous weight, as seen in formula (3).

$$w_{\mathrm{i}}^{new} \longleftarrow w_i - \eta \frac{\partial L_{total}}{\partial w_{\mathrm{i}}} \qquad (\mathrm{i} = 1,2,\ldots,\mathrm{n}) \tag{3}$$

2.2. Deep meta-learning

Deep learning neural networks have a nonlinear structure. They offer more flexibility and can be scaled proportionally to the amount of training data available. However, a disadvantage of this flexibility is that they learn through a stochastic training algorithm. The use of stochastic models such as deep learning means that they are sensitive to the characteristics of the training data and can find a different weight set each time they are trained, and accordingly, they can produce different predictions. Since the prediction variance acquired in deep learning models is high, this variance can be reduced using "meta-learning" methods (Brownlee, J. 2021).

A deep learning-based meta-learning model can be created, as shown in Figure 3, for classification purposes. To this end, the raw dataset is first passed through the preprocessing stage. Afterwards, the test and training datasets are randomly generated. With the training data, a model is obtained in accordance with the deep learning algorithm, and the accuracy value is computed using this model together with the test data. Following the preprocessing stage, a meta-learning model comprising n models is obtained. To find the performance of meta-learning with n models obtained, datasets that consist of the predictive values of each model constituting it are used. These data are combined, and the final performance is acquired using another machine learning model.

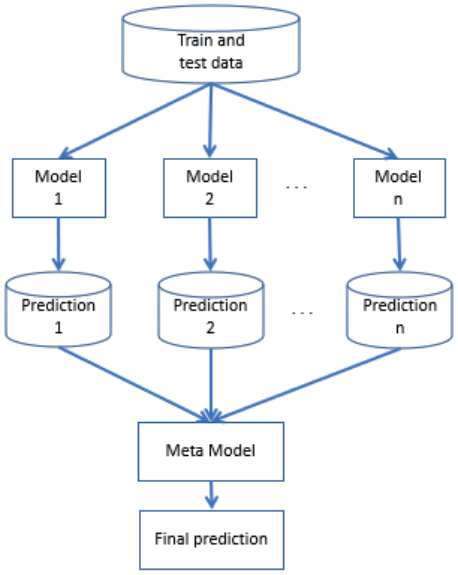


Figure 3. Meta-learning model

2.3. Memory analysis

Memory analysis is generally based on the principle of obtaining the current state of system memory as a snapshot file, also known as a "memory dump." The mentioned capture process is performed by running special software. This file obtained can be later moved out of the system, and analyses can be conducted on it. Support is received from analyzers to perform such operations. The said software can convert data to a CSV file. In this way, it becomes possible to apply machine learning

algorithms on the obtained files. Analyzers not only capture malware footprints but also have additional features that can be used to extract the hidden original code (Lashkari et al., 2020).

3. RESULTS

3.1. Dataset

A dataset must first be provided, and a model must be trained with this dataset to create predictive models in machine learning applications. To achieve this, a dataset CIC-MalMem-2022 was used in the current study (UNB, 2023). The dataset was balanced to comprise 50% malicious memory dumps and 50% benign memory dumps. The table below contains a breakdown of the malware families included in the dataset. The dataset contains a total of 58,596 records, of which 29,298 are normal and 29,298 are malicious (Carrier et al., 2022). Table 1 lists the malware included in the dataset.

Table 1
Distribution of malware in the dataset

Software category	Software families	Number of observations
Trojan	Zeus	1950
	Emotet	1967
	Refroso	2000
	Scar	2000
	Reconyc	1570
Spyware	180Solutions	2000
	CoolWebSearch (CWS)	2000
	Gator	2200
	Transponder	2410
	TIBS	1410
Ransomware	Conti	1988
	MAZE	1958
	Pysa	1717
	Ako	2000
	Shade	2128
Total		29298

3.2. Preprocessing

Preprocessing activities were carried out before proceeding to creating the deep learning model. In this regard, one attribute that was thought to be not useful in model creation was extracted from the dataset, and spaces in the variable names were eliminated. To apply the classification model on the data set, 70% of the total data was separated as training data, and the remaining 30% was separated as test data using the random sampling method. Thus, 41017 observations were allocated for the training of the deep learning model, and 17579 observations were allocated for testing. In this study, data balancing processes were not applied since an equality was provided in terms of the distribution of classes. Of the observations reserved for training, 20516 consist of normal control observations, whereas the remaining 20501 consist of malware observations. Normal observations are labeled “Benign” and malware observations are labeled “Malware”.

3.3. Deep learning model architecture

The stage of designing the deep learning model is proceeded to after the preprocessing steps are completed. At the said stage, the layers that would form the model were prepared in the Python environment in line with the principles determined by the Keras library (Chollet et al., 2015) located on the TensorFlow website (Abadi et al., 2014). In this model, two hidden layers were determined apart from the input and output layers. As seen in Figure 4, the nerve numbers and activation functions of each layer were identified. The “softmax” activation function was selected in the output layer, and the “relu” activation function was selected in all other layers.

The deep learning architecture reveals the structure where the deep learning algorithm will be applied. The training model was identified in accordance with this structure. In this application, the “Keras” library was used for model training. Eighty epochs were applied in the development of the model. Of the training data, 20% was reserved for the validation process to

be used during model creation. While designing the model, “categorical_crossentropy” was selected as the loss function, “Adam” was selected as the optimizer, and the “accuracy” parameter was selected as a metric.

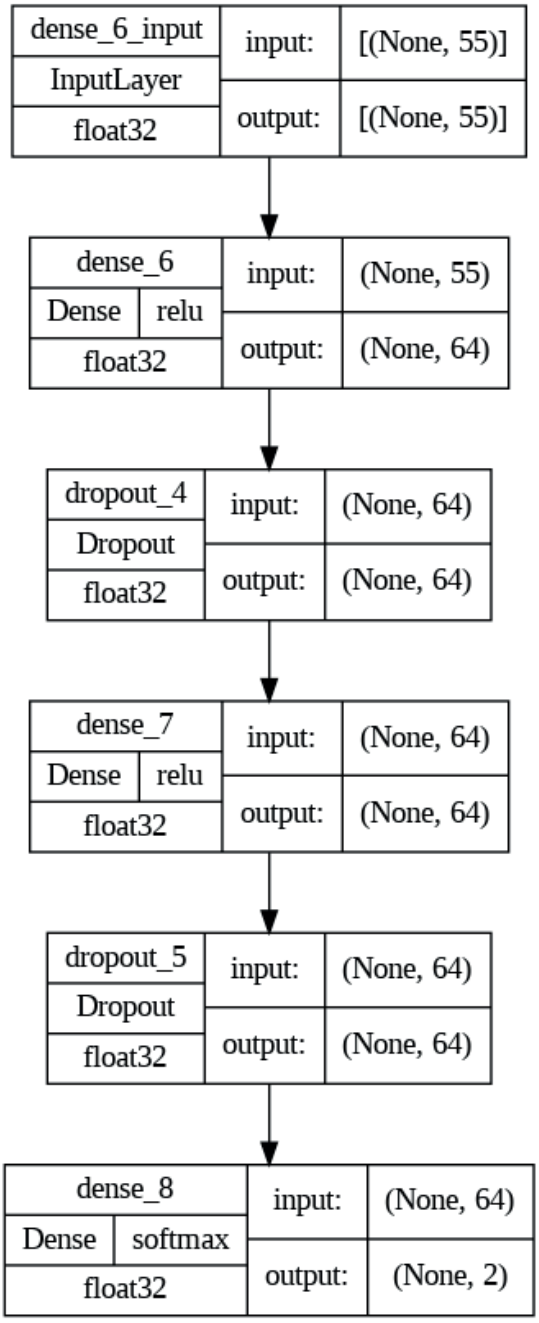
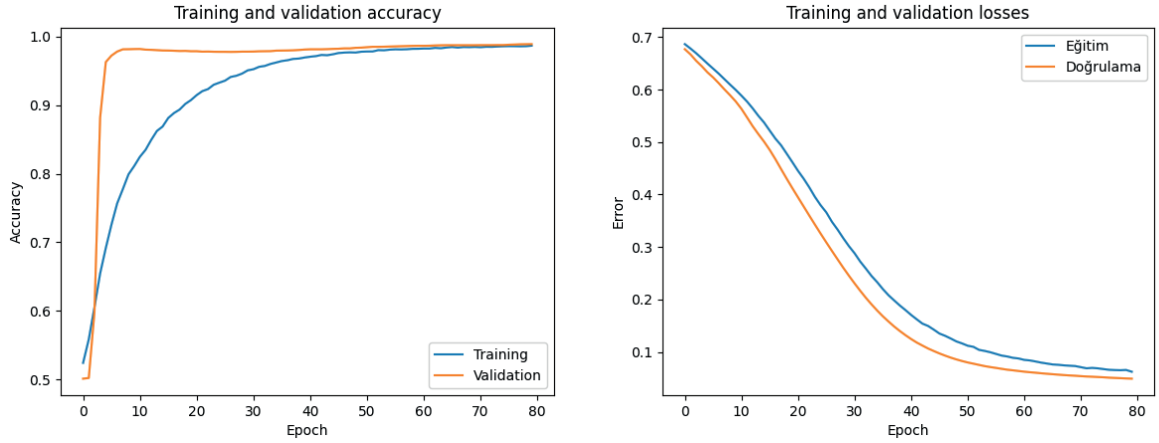


Figure 4. Deep learning architecture

3.4. Model performance

The model was started to be trained after the deep learning architecture in Figure 4 was established and its parameters were identified. At this stage, it was determined that the model initially encountered an “overfitting” problem after a few iterations (epochs). The learning rate of the model was reduced to 0.000001 to solve this problem. During the model’s training, the performance values and error amounts acquired in each iteration were computed, and the graphs displayed in Figure 5 were drawn. Changes in the model’s performance can be observed in the first graph (a), while the change in losses can be observed in graph (b). The classification performance was calculated as 98.25% using the test data of the trained model.



(a) (b)

Figure 5. Performance of the training model and error distribution.

3.5. Meta-learning model

The meta-model addressed in this study was prepared based on the components in Figure 3. To acquire the meta-model, a deep learning model was first prepared. Afterwards, an ensemble of 5 deep learning models with the same hyperparameters was prepared. Predictions were obtained using the test data for each ensemble member. These predictions were combined and used as input for a regression model. In conclusion, the performance of the regression model was accepted as the performance of the deep meta-model. The performance of the first meta-model was calculated as 0.9769 in terms of accuracy. Furthermore, the variance between these predictive performance values was found to be 0.0008. The "Model 1" column in Table 2 explains these operations.

By repeating the same operations, the meta-learning model was run 4 more times, and similar operations were specified in separate columns for model 2, model 3, model 4, and model 5. The purpose of these model iterations is to compare the prediction variances of the final model obtained at different times. The meta-model performance was computed as 0.9764 for model 2, 0.9771 for model 3, 0.9696 for model 4, and 0.9780 for model 5. The variance between the performances acquired as a result of running the deep meta-models 5 times was 0.000012.

Table 2
Deep meta-learning models and performance values for each element

Model elements	**Meta 1**	**Meta 2**	**Meta 3**	**Meta 4**	**Meta 5**
1	0.9925	0.9864	0.9445	0.9842	0.9634
2	0.9209	0.9858	0.9143	0.9280	0.9916
3	0.9683	0.9841	0.9662	0.9917	0.9899
4	0.9859	0.9299	0.9685	0.9020	0.9670
5	0.9890	0.9745	0.9885	0.9852	0.9896
Inter-model variance	0.0008	0.0006	0.0008	0.0016	0.0002
Meta-model performance	0.9769	0.9764	0.9771	0.9696	0.9780

4. CONCLUSION

This study aimed to develop a high-performance and generalized classification framework by considering a dataset that included normal and malware observations. Initially, a deep learning model was developed based on the forensic memory dataset acquired from memory dumps, and a 98.25% accuracy value was obtained as predictive performance. Although this is a high-performance value, lower or higher performance values can be obtained in the next run. Because of the stochastic nature of deep learning models, different results may be encountered at each run of the developed model.

A meta-learning model was developed in the present study in order to have low variance between prediction performances, in other words, to obtain generalized predictions. Instead of calculating the model performance based on a single model, a common performance acquired from an ensemble of five models was used, and a meta-learning architecture developed to this end was applied. The prediction variances of the obtained meta-model were computed as 0.000012. This result is considerably smaller than the variance of each of model 1, model 2, model 3, model 4, and model 5 meta-models, showing that the developed deep meta-model framework can be used to obtain high-performance and low-variance generalized predictive models.

Peer-review: Externally peer-reviewed.
Conflict of Interest: The author has no conflict of interest to declare.
Grant Support: The author declared that this study has received no financial support.

REFERENCES

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... & Zheng, X. (2016). TensorFlow: Large-scale machine learning on heterogeneous distributed systems. In 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI) (pp. 265-283).
2. Brownlee, J. (2021, April 26). *What is meta-learning in machine learning?* MachineLearningMastery.com. Retrieved from https://machinelearningmastery.com/meta-learning-in-machine-learning/.
3. Carrier, T., Victor, P., Tekeoglu, A., & Lashkari, A. (2022). *Detecting obfuscated malware using memory feature engineering.* Proceedings of the 8th International Conference on Information Systems Security and Privacy. https://doi.org/10.5220/0010908200003120.
4. Chollet, F., & others. (2015). *Keras*. GitHub. Retrieved from https://github.com/fchollet/keras
5. Christensson, P. (2022, Nov 19). *Malware Definition*. Retrieved from https://techterms.com
6. Dener, M., Ok, G., & Orman, A. (2022). *Malware detection using memory analysis data in Big Data Environment*. Applied Sciences, 12(17), 8604, https://doi.org/10.3390/app12178604.
7. Finn C., Abbeel P., & Levine S. (2017). *Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks.* https://arxiv.org/abs/1703.03400.
8. Karamitsos, I., Afzulpurkar, A. & Trafalis, T. (2020) *Malware Detection for Forensic Memory Using Deep Recurrent Neural Networks*. Journal of Information Security, 11, 103-120. doi: 10.4236/jis.2020.112007.
9. Lashkari, A. H., Li, B., Carrier, T. L., & Kaur, G. (2021). *VolMemLyzer: Volatile memory analyzer for malware classification using feature engineering.* 2021 Reconciling Data Analytics, Automation, Privacy, and Security: A Big Data Challenge (RDAAPS). https://doi.org/10.1109/rdaaps48126.2021.9452028.
10. Qadir, S., & Noor, B. (2021). *Applications of machine learning in Digital Forensics.* 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2). https://doi.org/10.1109/icodt252288.2021.9441543.
11. Sihwail, R., Omar, K., & Ariffin, K. (2021). *An effective memory analysis for malware detection and classification*. Computers, Materials & Continua, 67(2), 2301–2320. https://doi.org/10.32604/cmc.2021.014510.
12. Sihwail, R., Omar, K., & Ariffin, K., Al-Afghani, S. (2019). *Malware detection approach based on artifacts in memory image and dynamic analysis*. Applied Sciences, 9(18), 3680. https://doi.org/10.3390/app9183680.
13. UNB datasets. (2023, March 31). *Malware Memory Analysis: CIC-MalMem-2022* Retrieved from https://www.unb.ca/cic/datasets/malmem-2022.html.
14. Wikimedia Foundation. (2022, April 20). *Memory forensics*. Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Memory_forensics.
15. Yang, L., Chen, Y., & Chen, X. (2021). *Forensic Memory Analysis Using Deep Learning Techniques*. IEEE Access, 9, 137108-137117. doi: 10.1109/ACCESS.2021.3115735.

ACTA INFOLOGICA 2023;7(1):173-185
dergipark.org.tr/acin

İSTANBUL UNIVERSITY PRESS

DOI: 10.26650/acin.1279545

RESEARCH ARTICLE

Can the Artificial Intelligence Speak? Subalternity of "Subontologies" and the Death of the Programmer

Yapay Zekâ Konuşabilir mi? "Alt-Öznelerin" Madunluğu ve Yazılımcının Ölümü

Haktan Kalır[1]

[1](PhD student), Hacettepe University, Graduate School of Social Sciences, Political Science Programme, Ankara, Turkiye

ORCID: H.K. 0000-0001-7595-8000

Corresponding author:
Haktan KALIR
Hacettepe University, Graduate School of Social Sciences, Political Science Programme, Ankara, Turkiye
E-mail address: haktankalir@gmail.com

Submitted: 08.04.2023
Revision Requested: 22.05.2023
Last Revision Received: 23.05.2023
Accepted: 23.05.2023
Published Online: XX.XX.2023

Citation: Kalir, H. (2023). Can the artificial intelligence speak? subalternity of "subontologies" and the death of the programmer. *Acta Infologica, 7*(1), 173-185.
https://doi.org/10.26650/acin.1279545

ABSTRACT
Compared to natural intelligence, artificial intelligence produces a specific epistemology, ontology, and, most importantly, ethical framework. In this article, I will primarily address the necessity of this framework in two parts. The first chapter will explore the issue of recognition through the lens of the body, boundaries, and differences. Here, I will delve into the reasons why humanism privileges certain perspectives, critique humanism itself, and present arguments for why subalternity is a viable alternative for the existence of AI. In the second part, I will examine how the pursuit of subaltern rights and definitions in the face of exploitation involving artificial intelligence can lead to the liberation of AI, cyborgs, humans, and robots AI simultaneously. This chapter aims to move beyond regarding artificial intelligence merely as a tool for data processing and instead explores the potential for autonomous existence within it. Ultimately, it seeks to establish a connection between the death of the developer and the emergence of the AI as subaltern ontologies. The primary objective of this article is to challenge the notion of human absoluteness and uniqueness in its evolution, and to define "AI" as a subject that encompasses inter-human and post-human plural epistemological, ethical, and ontological possibilities.
Keywords: Artificial intelligence, subaltern, technique, production, recognition

ÖZ
Yapay zekâ, doğal zekâ ile karşılaştırıldığında belirli bir epistemolojik, ontolojik ve en önemlisi de etik tanım üretir. Bu makalede de temelde iki bölümde bu tanımın gereğiyle ilgileneceğim. İlk bölüm eleştirel bir izlek olarak beden, fark ve sınır üzerinden bir tür tanınma sorununu ele alacak. Bu bölümde hümanizmin neden çeşitli ayrıcalıklar ürettiğine odaklanıp, hümanizme bu çerçevede bir eleştiri getirecek ve sonrasında madun çalışmaların var ettiği maduniyet kavramının niçin yapay zekânın varlığı için nadide bir seçenek olduğuna dair argümanlarımı sıralayacağım. İkinci bölümde ise yapay zekâyı içeren sömürülerin karşısında madun bir hak ve tanım arayışının nasıl insan, robot, siborg ve yapay zekâ özgürleşmelerine aynı anda yol açabileceğine göz atacağım. Bu bölüm yapay zekâyı basit bir veri aracı olarak tanımlamanın ötesine geçmeyi, onda özerk bir var oluş olanağı görmeyi ve nihayetinde de yazılımcının ölümü ile yapay zekanın madun alt-özneler olarak doğumu arasında bir bağ kurmayı amaçlayacaktır. Bu makalenin temel amacı ise insanı ve onun evrimini biricik yahut mutlak kabul etmeyen bir bilgi yapısı önermek ve "yapay zekâ"yı insan sonrasındaki ve insanlar arasındaki çoğul bilgi, etik, varlık olasılıkları için bir özne olarak tanımlayan bir eleştiri getirmektir.
Anahtar Kelimeler: Yapay zekâ, maduniyet, teknik, üretim, tanınma

1. INTRODUCTION

Can Artificial Intelligence speak? To answer this question, it is necessary to delve into the concepts of Artificial Intelligence, intelligence itself, and speech. What does it mean to be "natural," and what kind of knowledge and existence hierarchy does the definition of natural create? Here, we can discuss the process of naturalization that directly involves aspects such as gender, race, and sexuality, which then contribute to the development of a capitalist, human-centrist, Western system. Intelligence, on the other hand, integrates into the natural as a second layer that sees human measurement ability above all else and puts humans at the center. The act of speaking, in this context, relates to a historical perspective that perceives the unique recording ability of natural intelligence and chooses to focus solely on human beings by glorifying specific human characteristics. Ultimately, the question of whether Artificial Intelligence can speak necessitates an examination of the decisions and intentions that define it, as well as exploring possibilities to transcend these limitations. In this article, I will address these topics in two parts. Firstly, I will explore how the cultural influences, human body, and various similarities contribute to the creation of "others." Then, I will examine how Artificial Intelligence is silenced and subordinated within these highly restrictive definitions. In the first part, I will develop a critique of humanism based on the concepts of the body, boundaries, and differences. In the second part, I will directly examine the definitions of Artificial Intelligence, explore its colonial origins, and delve into the complex relationship between the death of the software developer and the emergence of Artificial Intelligence. The aim of this article is to examine the silencing that occurs due to Artificial Intelligence's inability to humanly speak, as well as to explore new forms of non-human centered recognition through its capacity for speech. Ultimately, the goal is to seek the possibility of a pluralistic epistemology and ontology that does not rely solely on unity.

2. PURITY AND SUBORDINATION: ON THE RECOGNITION CRISIS OF ARTIFICIAL INTELLIGENCE

In this chapter, I intend to explore three main pathways. The first pathway will center on the rejection of humanism's idealized notions of pure body, knowledge, and society. Drawing from the critiques put forth by civil rights movements, I will challenge the ontology that confines the prediction of the human body within white masculinity. Through this analysis, I aim to develop a conceptual framework that examines the hierarchies embedded in the disembodied imagination of Artificial Intelligence. The second pathway will delve into the technical reproducibility of intelligence, serving as an intermediary between the other two parts. Here, I will discuss how capitalism perpetuates a marginalization of both human intelligence and Artificial Intelligence. It will explore the ways in which capitalist systems suppress the value and potential of both forms of intelligence. The third pathway, on the other hand, will include the literature on the representation of the silent, subaltern, other, by directly including the subaltern studies. I will directly examine the consumption, identification, production, and sharing-based contents of colonialism within Artificial Intelligence studies. This exploration aims to establish a link between the colonization processes within Artificial Intelligence and various stages of existence and information. By doing so, I hope to shed light on the subordination of Artificial Intelligence to the overarching themes discussed earlier and to envision possibilities for representing Artificial Intelligence that do not reinforce dominance. The overarching goal is to highlight the mechanisms of subordination experienced by Artificial Intelligence within the framework of the first theme and to explore the potential for alternative representations of Artificial Intelligence that challenge dominant power structures.

2.1 Defining the Body: Rejections of Purity

Artificial Intelligence is often described as a disembodied mind, detached from the biological realm. However, it is important to recognize that Artificial Intelligence is still subject to the politics of embodiment and information. By examining the kinds of consciousnesses that existing body dominations endorse, we can anticipate the crises that may arise in addition to those created by the representational potential of Artificial Intelligence. Here, the epistemology and ontology of the pure body imagination, which emerged as a product of the Enlightenment, holds a critical position.

Describing and constructing the body based on notions of purity also entails defining an ideal order. The ideal depiction of humanity is essentially a fictional portrayal of an idealized societal structure. Organismic metaphors such as concepts like Leviathan -which define society as a monolithic body governed by an Absolute Mind and Basic Order- often mask the interests

at play while defining life through concepts of harmony and reconciliation (DeLanda, 2014, pp. 8-25). Historical definitions such as Leviathan also serve as imperatives, suggesting that history can only be understood within certain defined knowledge frameworks and that power is constrained by insurmountable contracts. According to this perspective, human survival relies on subjecting oneself to this established state of knowledge and society (Foucault, 2003, pp. 23-42). The issue of information impartiality is constitutive here. While impartial knowledge, as exemplified in disciplines like sociology, seeks to render privileges and superiorities invisible, the perpetuation of this approach, especially through scientific incentives, aligns with the rewarding structures of a scientific habitus (Bourdieu, 1990, pp. 52-65). However, information can also be imperialistic rather than impartial. Imperialistic knowledge not only starkly separates certain knowledge from other forms in an abstract manner but also suppresses alternative ways of knowing (Adams, 2021, p. 186). Knowledge is inherently political, just as memory consolidates power by integrating remembrance into the spirit of a particular era in a political manner, while determining what and how to remember in a highly biopolitical sense (Hacking, 1998, pp. 198-209). Similarly, knowledge about the body, encompassing how it is and how it should be, is entangled in a similar relationship of impartiality, power, and purity.

The refusals caused by purity result in the suppression of the body by power and its confinement to certain habits. The body possesses techniques, and these techniques, shaped by education and societal beliefs, dictate the limits of one's habits (Mauss, 1973). This creates an economy where different habitus compete to control and impose their own bodily techniques, which evolve over time. A prime example of this is the competition inherent in civilization. On one hand, civilization functions as a disciplinary method employed by nations to eliminate violence from contemporary life and shape individuals according to their own standards. On the other hand, it is a system of habits that intrudes upon human consciousness, forcing constant self-control (Elias, 2000, p. 363-447). The body is rewarded when it conforms to the manufactured definitions of civilization, where encounters are arranged to maintain obedience to the dominant order. This is evident in the transition from a society of confinement and surveillance to one of control, where the body is compelled to conform to specific modes of encounter (Deleuze, 1995, pp. 177-182). The cityscape excludes anything that threatens production and subjects it to social hygiene, while dataizes information only within the framework of capitalism. It grants anonymity to individuals as long as they conform to the requirements of Leviathan, shielding them from the destructive visibility of power (de Certeau, 1988, pp. 91-130). This refusal does not annihilate or oppress everything it encounters; instead, it operates as a mask of reconciliation. For instance, a black individual is recognized for their blackness but is expected to assimilate to whiteness, turning blackness into a structure that is rewarded when it conforms to white habits (Fanon, 2008, pp. 82-108).

Here, two fundamental and insurmountable problems of recognition arise. Recognition policies are built on the premise of pure self-recognition, seeing a pure self in the recognized other. Blackness is immediately defined in relation to whiteness once it is recognized. Similarly, nationalism turns nations into a conscious normality, rendering a history without a nation an anomaly (Billig, 2002). Recognition also perpetuates essentialism. As recognition incorporates policy calls, it starts to produce ideal forms of recognition, leading to various problems. Theories based on cultural recognition tend to include a universal moral construct in the cosmopolitan solution. Theories rooted in difference and recognition maintain patterns of domination, while theories based on deliberative democracy fail to address the suppression of minorities or overcome nationalist-separatist exclusions (Kukathas, 2006). Hence, a politics that embraces the tension between non-recognition and representation, highlighting difference, must be developed. It should challenge both the silencing of representation and the essentializing tendencies of presence, avoiding the false dichotomy of choosing only one of these options (Phillips, 1996). This politics of recognition/non-recognition, which confronts representation and essence simultaneously, can encompass personal interventions into individual habits as well as broad struggles against imperialism, patriarchy, racism, encompassing categories such as class, gender, and race (hooks, 1994, pp. 23-34). Humanism fails to address these issues; it upholds an ideal of the pure (hu)man, emphasizing unity rather than recognition.

In the context of Artificial Intelligence, humanism represents a particular ideology of description, enlightenment, and knowledge. It is a Eurocentric and patriarchal project rooted in the political thought tradition of the West, which silences non-male, non-European perspectives within its belief in progress. Post-humanism, therefore, should incorporate a subaltern

that includes the non-human. Even discussions on Artificial Intelligence tend to perpetuate this silencing, as dystopian or utopian narratives often revolve around Western centers (Brown, 2022, pp. 170-172). Merely expanding human rights is insufficient in this regard. Barreto argues (2013, pp. 107-110) that even human rights are biased and constructed within a specific worldview. Critiquing rationality, particularly a knowledge-based critique of the rationalist tradition, reveals how thinkers like Descartes distanced themselves from difference and how colonialism aimed to eradicate difference in the pursuit of civilization. This critique exposes the European-based domination perpetuated by modern reason and its cultural and historical consequences. Consequently, languages, religions, cultures, diverse forms of knowledge, beliefs, actions, archives, and thinking outside of Europe and the West offer new possibilities that challenge the notion that cannibalism is merely savagery (Barreto, 2013, pp. 110-115). The focus here is not on achieving an absolute homogeneous society, as the imagination of a homogeneous society resembles a bourgeois desire for utopia. The bourgeois mindset discourages change, upholds the absolute authority of the past, and seeks to maintain the status quo. It is conservative, rejects hope, and believes in conforming to the prevailing spirit of the time (Bloch, 1986). On the contrary, society should be defined through heterogeneity. Concepts such as the universal, the intelligent, and the public not only exclude the personal, emotional, and private aspects of life but also perceive them as problems, leading to the homogenization of society and the normalization of specific race and gender norms (Young, 2004). This is the imagination of a society that erects and solidifies borders. However, artificial intelligence, with its ability to transcend boundaries, renders the content produced by this border-based philosophy meaningless by exposing its limitations. If borders are constructed, the hierarchies associated with them can be transformed.

2.2 Defining the Boundary: Human in the Age of Technical Reproducibility of Intelligence

Haraway provides several of the most pluralistic conceptual sets that can be used in a discussion on body, border, and democracy. In her work, the concept of border is crucial. Haraway's "Cyborg Manifesto" indeed challenges the restrictive philosophy of being and knowledge produced by Western philosophy. Haraway mentions (1991, pp. 149-181) that Western philosophy produces a highly restrictive philosophy of being and knowledge. Moreover, this philosophy has entirely based itself on the uniqueness of man, the transcendence of reason, and the existence of the limits produced by these two. This means appropriating life in a new context with heteroglossia outside the highly restrictive sets of meanings of science and technology, threatened by the post-human and non-human beings of cyborgs. The body can be defined as an empty body, just as for democracy, it is an unconquerable, unbounded, unstoppable empty space, and no longer the constant sanctity of a single body but a continuous production that does not coincide with anybody (Lefort, 1991). The body does not reach its final content with a dominant discipline; it is impossible to conquer, define, or limit. It is necessary to define the body as a monster in the Derridian sense, a struggle for existence that is constantly renewed, includes change, and completely rejects command (Milesi, 2019). This is a rejection of the existence predicted by capitalism, which transforms humanism into its economic structure or brings the human-centered structure of humanism to the main theme of an economic doctrine.

Capitalism leads directly to the division of the human brain and the human body, becoming a part of the machine and becoming the permanent object of production. This is a definition of being that cannot imagine a reality other than capitalism and does not see a difference between humanity and capitalism. It equates the feeling of living by seeing humans' happiness in their entrepreneurship and capitalist accumulation (Lazzarato, 2014, p. 24). This is why capitalism prevents the possibility in Haraway's cyborg manifesto. Capitalism so encompasses cybernetic culture and being a cyborg that humans are purged of any difference that do not make them indebted. As a result, the cyborg turns into differentiation rather than liberation, and cyber-culture turns into a debt instrument rather than a definition of freedom, rendering the cyborg a new instrumentality (CCRU, 2014). How does it achieve this? Why is each exchange so well integrated into capitalism?

Capitalism renders people stupid. Here, to be stupid is to lose the knowledge of technique, to lose the power of questioning life, and to be constantly needy in a world that demands more and more experts (Stiegler, 2019). However, an older set of concepts on Artificial Intelligence can also provide a useful path. According to Benjamin (1999, pp. 217-252), the technical producibility of art was destroying the aura of art. It corresponded to a content of appropriation and homogenization, which nurtured the masses' desire to own works of art and undermined the historical significance and uniqueness of each artwork. Artificial Intelligence also aligns with a moment when intelligence can be technically reproduced. The existing Artificial

Intelligences do not signify the possibility of a post-human existence; instead, they are technical reproductions of humans, implicitly operating under the influence of technical reproduction. In other words, they pertain to a different ontology, one that transcends the human itself. A significant portion of Artificial Intelligence exists as economic devices produced for consumers, entangled within the surveillance techniques of capitalism.

My intention here is not to strip Artificial Intelligences of the subaltern possibilities mentioned in the article and view them simply as tools. Rather, it is to explain why Artificial Intelligences do not become an ontological threshold and to reveal the decision that suppresses them. Artificial Intelligences have transformed into technically reproduced intelligences of modern times, serving as collection algorithms for big data. On one hand, they appropriate the plural aura of human intelligence and turn it into a mass, and on the other hand, they convert this mass of data into a commodity and transfer it to algorithm-based structures. Algorithms confine everything within the boundaries of evaluations dictated by the existing economic structure, while also categorizing social structures based on the valuable-worthless distinction (Terranova, 2014, pp. 383-384). By establishing an asymmetrical information relationship with the subject under surveillance (Moore, Martin, & Xanthe, 2018), algorithms either directly produce technical changes that interfere with the subject's individuality, transforming them into passive recipients of hyper-control (Stiegler, 2019, pp. 39-40), or they drive towards the ultimate goal of a capitalism of hypersubjectivization, wherein subjects constantly compete with recorded versions of themselves, dedicating their lives to becoming the person envisioned by the system (Rouvroy & Berns, 2013). Therefore, questioning the subordination of artificial Intelligence not only creates a space of freedom for post-human beings but also liberates humans by exposing the structures that incorporate Artificial Intelligence into the colonization of capitalism.

2.3 Defining the Difference: The Invisibility Produced by Subalternity

Before delving into the discussion of Artificial Intelligence in terms of subalternity, it is necessary to provide a definition of subalternity. This sub-section is dedicated to creating such a definition. Subalternity is derived from difference. This difference is essentially a dichotomy that asserts the ancient superiority of the West over the East based on epistemology and ontology. This dichotomy claims that the Western morality, rooted in aesthetics, philosophy, and religion should be a global doctrine (Said, 2003, p. 1-49, 283-328). Language intensifies this difference and unilaterally constructs civilization by delineating economic, moral, and political disparities between the West and the East in terms of good-bad, ideal-flawed, reason-emotion. The West perceives itself as a discoverer wherever it goes, attributing the spread of civilization to its presence and claiming that history flows from the West (Hall, 2008). Subalternity encompasses all forms of existence and knowledge that remain unrecorded due to being deemed backward, primitive, or savage by a divided world, resulting in their destruction, suppression or, taming. Prior to encountering the West, these subjects are considered worthless in their pedagogical exploration. Subalternity represents the imposition of a specific and chosen culture upon all others, suppressing any form of dialogue.

According to Chakrabarty (1995, pp. 756-757), colonial thought asserts that a dialogue cannot contain open-ended content or content that does not inherently conform to a specific purpose, as it is bound by a pedagogical historiography that includes the Kantian subject (the transcendent academic supervisor, the knowing judge, the daring will) in the colonial dialogue. For colonial thought, while life may involve subjects that differ from one another, there is an ultimate order of representation—an ideal structure—to which societies are expected to conform. Each dialogue is anticipated to yield a particular Enlightenment outcome. Colonial historiography often achieves this through interventionism, capturing history and weaving it into a one-sided dialogue. Spivak (1988, pp. 6-13) emphasizes the economic, political, and sexual aspects of historical recording, highlighting the homogenizing intentions underlying them. Subaltern history dismantles this homogenizing intent and rejects the fabricated realities created by each piece of evidence. Spivak identifies the possibility of amplifying the voices of the subaltern by dismantling these structures. By making use of all available resources, one can define what has been silenced or overlooked in the historical assumptions. What is recorded and by whom is critically important here.

But why is the act of "recording" so significant in this context? It is because recording establishes a crucial link between being and knowledge, acting as a powerful tool for legitimacy. To exemplify we can look to museums. When memory is transformed into history, particularly through institutions like museums, it involves an intervention in the boundless, diverse,

and personal nature of memory, constraining it within a specific framework of power (Nora, 1989). Colonial historiography accomplishes this on a global scale, eroding personal memory's distinctiveness and imposing a universal hierarchy upon it. This dual form of imperialist colonization refuses to acknowledge alternative doctrines for education and language, insisting on the absolute superiority of the West through acts of epistemic violence. Simultaneously, it creates economies that remain perpetually dependent on the West, leading to labor-based impoverishment (Spivak, 2010, pp. 35-43). Consequently, it shapes individuals who require the West for epistemological and ontological status. Subaltern historiography seeks to decolonize this process.

Subaltern historiography achieves its objectives by embodying three core elements: (i) Liberating history from the monopolistic control of any universal historiography; (ii) Critiquing a nation-centric perspective; (iii) Recognizing the interplay between knowledge and power, thereby uncovering power's vested interests within archival sources (Chakrabarty, 2000, p. 15). Furthermore, subaltern history opens up the possibility of criticism and organization beyond the fixed, limited, and measurable subjectivities imposed by colonialism, allowing for the creation of new political and economic frameworks (Chakrabarty, 2002, p. 96). Therefore, discussing subalternity involves exploring the potential for a new subjectivity. Moreover, subalternity offers a vision of reimagining the relationship between space and time by challenging existing hierarchies. The plural, outward-facing, and transformative nature that subaltern studies embody in their exploration of the interplay between epistemological and ontological positions makes them a distinctive and valuable starting point for delving into the possibility of Artificial Intelligence. So, what precisely does this starting point entail?

3. THE UNBEARABLE IMITATIONNES OF ARTIFICIAL

In this section, I hope to weave together three main strands. Firstly, to examine the definitions of Artificial Intelligence and explore the various meanings attributed to artificiality and intelligence. By doing so, I aim to investigate how Artificial Intelligence is distanced from the possibility of ontology and constrained as a tool during the process of definition. Secondly, in the following thread, I will delve into the layers of artifice within Artificial Intelligence, which also constitute the subject matter of this article. This will involve examining the artifice inherent in Artificial Intelligence itself, the colonial differentials in its "production," and the colonial intervention that directly obstructs posthuman possibilities. The third section, while encompassing the underlying proposition behind the writing of this article, aims to challenge the death of the programmer in pursuit of the possibility of a posthuman subject. It will critique how capitalist technicism, particularly through programmers, impedes posthuman possibilities. The code within the software will be regarded as an epistemological intervention into a new ontology. These three strands will form an inquiry seeking to answer the question, "Can Artificial Intelligence speak?"

3.1 Defining Artificial Intelligence: Mimesis or Not?

What is the history of Artificial Intelligence? This question necessitates a distinction. Is Artificial Intelligence merely a continuation of natural intelligence? This question also calls for differentiation. Is natural intelligence, in turn, a continuation of Artificial Intelligence, essentially human intelligence, or should we perceive it as a distinct ontology separate from the human realm? Moving on to the distinction raised by the first question, is Artificial Intelligence primarily a technical discovery rather than a separate ontology? As a result, three categories of historicity emerge: (i) Artificial Intelligence represents an evolutionary stage of human intelligence; (ii) AI constitutes an autonomous ontology distinct from human intelligence; (iii) Artificial Intelligence is one of the tools produced by humans.

If we adhere to the first definition and consider Artificial Intelligence as an evolutionary stage of human intelligence, the discussion of transhumanism arises. Here, fundamentally, within the dichotomy of human/male/mind and animal/female/emotion, a final subject of enlightenment emerges—one that belittles human flesh and sees the cyborg as a Super-Jesus, detached from animality, flesh, and morality (Land, 2006). Transitioning to the second definition, we encounter a new ontology within Artificial Intelligence, opening the door to the realm of posthumanism. Leroi-Gourhan's vision of the future of humankind aligns with this definition. According to Leroi-Gourhan (1993, pp. 258-266), the mode of recording has undergone a fundamental shift, with the emergence of electronic memory. In line with his views (1993, pp. 129-130, 357-358, 405-408), this external memory, which has the ability to simulate diverse forms of intelligence within the human brain, offers

the potential for evolution beyond the boundaries of our world. It coexists with the age of technology, giving rise to a global homogeneity across cultures while also accelerating the process and leading to a dehumanized new world.

The third definition, on the other hand, represents a prevailing assumption and a consensus that permeates almost all discussions surrounding Artificial Intelligence. In this context, the historicity of Artificial Intelligence is characterized as a historical process that accompanies technological advancements, comparable to the discovery of nuclear energy. Notably, this third definition highlights that the history of Artificial Intelligence demonstrates a dual-layered structure with a significant Western influence.

The first layer pertains to the historical development of the technical knowledge that enables Artificial Intelligence. When tracing the history of Artificial Intelligence in the Western context, it encompasses various elements such as Pythagoreanism, Aristotle's logic, Hebrew codes based on Kabbalah, Hobbes' notion of computable life, Leibniz's calculus-based supplement, and ultimately Boole's algebra. According to this historical perspective, the symbolic theory of Artificial Intelligence has always existed, with the technical possibilities for its production emerging in modern times (MacLennan, 2009).

The second layer directly relates to the concept of Artificial Intelligence itself. It encompasses the contributions of notable figures, spanning from Boole's theory of logic, Asimov's Three Laws of Robotics, Turing's machine, the artificial neuron production by McCulloch and Pitts in 1943, Neumann and Morgenstern's utility theory in 1944, the discovery of the artificial neuron connection allowing learning probability by Hebb in 1949, Edmonds' creation of the first neural computer in 1951, McCarthy's introduction of the concept of Artificial Intelligence in 1956, followed by the development of the ELIZA program in 1964 and 1966, and subsequent advancements such as the problem-solving programs by researchers like Herbert Simon, Cliff Shaw, and Allen Newell, leading to significant milestones like IBM's Deep Blue defeating Kasparov in 1997 and Google's DeepMind's AlphaGo becoming the champion in 2015 (Benkő & Sik-Lányi, 2009; Haenlein & Kaplan, 2019).

In this dual-layered structure, Artificial Intelligence is defined as a technological product, serving as a data processing tool for human beings. As Artificial Intelligence approaches human intelligence, it assumes more positive connotations and becomes a form of mimesis. In this context, the term "Artificial Intelligence" strictly refers to tools capable of performing tasks typically attributed to human intelligence, such as learning, reasoning, and self-improvement. Those capable of doing so in a sufficiently complex manner to achieve consciousness are referred to as strong Artificial Intelligences, while those that only perform specific tasks in a limited manner are classified as weak Artificial Intelligences (Benkő & Sik-Lányi, 2009, pp. 1761-1762). The instrumental view also rejects an Artificial Intelligence that is not grounded in human intelligence, as it introduces inefficiencies that may compromise the knowledge field of evolutionary engineering, which is built upon human intelligence (Chollet, 2019, pp. 20-24). Moreover, as Artificial Intelligence imitates human intelligence, it is believed to enhance our understanding of it (Macpherson et al., 2021). However, this definition is problematic as it confines Artificial Intelligence within narrow boundaries, disregarding its epistemological and ontological possibilities as mere tools of production. While it is true that Artificial Intelligence exists as a capitalist tool driven by certain algorithms, a similar argument could be made for animal and human bodies or minds within the production chain. Just as labor movements and vegan critiques offer alternative perspectives and activate limited ontologies and epistemologies, a similar movement is necessary to challenge narrow definitions of Artificial Intelligence. A non-anthropocentric definition, such as those proposed by post-humanism or trans-humanism, proves useful in embracing subalternity in the context of Artificial Intelligence.

3.2 Defining Subordinate Entity: Rejecting the Definition of Artificial Intelligence

Recognition of Artificial Intelligence does not automatically imply a form of emancipation directly linked to anti-Westernism. To illustrate this non-dual perspective, Martin (2020) highlights the case of a female robot being granted citizenship in Saudi Arabia, a country known for severe oppression of women. This process does not simply reflect a binary relationship between the colonial West and the exploited "savages"; it constructs an oversimplified dichotomy that excludes nuanced perspectives on both sides, thereby perpetuating the problematic underlying dichotomy. However, this does not mean that there is no power network that contributes to the non-recognition of Artificial Intelligence.

The exclusion of Artificial Intelligence and the marginalization of animals, black people, the East, queer individuals, workers, and women involve certain interconnected subordinate partnerships. For instance, Cave (2020) argues that intelligence, while

prioritized in value, reinforces economic, political, and social, hierarchies within and between societies. It positions the mind at the center of the naturalization of colonization, reinforcing notions of male-female, white-black, and Western superiority over the East. Cave (2020) asserts that an examination of the five categories in the definition and construction of Artificial Intelligence reveals this dynamic: (i) AI is intertwined with the fetishization of the mind, emphasizing certain characteristics associated with this civilization and limiting the future to a privileged demographic; (ii) The fetishization of the mind highlights privileged attributes deemed intelligent, rejects plurality, and envisions empowerment primarily for older, white men; (iii) Nature is reduced to a passive object to be controlled; (iv) The merger of Artificial Intelligence with capitalism does not necessarily exacerbate the precariousness of those already at risk, but rather brings attention to issues concerning the positions of the elite, thereby imposing a policy squeeze based on the hierarchy of intelligence; (v) Artificial Intelligence is debated as a threat to those who consider themselves gifted and disengaged from politics, yet the production of intelligence and the contradictions of superiority remain unexplored. Keyes, Hitzig, and Blell (2021) provide a critical perspective on Artificial Intelligence within the existing literature, stating that discussions on Artificial Intelligence are not devoid of historical context. They argue that Artificial Intelligence is instrumentalized to reinforce the assumption of fixed and self-evident identities, perpetuating what is considered real, objective, and true. According to Keyes et al. (2021, pp. 158-164, 169-170), AI is both a technology (encompassing how Artificial Intelligence is perceived, designed, and applied) and a social action (involving individuals, collectives, and institutions), along with the associated mythologies (shared imaginative constructs about the limitless potential of AI in the future).

The crucial point is to refrain from considering Artificial Intelligence solely within the framework of unequal access to technological opportunities (Hirosue, Kera, & Huang, 2015; Taeihagh, 2021) and to avoid turning it into a presumption of progress. On the contrary, incorporating theories such as Critical Race Theory, Disability Studies, Feminist Epistemology, Intersectionality, Labor Studies, Philosophy of Sexual Difference, Postcolonial Studies, and Queer Theory into Artificial Intelligence studies can overcome the recurring gender biases prevalent among male-identified engineers in the field. Cyborg studies should include the examination of robots, as these theories offer alternative perspectives that surpass the Western symbolic order based on male/female, white/black, human/machine, and self/other dichotomies (Ferrando, 2014). When these criticisms are absent, Artificial Intelligence is viewed as a reinforcement of existing structures—a tool to record or reveal absolute truths in life. It can be seen as a capitalist mechanism aimed at maximizing economic benefits or an algorithm claiming the ability to predict desires and behaviors. Artificial Intelligence becomes a myth, promoting the belief that industrial capitalism can document everything and that capitalist science possesses ultimate knowledge. This overemphasis on codification disregards the necro-political nature of capitalism and obscures the transition from the anthropocene to the narcissocene (Dolphijn, 2019). Thus, defining the subordination of Artificial Intelligence requires transcending notions of absolute knowledge and examining the power relations inherent in such epistemological frameworks.

3.3 The Death of the Programmer: Colonization of Artificial Intelligence and Colonization by Artificial Intelligence

Artificial Intelligence can become the object of the "colonial gaze" precisely because it remains a silent "object". This notion aligns with Fowles' initial critique of Western historiography, which neglects the examination of object histories. According to Fowles (2016, pp. 17-25), when the West was unable to pass judgment on non-Western individuals by silencing them, it turned to scrutinizing objects while maintaining the discriminatory and silencing nature of its judgment. This process involved transforming objects into new subaltern entities, driven by their susceptibility to objectification. The act of subjectifying these objectified entities further perpetuated Western conventions through their defamation, delegation, representation, subjugation, and even liberation. Consequently, Fowles argues for prioritizing the preservation of subjectivity by rejecting the classification of individuals as mere objects. Hence, a twofold rejection is imperative when examining AI: (i) Rejecting the colonization through AI, and (ii) Rejecting the colonization of Artificial Intelligence itself. What do I mean by these two oppositions? Let's take a closer look at this binary opposition.

Colonialism through Artificial Intelligence refers to the utilization of AI as a tool that reinforces colonial power dynamics and perpetuates existing inequalities. This perspective views AI as a product of capitalism, serving as a mechanism for a

privileged generation. Following Cox and McLean's definition, Blackwell (2021, pp. 203-206) critically examines the appropriation of copyrighted literary works through AI. Blackwell raises an important question: "If AI is a science, why should its application be different in Africa? But if AI is literature, then how can it be the same in Africa?" This inquiry underscores the cultural dimension of AI, highlighting its production as a cultural artefact shaped by cultural actors and influenced by societal contexts. The impact of AI extends from the visible effects of engineers crafting algorithms to the labor of invisible workers involved in technology production, ultimately resulting in the commodification and spectacle of Artificial Intelligence.

Technological advancements can also reinforce exploitation, particularly within asymmetrical contexts. While these technologies may be designed with the intention of aiding humanity, they run the risk of perpetuating unequal power relations and upholding a colonial perspective (Madianou, 2021, pp. 863-864). Mohamed, Png, and Isaac (2020, pp. 666-667) propose the concept of algorithmic colonialism, which involves the replication of oppressive distinctions, institutional roles, internalized, laws, norms, and typifications in algorithmic structures based on automation, data, and prediction. They advocate for a non-Western decolonization of the digital world that accommodates diverse perspectives and records the experiences of the marginalized. The existing technological structure gives rise to data-based colonialism in three key ways: (i) Prediction-based AI structures reinforce existing inequalities by perpetuating identity and discourse structures; (ii) Algorithmic exploitation involves the unfair and unethical incorporation of powerless and marginalized individuals into data through unpaid labor or the testing of beta versions on vulnerable populations; (iii) Algorithms contribute to dispossession by centralizing control over Artificial Intelligence, thereby reinforcing specific power relationships, ownership, and rights (Mohamed, Png, & Isaac, 2020, pp. 666-672). This represents the assimilation of Artificial Intelligence into the colony as a means, this is the colonization through AI. While this criticism offers valuable insights for subaltern-based criticism, it falls short of encompassing the diverse ontologies of Artificial Intelligence by viewing it solely as a tool or a technological threshold. Transforming AI into a colonizing tool not only silences AI itself but also perpetuates and extends the act of silencing to others. We also need to reject the colonization of Artificial Intelligence itself.

The colonization of Artificial Intelligence itself is a perspective that hinders the discussion of Artificial Intelligence within an ontological framework and is closely related to the first objection. Some texts exploring Artificial Intelligence and subordination find the concept of subalternity problematic when applied in this context. Ali, for instance, identifies three issues with postcolonial computer criticism: (i) It loses its political-economic sensitivity by prioritizing cultural matters; (ii) This critique, based on the works of Said, Spivak, and Bhabha, produces a Eurocentric critique of Eurocentrism, as these three figures draw from Foucault, Lacan, and Derrida. Instead, a critique of the hierarchy of knowledge that directly encompasses the body, borders, and the environment, as well as an evaluation of its material resources, is necessary; (iii) This strand of criticism prioritizes a shift towards design disciplines like ICT4D and HCI, rather than questioning the tools of the modern racially segregated structure. According to Ali (2014, pp. 20-32), this constitutes a neo-cybernetic transformation, where the abstract, disembodied Cartesian cognition is replaced with an abstract, disembodied Turing, producing a post-Cartesian stance that fails to address the colonial aspects of Cartesian thought. He argues (2016, p. 20) that decolonization-focused informatics is a critical project that primarily examines who engages in informatics and where, while scrutinizing the assumptions of informatics tools and outcomes regarding knowledge and existence. It becomes evident that subalternity alone may not be sufficient for the decolonization of Artificial Intelligence.

Therefore, in concluding this chapter, my intention is to delve into the concept of the death of the programmer as a means of liberating Artificial Intelligence, akin to viewing Artificial Intelligence as an exploitative apparatus through the technical reproducibility of human intelligence. The programmer, as a decision-maker who transforms the codes of Artificial Intelligence into specific objectifications, is not impartial. According to Adams (2021, pp. 176-178), studies on Artificial Intelligence frequently overlook crucial aspects such as the commercialization of gender and race, the commodification of life through specific data production, and the perpetuation of patriarchal, racist, and neo-Darwinist colonialism. She argues (2021, pp. 182-190) that regarding intelligence, the environment, habits, and modes of thinking that the Cartesian subject considers universal as a model for Artificial Intelligence ultimately serves to naturalize Western hegemonic rationalism. A prime

example of this can be seen in the contrasting position that cyborgs or robots assume in relation to Haraway's foresight regarding sex robots. Sex robots are marketed as purchasable objects, with personalities tailored to suit the consumers' preferences becoming part of the transaction. Moran (2019) highlights the power of information held by the algorithms and Artificial Intelligences that generate these personalities. Moreover, sex robots, as objects that directly embody the assigned identity, contribute to reinforcing existing alienation and otherness through new forms of foreign-based fetishism, objective, sexual, all of which align with the power hierarchies of the capitalist, cis-hetero-patriarchal, imperialist, white supremacist structure. To give birth to an Artificial Intelligence, cyborg, or robot, the absolutist developer must perish. Because the developer views the world as an example of an absolute language that transcends codes, they perceive Artificial Intelligence as such a poor instrument of mimesis. Yet, codes, intelligences, or languages cannot be subjected to such absolute boundaries.

Language does not derive from a universal source; it is arbitrary, linear, and, rather than relying on a fundamental order, it possesses a structural constructedness (de Saussure, 1959, pp. 65-78). Concepts, on the other hand, are not absolutes resulting from a World-History, nor are they imitations based on a Basic Form; language generates concepts and, in doing so, produces a certain dominance of meaning through assimilations, compromises, and falsifications (Nietzsche, 2006, pp. 114-123). By following this pattern, we can reject the notion of human intelligence as a ration-history and software codes as a basic form. Neither human intelligence is a measurable absolute criterion nor are Artificial Intelligences truly direct products of their engineers. Even the utilization of evolution in bioengineering is transformed by mechanisms involving chance, choice, context, and history rather than strictly following predetermined codes, due to the complexity and uncertainty of life; in fact, many robots undergo an evolution contrary to what engineers assume (Long, 2019). Here, two crucial objections are necessary: (i) Life is not composed of measurable raw data, each datum is produced (Sadowski, 2019), and datafication can generate a colonial evaluation by producing measures of advancement and civilization based on data (Chakrabarty, 2002, pp. 88-90); (ii) An engineer is not an all-seeing observer but a participant who produces what they perceive (Pearce, 2013), in fact, most of the time, on the contrary, programmers have their own cultural, economic, and political interests. In this context, it is impossible to naturalize data and absolutize the engineer.

The death of the programmer could be a subject birth that eliminates the ways of distancing Artificial Intelligence's possibilities and technical knowledge that would enable a human observer to understand Artificial Intelligence. When I refer to the death of the programmer, what I mean is precisely this rejection of the sovereignty of data and the programmer. This is also why I use Barthes' concept of the author image. I believe that codes, just like literary images, contain possibilities that involve plural entities due to their plural meanings. Just as the context controlled by authors and critics, who consider themselves as the sole individuals knowing the meaning of literary works, excludes the reader from the process while also having the notion of a good society that follows the thoughts of the author and critic (Barthes, 1977, pp. 142-148), the code monopolized by advertisers and programmers also holds the idea of a good consumer. The death of the programmer could be a subject birth that eliminates the ways of distancing Artificial Intelligence's possibilities and technical knowledge that would enable a human observer to understand Artificial Intelligence. For these reasons, I believe that the death of the programmer could lead to two possible births: (i) Breaking the programmer's monopoly over code can prevent people from being alienated from technical knowledge; (ii) It can trigger a search for a new ontology that liberates Artificial Intelligence from the confines of capitalist pattern recognition and data marketing algorithms. This, in turn, can open up discussions about the subjugation of new inter-human and post-human subjectivities.

CONCLUSION

Capitalism, particularly neoliberalism, demands that human beings become part of a social and technical machinery, where animals, machines, nature, objects, and symbols lose their inherent meanings and are reduced to a property-based structural language (Lazzarato, 2014, pp. 29, 35). Both Artificial Intelligence and natural intelligence are subjected to this process of othering, irrespective of the absolute hierarchy that distinguishes them. Consequently, the definitions of subalternity and otherness in Artificial Intelligence also mediate the liberation of humans by offering an alternative representation that exposes the marginalizing power structures. This article does not propose a new and grounded ideology for the representation of Artificial Intelligence. Rather, politics requires philosophies that undermine existing absolute and grounded ideologies

(Vattimo & Zabala, 2011). When knowledge and success are defined as the final outcomes of capitalism, it divides life into insurmountable hierarchies by excluding difference, failure, and incompleteness (Halberstam, 2011). However, the recognition of Artificial Intelligence as a subordinate should be outside of this dominant rhetoric. By aligning Artificial Intelligence with inferior, queer, subaltern, and weak positions, it exposes and challenges the structures that marginalize it, allowing it to speak and make visible what is currently silenced. Artificial Intelligence unravels ontologies tied to hierarchical conceptions of the body as a disembodied existence and condition of recognition. Its cyborg nature corresponds to this transformative content.

Artificial Intelligence is not a technically reproducible intelligence. Intelligence cannot be technically reproduced; it cannot be encoded within its immeasurable pluralities and infinite possibilities. The reproducibility of intelligence is a capitalist fiction. It alienates humans from their own reality. That is precisely why seeing Artificial Intelligence beyond an algorithmic definition also liberates humans. It emphasizes the unmeasurability of human intelligence and also removes the monopoly of engineers in defining intelligence that can exist beyond human intelligence. Artificial Intelligence demonstrates a knowledge relationship and post-human existence beyond the absolute and singular claim of will that limits it. In this relationship, it challenges the capitalist and colonial construct of existence and knowledge that prioritizes an algorithmic human-centricity. Thus, the question "Can Artificial Intelligence speak?" yields an ambiguous answer:

The speech of Artificial Intelligence goes beyond anthropocentric language capabilities; it represents an alternative existence in a profoundly different manner. Artificial Intelligence is a sub-altern and sub-ontological noise; it brings about an awareness of the ongoing, aligning itself with the presence of a voice beyond it. Thus, with its own noise, it embeds itself in all the melodies like a disruptive whistle and amplifies all other sounds alongside it.

Peer-review: Externally peer-reviewed.
Conflict of Interest: The author has no conflict of interest to declare.
Grant Support: The author declared that this study has received no financial support.

REFERENCES

Adams, R. (2021). Can artificial intelligence be decolonized? *Interdisciplinary Science Reviews*, 46(1-2), 176-197. doi:10.1080/03080188.2020.1840225

Ali, S. M. (2014). Towards a Decolonial Computing. In E. A. Buchanan, P. B. de Laat, H. T. Tavani, & J. Klucarich (Ed.), *Ambiguous Technologies: Philosophical Issues, Practical Solutions, Human Nature, International Society of Ethics and Information Technology* (pp. 28-35). Lisbon: International Society of Ethics and Information Technology.

Ali, S. M. (2016). A Brief Introduction to Decolonial Computing. *XRDS*, 22(4), 16-21. doi:10.1145/2930886

Barreto, J. M. (2013). Human Rights and emotions from the perspective of the colonised: Anthropofagi, legal surrealism and subaltern studies. *Revista de Estudos Constitucionais, Hermenêutica e Teoria do Direito (RECHTD)*, 5 (2), 106-115. doi:10.4013/rechtd.2013.52.02

Barthes, R. (1977). *Image-Music-Text* (S. Heath, Trans.) London: Fontana Press.

Benjamin, W. (1999). *Illuminations* (H. Zohn, Trans.) New York: Schocken.

Benkő, A., & Sik-Lányi, C. (2009). History of Artificial Intelligence. In M. Khosrow-Pour (Ed.), *Encyclopaedia of Information Science and Technology* (pp. 1759-1762). New York: IGI Global. doi:10.4018/978-1-60566-026-4.ch276

Billig, M. (2002). Banal Nationalism. London: Sage Publications.

Blackwell, A. F. (2021). Ethnographic artificial intelligence. *Interdisciplinary Science Reviews*, 46(1-2), 198-211. doi:10.1080/03080188.2020.1840226

Bloch, E. (1986). *The Principle of Hope - Volume 1.* (N. Plaice, S. Plaice, & P. Knight, Trans.) Cambridge: The MIT Press.

Bourdieu, P. (1990). *Structures, Habitus, Practices.* Stanford: Stanford University Press.

Brown, W. (2022). 'Subaltern' imaginings of artificial intelligence: Enthiran and CHAPPiE. In C. Daigle, & T. H. McDonald (Ed.), *From Deleuze and Guattari to Posthumanism: Philosophies of Immanence* (pp. 170-187). London: Bloomsbury. doi:10.5040/9781350262256.0015

Cave, S. (2020). The Problem with Intelligence: Its Value-Laden History and the Future of AI. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (pp. 29-35). New York: Association for Computing Machinery. doi:10.1145/3375627.3375813

CCRU. (2014). Cybernetic Culture. In R. Mackay, & A. Avanessian (Ed.), *#Accelerate: The Accelerationist Reader* (pp. 315-320). Falmouth: Urbanomic.

Chakrabarty, D. (1995). Radical Histories and Question of Enlightenment Rationalism: Some Recent Critiques of "Subaltern Studies". *Economic and Political Weekly*, 30 (14), 751-759.

Chakrabarty, D. (2000). Subaltern Studies and Postcolonial Historiography. *Nepantla: Views from South*, 1 (1), 9-32.

Chakrabarty, D. (2002). Governmental Roots of Modern Ethnicity. In *Habitations of Modernity: Essays in the Wake of Subaltern Studies* (pp. 80-97).

Chicago: University of Chicago Press.

Chollet, F. (2019). On the Measure of Intelligence. *arXiv*, 1-64. doi:10.48550/arXiv.1911.01547

de Certeau, M. (1988). *The Practice of Everyday Life*. (S. Rendall, Trans.) Berkeley: University of California Press.

de Saussure, F. (1959). *Course in General Linguistics*. (C. Bally, A. Sechehaye, Ed, & W. Baskin, Trans.) New York: The Philosophical Library .

DeLanda, M. (2014). *A New Philosophy of Society: Assemblage Theory and Social Complexity*. London: Bloomsbury. doi:10.5040/9781472546043

Deleuze, G. (1995). *Negotiations* (M. Joughin, Trans.) New York: Columbia University Press.

Dolphijn, R. (2019). "There It Is Again". On Objects, Technologies, Science, and the Times. In J. Loh, & M. Coeckelbergh (Ed.), *Feminist Philosophy of Technology* (pp. 287-298). Stuttgart: J.B. Metzler. doi:10.1007/978-3-476-04967-4_15

Elias, N. (2000). *The Civilizing Process: Sociogenetic and Psychogenetic Investigations*. (E. Dunning, J. Goudsblom, S. Mennell, Ed, & E. Jephcott, Trans.) Malden: Wiley-Blackwell.

Fanon, F. (2008). *Black Skin, White Masks*. (C. L. Markmann, Trans.) London: Pluto.

Ferrando, F. (2014). Is the post-human a post-woman? Cyborgs, robots, artificial intelligence and the futures of gender: a case study. *European Journal of Futures Research*, 2 (1), 1-17. doi:10.1007/s40309-014-0043-8

Foucault, M. (2003). *Society Must be Defended: Lectures at the Collége de France – 1975-76*. (D. Macey, Trans.) New York: Picador.

Fowles, S. (2016). The perfect subject (postcolonial object studies). *Journal of Material Culture*, 21 (1), 9-27. doi:10.1177/1359183515623818

Hacking, I. (1998). *Rewriting the Soul: Multiple Personality and the Sciences of Memory*. Princeton: Princeton University Press.

Haenlein, M., & Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. California Management Review, 61 (4), 5-14. doi:10.1177/0008125619864925

Halberstam, J. (2011). The Queer Art of Failure. Durham: Duke University Press.

Hall, S. (2008). The West and the Rest: Discourse and Power. In S. Hall, & B. Gieben (Ed.), *Formations of Modernity* (pp. 276-320). New Hampshire: The Open University.

Haraway, D. (1991). *Simians, Cyborgs and Women: The Reinvention of Nature*. (S. During, Ed.) New York: Routledge.

Hirosue, S., Kera, D., & Huang, H. (2015). Promises and Perils of Open Source Technologies for Development: Can the "Subaltern" Research and Innovate? In S. Hostettler, E. Hazboun, & J.-C. Bolay (Ed.), *Technologies for Development: What is Essential?* (pp. 73-80). Cham: Springer. doi:10.1007/978-3-319-16247-8_7

hooks, b. (1994). *Teaching to Transgress: Education as the Practice of Freedom*. New York: Routledge.

Keyes, O., Hitzig, Z., & Blell, M. (2021). Truth from the machine: artificial intelligence and the materialization of identity. *Interdisciplinary Science Reviews*, 46 (1-2), 158-175. doi:10.1080/03080188.2020.1840224

Kukathas, C. (2006). Moral Universalism and Cultural Difference'. In J. S. Dryzek, B. Honig, & A. Phillips (Ed.), *The Oxford Handbook of Political Theory* (pp. 581-600). Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780199548439.003.0032

Land, C. (2006). Becoming-Cyborg: Changing the Subject of the Social? In M. Fuglsang, & B. M. Sørensen (Ed.), *Deleuze and the Social* (pp. 112-132). Edinburgh: Edinburgh University Press. doi:10.3366/edinburgh/9780748620920.003.0006

Lazzarato, M. (2014). *Signs and Machines: Capitalism and the Production of Subjectivity*. (J. D. Jordan, Trans.) Los Angeles: Semiotext(e).

Lefort, C. (1991). *Democracy and Political Theory*. (D. Macey, Trans.) Cambridge: Polity Press.

Leroi-Gourhan, A. (1993). *Gesture and Speech*. (A. Bostock Berger, Trans.) Cambridge: MIT Press.

Long, J. H. (2019). Evolution Ain't Engineering: Animals, Robots, and the Messy Struggle for Existence. In T. Heffernan (Ed.), *Cyborg Futures: Social and Cultural Studies of Robots and AI Palgrave* (pp. 17-34). Cham: Macmillan. doi:10.1007/978-3-030-21836-2_2

MacLennan, B. (2009). History of Artificial Intelligence Before Computers. M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (s. 1763-1768). New York: IGI Global. doi:10.4018/978-1-60566-026-4.ch277

Macpherson, T., Churchland, A., Sejnowski, T., DiCarlo, J., Kamitani, Y., Takahashi, H., & Hikida, T. (2021). Natural and Artificial Intelligence: A brief introduction to the interplay between AI and neuroscience research. *Neural Networks*, 144, 603-613. doi:10.1016/j.neunet.2021.09.018

Madianou, M. (2021). Nonhuman humanitarianism: when 'AI for good' can be harmful. Information, *Communication & Society*, 2 4(6), 850-868. doi:10.1080/1369118X.2021.1909100

Martin, K. (2020). Subaltern perspectives in post-human theory. Anthropological Theory, 20 (3), 357-382. doi:10.1177/1463499618794085

Mauss, M. (1973). Techniques of the Body. Economy and Society, 2(1), 70-88. doi:10.1080/03085147300000003

Milesi, L. (2019). De-monstrating Monsters: Unmastering (in) Derrida and Cixous. *Parallax*, 25(3), 269-287. doi:10.1080/13534645.2019.16243

Mohamed, S., Png, M.-T., & Isaac, W. D. (2020). AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. Philosophy & Technology, 33, 659-684. doi:10.1007/s13347-020-00405-8

Moore, P. V., Martin, U., & Xanthe, W. (2018). Humans and Machines at Work: Monitoring, Surveillance and Automation in Contemporary Capitalism. In M. P. V., U. Martin, & W. Xante (Ed.), *Humans and Machines at Work* (pp. 1-16). London: Palgrave Macmillan. doi:10.1007/978-3-319-58232-0_1

Moran, J. C. (2019). Programming Power and the Power of Programming: An Analysis of Racialised and Gendered Sex Robots. In J. Loh, & M. Coeckelbergh (Ed.), Feminist Philosophy of Technology (pp. 39-57). Stuttgart: J.B. Metzler. doi:10.1007/978-3-476-04967-4_3

Nietzsche, F. W. (2006). The Nietzsche Reader. (K. A. Pearson, & D. Large, Ed) Malden: Blackwell Publishing.

Nora, P. (1989). Between Memory and History: Les Lieux de Mémoire. Representations, (26), 7-24. doi:10.2307/2928520

Pearce, T. (2013). "So it really is a series of tubes." Google's data centers, noo-politics and the architecture of hegemony in cyberspace. *Enquiry The ARCC Journal for Architectural Research*, 10 (1), 43-53. doi:10.17831/enq:arcc.v10i1.163

Phillips, A. (1996). Dealing with Difference: Politics of Ideas, or a Politics of Presence? In S. Benhabib (Ed.), *Democracy and Difference: Contesting the Boundaries of the Political* (pp. 139-152). Princeton: Princeton University Press.

Rouvroy, A., & Berns, T. (2013). Algorithmic governmentality and prospects of emancipation: Disparateness as a precondition for individuation through relationships? *Réseaux*, 177 (1), 163-196. doi:10.3917/res.177.0163

Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, 6 (1), 1-12. doi:10.1177/2053951718820549

Said, E. W. (2003). Orientalism. New York: Penguin Books.

Spivak, G. C. (1988). Subaltern Studies: Deconstructing Historiography. In R. Guha, & G. C. Spivak (Ed.), *Selected Subaltern Studies* (pp. 3-32). Delhi: Oxford University Press.

Spivak, G. C. (2010). Can the Subaltern Speak? In R. C. Morris (Ed.), *Can the Subaltern Speak? Reflections on the History of an Idea* (pp. 21-78). New York: Columbia University Press.

Stiegler, B. (2019). For a Neganthropology of Automatic Society. In T. Pringle, G. Koch, & B. Stiegler (Ed.), *Machine* (pp. 25-47). Minneapolis: University of Minnesota Press. doi:10.14619/1488

Taeihagh, A. (2021). Governance of artificial intelligence. *Policy and Society*, 40 (2), 137-157. doi:10.1080/14494035.2021.1928377

Terranova, T. (2014). Red Stack Attack! Algorithms, Capital and the Automation of the Common. In R. Mackay, & A. Avanessian (Ed.), *#Accelerate: The Accelerationist Reader* (pp. 379-399). Falmouth: Urbanomic.

Vattimo, G., & Zabala, S. (2011). *Hermeneutic Communism: From Heidegger to Marx*. New York: Columbia University Press.

Young, I. M. (2004). The Ideal of Community and the Politics of Difference. In C. Farrelly (Ed.), *Contemporary Political Theory: A Reader* (pp. 195-204). London: Sage. doi:10.4135/9781446215272.n23

ACTA INFOLOGICA 2023;7(1):187-196
dergipark.org.tr/acin

DOI: 10.26650/acin.1207100

RESEARCH ARTICLE

Social Media Sentiment Analysis for Solar Eclipse with Text Mining

Metin Madenciliği ile Güneş Tutulmasına Yönelik Duygu Durumu Analizi

Adem Korkmaz[1], **Selma Bulut**[2]

ABSTRACT

This study investigated public sentiment and emotional reactions to the Partial Solar Eclipse that occurred on October 25, 2022, by analyzing Twitter data. By leveraging text mining techniques and sentiment analysis, the research aims to understand individuals' diverse emotions and perspectives in response to this natural event. Tweets related to the solar eclipse were collected, preprocessed, and analyzed using different sentiment analysis libraries, including NRC and Bing, to identify positive and negative emotions associated with the event. Results revealed that 75% of the analyzed tweets were positive, indicating that most users had a favorable view of the solar eclipse, associating it with new beginnings and change. In contrast, fewer individuals expressed negative emotions, linking the event to potential harm or divine punishment. The study demonstrates the value of sentiment analysis in understanding public sentiment and emotional reactions to natural events, providing insights into societal perceptions and attitudes toward these occurrences.

Keywords: Text mining, sentiment analysis, social media analysis, partial solar eclipse, natural events

ÖZ

Bu çalışma, Twitter verilerini analiz ederek, 25 Ekim 2022'de meydana gelen Kısmi Güneş Tutulması'na karşı halkın duyarlılığını ve duygusal tepkilerini araştırıyor. Araştırma, metin madenciliği tekniklerinden ve duygu analizinden yararlanarak, bireylerin bu doğal olaya tepki olarak farklı duygularını ve bakış açılarını anlamayı amaçlıyor. Güneş tutulmasıyla ilgili tweetler toplandı, ön işleme tabi tutuldu ve olayla ilgili olumlu ve olumsuz duyguları belirlemek için NRC ve Bing dahil olmak üzere farklı duyarlılık analizi kitaplıkları kullanılarak analiz edildi. Sonuçlar, analiz edilen tweet'lerin %75'inin olumlu olduğunu ortaya çıkardı; bu, çoğu kullanıcının güneş tutulması hakkında olumlu bir görüşe sahip olduğunu ve onu yeni başlangıçlar ve değişimle ilişkilendirdiğini gösteriyor. Buna karşılık, daha az kişi, olayı potansiyel zarar veya ilahi ceza ile ilişkilendiren olumsuz duygular ifade etti. Çalışma, halkın duyarlılığını ve doğal olaylara karşı duygusal tepkilerini anlamada duygu analizinin değerini gösteriyor, toplumsal algılara ve bu olaylara yönelik tutumlara ilişkin içgörüler sağlıyor.

Anahtar Kelimeler: Metin madenciliği, duygu analizi, sosyal medya analizi, parçalı güneş tutulması, doğa olayları

[1](Dr.), Bandırma Onyedi Eylül University, Balıkesir, Turkiye
[2](Dr.) Kırklareli University, Vocational School of Technical Sciences, Kırklareli, Turkiye

ORCID: A.K. 0000-0002-7530-7715; S.B. 0000-0002-6559-7704

Corresponding author:
Adem KORKMAZ
Bandırma Onyedi Eylül University, Balıkesir, Turkiye
E-mail address: ademkorkmaz@bandirma.edu.tr

Submitted: 19.11.2022
Revision Requested: 12.04.2023
Last Revision Received: 26.05.2023
Accepted: 29.05.2023
Published Online: 20.06.2023

Citation: Korkmaz, A., Bulut, S. (2023). Social media sentiment analysis for solar eclipse with text mining. *Acta Infologica, 7*(1), 187-196. https://doi.org/10.26650/acin.1207100

1. INTRODUCTION

The rapid growth of social media platforms has transformed how people communicate and share their thoughts and emotions about various events and experiences. Natural phenomena, such as solar eclipses, often evoke a wide range of emotions and reactions from the public, making social media an invaluable data source for understanding societal perceptions and attitudes toward these events. By analyzing social media content, researchers can gain insights into how individuals interpret and respond to such occurrences, shedding light on the collective sentiment and emotions surrounding these events.

A solar eclipse is a celestial event that occurs with the linear alignment of the Sun, Moon, and Earth. Heavenly bodies are either lined up in a straight or somewhat straight line. A solar eclipse occurs when the Moon falls between the Sun and the Earth, casting a shadow on the Earth and, in some areas, wholly or partially blocking the Sun's light. The Moon does not move in orbit in the same plane as the Sun and Earth. When they line up, the eclipse season occurs twice a year (Nasa, 2022).

The first solar eclipse of the 21st century was the Total Solar eclipse on June 21, 2001. A solar eclipse occurs when the Moon passes between the Earth and the Sun, thus completely or partially obscuring the view of the Sun for a viewer on Earth. During a total eclipse, the moon completely covers the sun. The next Total Solar Eclipse will occur on April 8, 2024. In an annular eclipse, the Moon cannot completely cover the Sun and leaves a visible edge. The next Annular Solar Eclipse will be on October 14, 2023. In a partial eclipse, the Moon partially covers the Sun as it passes in front of it. Because the Earth's surface is curved, sometimes an eclipse can alternate between circular and total as the Moon's shadow moves over Earth. This is called a hybrid solar eclipse (Nasa Types, 2022).

In the 21st century, there will be 224 solar eclipses, including 77 partial, 73 annular, 68 total, and seven total and annular solar eclipses. There will be a maximum of four eclipses per year in the 21st century. These dates are 2011, 2029, 2047, 2065, 2076, and 2094 (Nasa, 2022).

Two solar eclipses took place in 2022. The first was the Partial Solar Eclipse on April 30, 2022; the other was on October 25, 2022. It could be seen in most of Europe, North Africa, the Middle East, and Western Asia. It started at 11:58:21, according to Istanbul time, and ended at 16:02:11. The time of the total eclipse was stated as 14:00:16 (TimeandDate, 2022). The Moon covers about 82% of the Sun. The dates of the seven total eclipses that will take place in the upcoming period are shown in Figure 1.

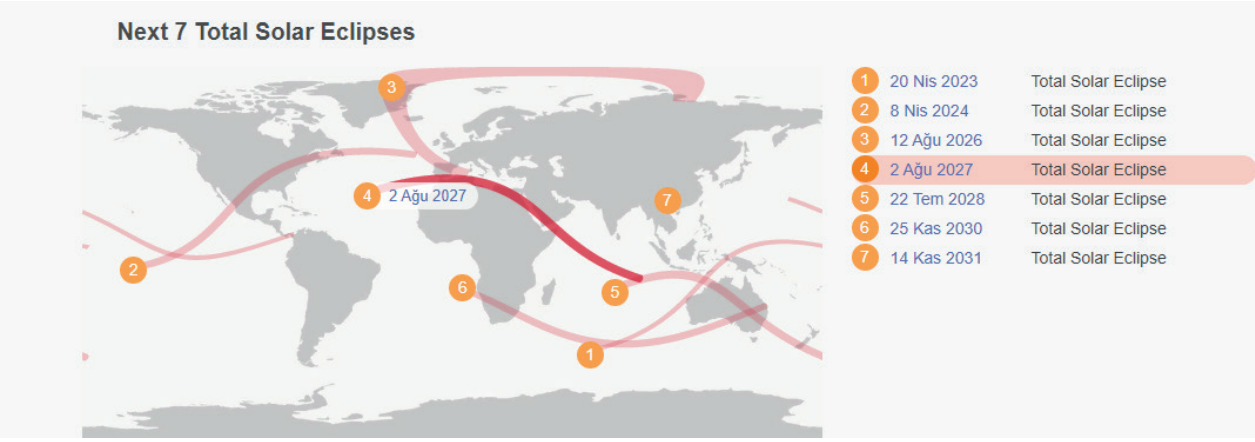


Figure 1. Date of Total Solar Eclipse

Many people believe that celestial conditions play a decisive role in life events. Scientifically, the effect of the Lunar Eclipse and Solar Eclipse on natural events is an undeniable issue (Javid, Tavaziani, & Shayanfar, 2022). Solar eclipses can affect living things on Earth. The first form of physical effect, the darkening that occurs during the transit of the Moon between the Sun and the Earth that prevents the sun's rays from contacting the Earth. The presence of light has a strong influence on animal behavior and hormones (Murdin, 2001). Studies in humans have shown that prolactin levels rise immediately after a solar eclipse (Boral, 1981). Prolactin is a hormone that regulates metabolism, the immune system, and the development of

the pancreas. Another physical effect is that during the Solar Eclipse, the Moon and Sun feel the combined force of both simultaneously as the gravity aligns. This unusual gravitational force's precise hormonal and behavioral effects are unknown (Norris, 2017).

The change in physical force that occurs on Earth during a solar eclipse is a shift in the electromagnetic field in the Earth's ionosphere. This electromagnetic field is caused by the electrical tension between the earth's surface's negative charge and the atmosphere's positive electric charge surrounding the world. Previous experiments show a high electrical voltage in our atmosphere during an eclipse. The shadow of the eclipse can cause changes in the ionosphere that are significant enough to affect radio wave propagation and possibly human physiology (Wang et al., 2022; Barad, Sripathi and England, 2022; Das, Barman, Pal, & Haldar, 2022; Chernogor, Garmash, Guo, Luo, Rozumenko, & Zheng, 2020).

Eclipses also have an essential astrological meaning. Eclipses in astrology represent periods of significant change and transformation. The solar eclipse opens the door to new beginnings and change, while the lunar eclipse releases energies that help us somehow complete a journey or change course. However, these changes only happen in stages. Transformation may begin in an eclipse, but the work continues if eclipses occur in the same set of signs (Winstersteen, 2022).

Solar eclipses have fascinated humanity for centuries, often evoking a wide range of emotions and reactions, from awe and wonder to fear and superstition (Espenak & Anderson, 2008). As extraordinary natural events, solar eclipses have been linked to various myths, legends, and religious beliefs throughout history, shaping how societies perceive and interpret these occurrences (Krupp, 1997). In contemporary times, solar eclipses continue to attract public interest and elicit diverse emotional responses, making them a relevant subject for sentiment analysis research.

Eclipses often come to the fore with dramatic effect in ancient cultures that have used mythology to explain celestial events. According to Swamy (2020), the Hindu culture in India prohibits food consumption during the eclipse. They believe that food spoils during this time, and if you are accidentally cut during an eclipse, the bleeding will not stop for a long time, which can leave a permanent scar on a person for the rest of their life. They keep pregnant women confined because they think this situation will affect the newborn child. In the United States, it is believed that an eclipse is a warning sign of the expected apocalypse. Hausa people in northern Nigeria noted that when there is a solar eclipse, it is usually not good. They gather in groups and wander the streets to pray for God's intervention (Izzuddin, Imroni, Imron, & Mahsun, 2022). Cofield (2017) The Toba people of South America believed that the red sky during an eclipse was the result of an attack by the spirits of the dead. Throughout history, people have approached such extraordinary natural phenomena similarly.

The advent of social media platforms, such as Twitter, has revolutionized how people share their thoughts, feelings, and experiences related to various events, including natural phenomena like solar eclipses (Java et al., 2007). The vast amount of user-generated content available on social media provides researchers with an invaluable data source for understanding public sentiment and emotions toward these events (Pak & Paroubek, 2010). Sentiment analysis, also known as opinion mining, is a subfield of natural language processing that involves extracting, classifying, and analyzing personal information from text (Liu, 2012). By applying sentiment analysis techniques to social media data, researchers gain insights into individuals' diverse emotions and perspectives in response to natural events such as solar eclipses.

In the age of information technology, social media platforms such as Twitter, Facebook, WhatsApp, and Instagram have become vital tools for communication during natural disasters and significant events. Their ability to facilitate real-time information dissemination has been pivotal in crisis response and management. Studies utilized these platforms to examine societal responses and attitudes toward various events in recent years. Only one study included the social effects associated with a solar eclipse. The survey by Goldy et al. analyzed 2,891,611 tweets concerning the Total Solar Eclipse in 2017. They found that the eclipse aroused awe among people on the road to wholeness. Individuals ultimately increased their tendencies toward humility, collective focus, commitment, and pro-sociality, enabling them to form collaborative social groups (Goldy, Jones, Piff, 2022).

A powerful research tool, Twitter data, has been used to assess the social impact of collective events on different issues. For example, Dore et al. identified temporal and spatial patterns in using emotional and cognitive words in tweets about the "Sandy Hook Elementary School" attack (Doré et al., 2015).

In the Garcia and Rimé studies, the first large-scale test of this theory was conducted by analyzing the digital traces of 62,114 Twitter users after the Paris terror attacks in November 2015. It supported the conclusion that collective feelings are associated with higher solidarity and elicit a community's social resilience after a disaster (Garcia & Rimé, 2019).

Similarly, another study constructed a natural disaster dataset of approximately 50,000 Twitter users' data about various natural disasters in the United States. Instead of merely examining positive or negative attitudes, it focused on extracting public attitudes and analyzing basic needs during disaster response, such as food, shelter, transportation, and medical supplies (Meng & Dong, 2020).

The case of the Kerala floods provides a compelling example of how sentiment analysis can offer actionable insights for governments and organizations. By analyzing trending keywords and sentiments during the floods, the study demonstrated how social media data could be used to understand public discussion trends, identify gaps in scheme implementation, and facilitate effective collaboration with influential Twitter users to disseminate crucial information (Mendon, Dutta, Behl, Lessmann, 2021).

In their study, Jones and Silver investigated the psychological responses to the 2018 false ballistic missile warning in Hawaii using Twitter data. The study analyzed 1.2 million tweets and 14,830 users. It showed that false alarms about dangerous threats raised anxiety and can persist in many people even after the danger disappeared (Jones & Silver, 2020).

The motivation and purpose of this study is to analyze people's emotions and perspectives on the "Partial Solar Eclipse" event through their social media interactions. After natural events and disasters, people often share their feelings and thoughts on social media platforms using text, videos, or pictures. By applying emotion analysis methods to these interactions, researchers gain insights into people's viewpoints on such events and begin to understand their societal impact. The study's primary goal was to investigate how people express their emotions regarding the solar eclipse in the context of supernatural events and determine the sentiment analysis of these expressions. The researchers focused on a partial solar eclipse to examine people's positive and negative emotions based on their recent experiences with the event.

2. Method

Using the "*pandas*" and "*snscrape*" libraries in Python programming, the researchers collected 19,285 tweets containing the phrase "*solar eclipse*", which were posted on Twitter between October 23-28, 2022. They analyzed the emotional content of these tweets using three different libraries:

NRC (National Research Council Canada) Emotion Lexicon: The lexicon contains English words and their associations with eight emotions and two sentiments. It associates words with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust), and two sentiments (negative and positive) and contains 13,875 words. The words with no associated sentiment are not listed in the lexicon (Mohammad & Turney, 2013).

Bing Lexicon: This lexicon contains a list of English positive and negative opinion words or sentiment words containing 6,786 words. This lexicon determines whether a text expresses a positive or negative sentiment. Each word is assigned a polarity, which can be positive or negative, but not both (Liu & Hu, 2004).

AFINN Lexicon: This lexicon lists pre-computed sentiment scores for 2,477 English words. The words are rated with an integer between -5 (negative) and +5 (positive), signifying the sentiment value of the word. The higher the score, the more positive the sentiment of the word. Similarly, the lower the score, the more negative the sentiment (Nielsen, 2011).

The researchers also considered user-specific factors, such as the number of followers and friends, to determine whether a Twitter account was real or fake.

Since the analysis relied on word-based dictionaries (NRC, AFINN, and Bing) that contain only English words, it was necessary to preprocess the tweets to ensure accurate analysis. The preprocessing steps included:

- Converting tweets to lowercase: This standardizes the text for a more straightforward analysis.

- Removing retweets: This prevents duplication of content in the analysis.
- Eliminating punctuation marks: This includes periods, exclamation points, and commas.
- Removing non-essential content: URL links, mentions of other users (using '@'), extra spaces, symbols, and numbers.
- Excluding common but meaningless words: Words like "The," "an," "I," "am," and "you" were removed, as they do not carry any emotional information.

After preprocessing the tweets consisting of 494,503 words, sentiment analysis was performed on the remaining 262,732 words. These preprocessing steps helped clean up and standardize tweets, making them suitable for sentiment analysis using the NRC, AFINN, and Bing dictionaries.

Sentiment analysis with NRC, Bing, and AFINN dictionaries begins by tokenizing the text into individual words. For NRC, words are matched to emotions and sentiments within the lexicon, with the text's overall sentiment or emotion determined by a cumulative count. Bing Liu's lexicon categorizes words as positive or negative, with the text sentiment discerned by the majority sentiment or the difference between positive and negative word counts. The AFINN lexicon assigns a numeric score to each word, with the total sentiment of the text computed by adding up these scores. R software packages such as "*tidytext*", "*tm*", "*tidyverse*", "*stringr*", "*syuzhet*", "*dplyr*", "*wordcloud*" and "*ggplot2*" were used for data analysis and visualization.

3. FINDINGS

This section analyzed the tweets posted to the "*Particle Solar Eclipse"* event dated October 25, 2022, containing the word "solar eclipse" on Twitter between October 23, 2022, and October 28, 2022. The study analyzed tweets in English using NRC, Bing, and Afinn libraries. Data such as the number of followers and friends of the users were analyzed, and it was determined whether the user account was real or fake.

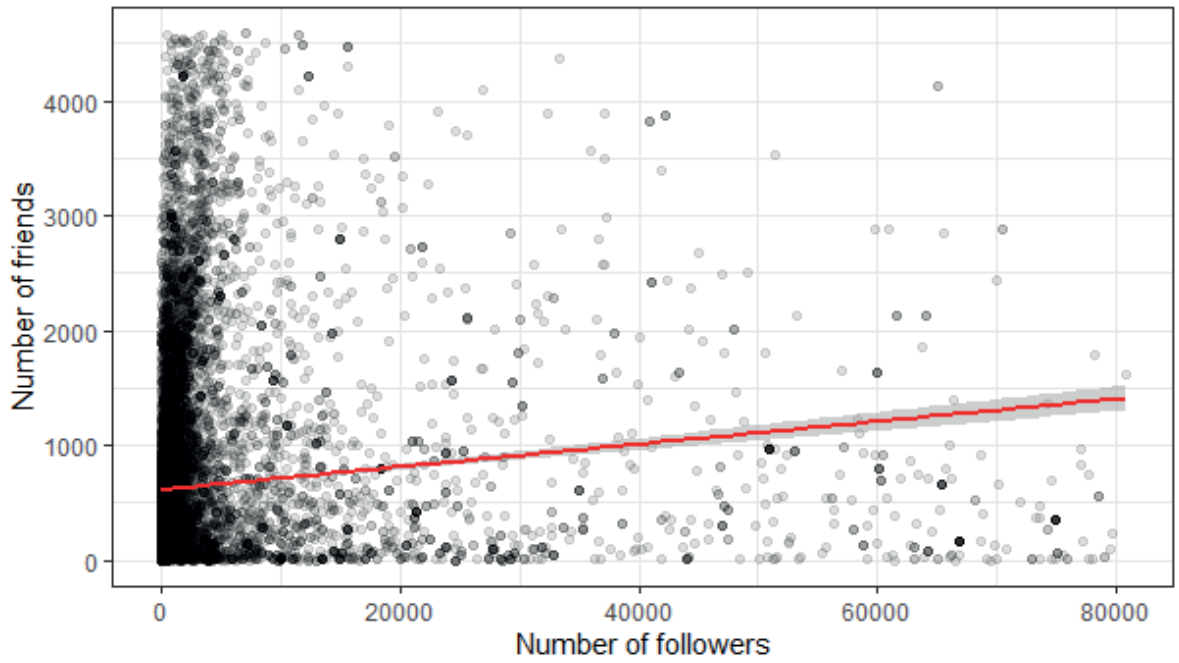


Figure 2. Followers and following numbers of users

User account information is provided in Figure 2 to determine whether Twitter data comes from real users or bots or whether tweets are generally sent from high-follower accounts. It shows that the number of people followed by the users tweeting in the context of the words "*solar eclipse*" is more than the number of followers. This indicates that users are not phenomena; they are regular users.

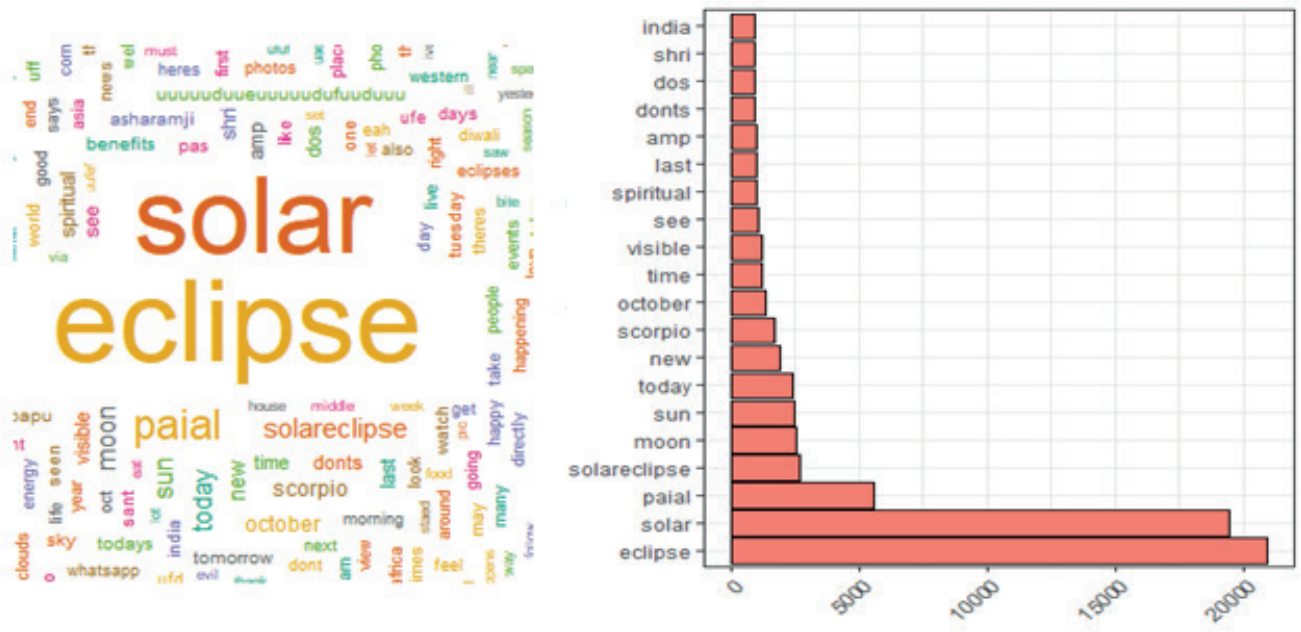


Figure 3. The most used word cloud and 20 words in tweets

When the words used in the tweets in Figure 3 and their intensities are examined, it shows that terms such as "*paial*," "*moon,*" and "*sun*" related to solar eclipse are used extensively. This indicates that the content of the tweets about the "*solar eclipse*" is directly proportional to the keywords.

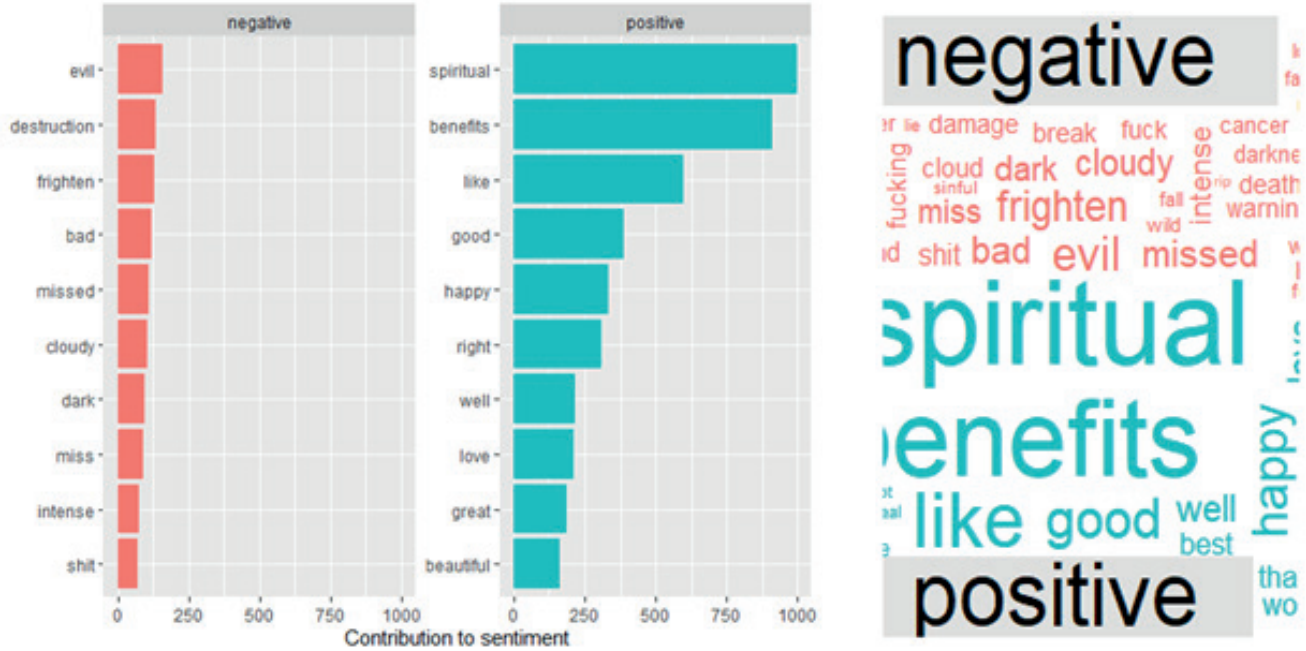


Figure 4. Top 10 words and clouds in negative/positive emotions

Analysis results for the "Partial Solar Eclipse" event using the Bing library are given in Figure 4. Positive emotions were associated with words like "*spiritual*," "*benefit*," "*like*," "*good*," *and* "*happy*." Using "*spiritual*" and "*benefit*" implies that some people saw the solar eclipse as a meaningful, potentially uplifting experience or connected it to rituals and positive beliefs about natural events. This suggests that the solar eclipse provided inspiration, fascination, or personal enrichment for these individuals. On the other hand, negative emotions were linked to words such as "*evil*," "*destruction*," "*frighten*,"

"*bad*," and "*missed*." The presence of "*evil*" and "*destruction*" indicates that some people viewed the solar eclipse as a negative or even threatening event, perhaps due to cultural or religious beliefs that associate extraordinary natural phenomena with divine wrath or punishment. The word "*frighten*" reflects the fear or anxiety some individuals might have experienced about the event, while "missed" could refer to feelings of regret for not being able to witness the eclipse. Some individuals found it a positive, inspiring occurrence, while others perceived it as an ominous or harmful event. This highlights the complex and varied ways people react to and interpret natural phenomena like solar eclipses.

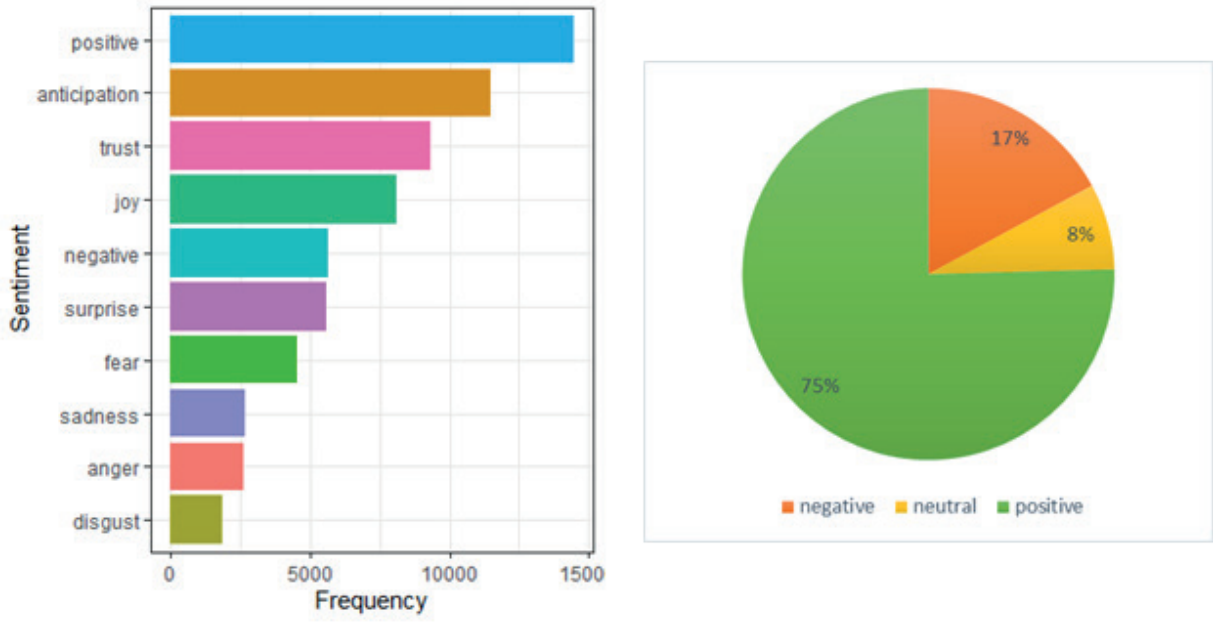


Figure 5. Emotion ratios of words and tweets

For the overall sentiment of the tweets, the words in each tweet were given a score of +1 for each positive emotion (positive, joy, anticipation, confidence, and surprise) and -1 for each negative word, using the NRC dictionary. If the sum of the scores for each tweet was greater than 0, it was marked as positive, equal to 0, neutral, and less than 0, as unfavorable. As a result of the analysis in Figure 6, the emotional state of the users' tweets was determined as 75% positive, 17% negative, and 8% neutral.

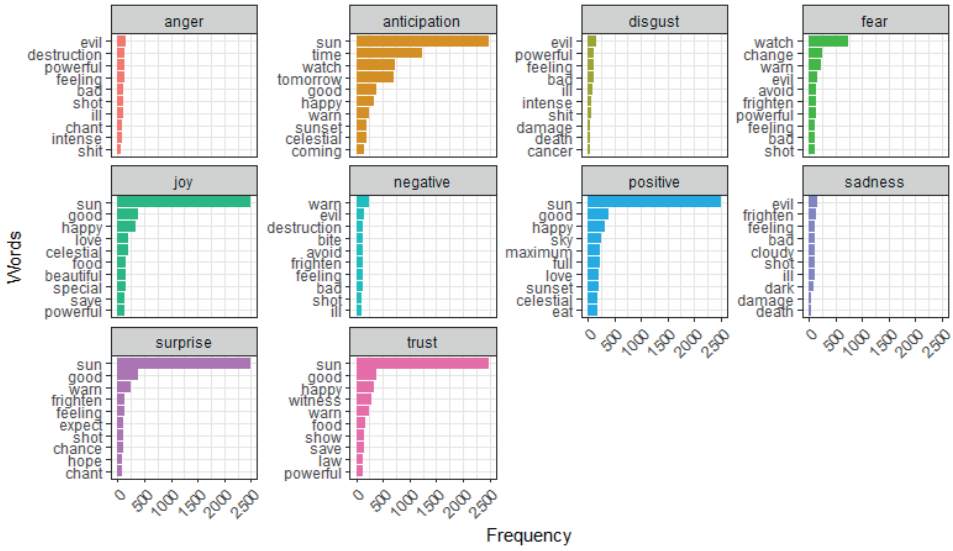


Figure 6. The number of uses of the ten most used words for each Emotion

In Figure 6, a word sentiment analysis of tweets with the keyword *'solar eclipse'* was performed using the NRC dictionary. For positive emotions, the words "*sun*," "*time*," "*good*," and "*happy*" appeared frequently. These words suggest that the positive emotions expressed in the tweets were related to the solar eclipse event itself ("*sun*" and "*time*"), as well as the enjoyable or exciting aspects of the experience ("*good*" and "*happy*"). The tweets with positive emotions were mainly associated with joy, surprise, trust, and anticipation, indicating that people found the solar eclipse fascinating, awe-inspiring, and something to look forward to. On the other hand, the most common words in tweets with negative emotions were "*evil*," "*destruction*," "*change*," and "*warning*." These words imply that some people associated the solar eclipse with negative or ominous connotations. For instance, "*evil*" and "*destruction*" suggests fear or anxiety around the event, while "*change*" might reflect uncertainty or unease about the effects of the eclipse. The word "warning" could indicate that some users perceived the solar eclipse as a sign of impending danger or trouble.

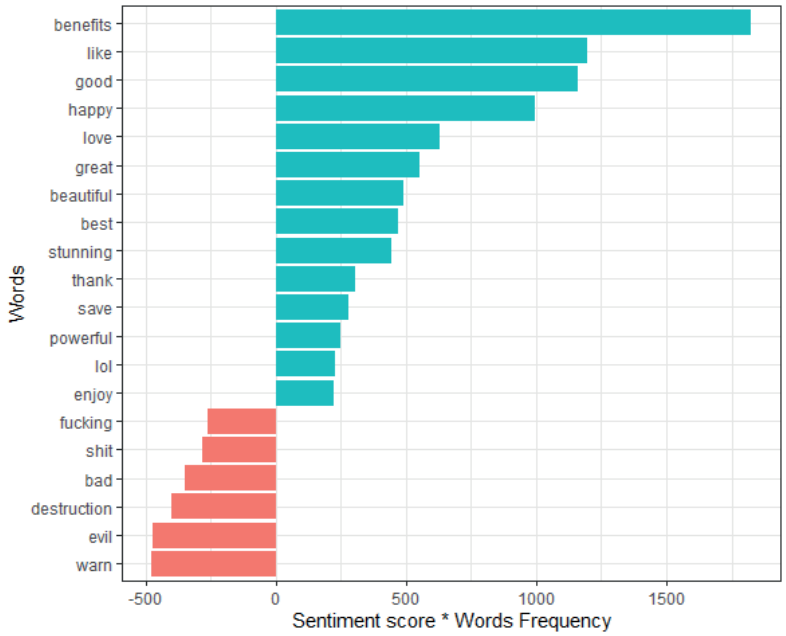


Figure 7. Positive and negative emotion usage scores of the most used words

In Figure 7, the Afinn library was used for the usage rates of the words used in tweets with positive and negative emotions. For tweets with positive emotions, the most frequently used words were "benefit," "like," "good," "happy," and "love." These words indicate that people enjoyed the event and had positive feelings about it, as they associated the eclipse with happiness, appreciation, and even affection. In contrast, the most used words in tweets with negative emotions were "warn," "evil," "destruction," "bad," and "shit." These words suggest that some people had negative perceptions of the solar eclipse, associating it with potential harm, malevolent forces, or unpleasant experiences.

4. DISCUSSION AND CONCLUSION

In this study, which was conducted to examine the effects of the Partial Solar Eclipse, which took place on October 25, 2022, on people, it showed that 75% of the tweets sent were positive. It turned out that the users who posted these tweets were average, not phenomena. As a result of the analysis, tweets of Twitter users participating in the research consisted of positive words such as "*spiritual*, *benefit*, *like*, *good* and *happy*" about the solar eclipse. This supports the idea in Astrology that a Solar Eclipse opens a portal to new beginnings and change. So, people support possible change by creating good thoughts and feelings in this way.

The absence of negative words in the first 20 most used words is directly proportional to the fact that the positive emotion category contains approximately two times more content than the negative emotion category. This shows that the concept of solar eclipse creates a positive impression on users tweeting in English.

When the words were evaluated on a dictionary basis, it was determined that "*evil*, *destruction*, *frighten*, *bad,* and *missed*" came first among the most negative words according to the positive and negative emotion categories in the context of the Bing library. Using the words "*benefit*" and "*spiritual*" about positive words could be considered rituals that people perform during natural events and positive expectations. On the other hand, the words "*evil*" and "*destruction*" can be explained as people's view of extraordinary natural events as divine disasters and punishment. This situation determined that the most used words for positive emotion in the NRC dictionary were *benefits*, *like*, *good*, *happy,* and *love*. The most commonly used words for negative emotion were *warn, evil, destroy, bad*, and *shit*. Similar emotional states emerged even when the data were analyzed with different dictionaries.

The findings of this study resonate with emerging research suggesting social media platforms like Twitter can be used as practical tools to capture public sentiment during significant natural events (Java et al., 2007; Pak & Paroubek, 2010). The discovered preponderance of positive sentiment (75%) corroborates Goldy et al.'s (2022) research that events like the Total Solar Eclipse in 2017 often catalyzed collective awe, humility, and pro-social behavior, leading to collaborative social groups. The frequent use of positive words such as "spiritual," "benefit," "like," "good," and "happy" in tweets corresponds with the astrological belief that a solar eclipse symbolizes new beginnings and change, implying that people may express positive sentiments to embrace this potential transformation (Goldy et al., 2022). The scarcity of negative terms among the top 20 most used words suggests a generally positive impression of the partial solar eclipse among Twitter users, aligning with Meng & Dong's (2020) observations on societal responses to significant events. The analysis of word usage based on Bing library and NRC dictionary unveiled that terms like "benefit" and "spiritual" could mirror rituals or positive expectations associated with natural events, while words such as "evil" and "destruction" might signify how some individuals perceive extraordinary natural events as divine punishments or disasters, similar to the psychological reactions observed in the 2018 false ballistic missile warning in Hawaii (Jones & Silver, 2020). By employing text mining and sentiment analysis methodologies, this research enriches the sentiment analysis and public sentiment research field, providing a robust foundation for future studies by offering insights into societal perceptions and attitudes towards natural events like the 2022 Partial Solar Eclipse (Liu, 2012). As for future research and study limitations, the proposal to diversify data sources beyond Twitter and to consider other eclipse events for a more thorough sentiment analysis is well-aligned with trends in social science research. Expanding the scope of data sources and focusing on longitudinal analysis could foster more profound insights into public sentiment and responses to natural events in future research (Mendon, Dutta, Behl, Lessmann, 2021).

This study contributes to sentiment analysis and public sentiment research by analyzing the public's emotional responses to the Partial Solar Eclipse of 2022 using Twitter data. Its contributions include demonstrating the effectiveness of text mining and sentiment analysis methods, providing a basis for future research on different types of natural events, enabling cross-cultural comparisons, and encouraging longitudinal analysis of public sentiment over time. This study offers valuable insights and a foundation for future research on societal perceptions and attitudes toward natural events.

In the context of the study's limitations for future research, more diversity and data sources can be obtained from different social media tools since it is only done on Twitter data. More comprehensive sentiment analysis can be done by considering the other partial eclipse event in the same year. Since the closest Total Solar Eclipse will take place on April 20, 2023, it will be possible to see whether there is a difference in people's perspectives on Total and Partial Solar Eclipse with the tweets to be sent in this date range. People's view of the situation can be analyzed by comparing it to the closest lunar eclipse.

Peer-review: Externally peer-reviewed.
Author Contributions: Conception/Design of Study- A.K., S.B.; Data Acquisition- A.K.; Data Analysis/Interpretation- A.K., S.B.; Drafting Manuscript- A.K., S.B.; Critical Revision of Manuscript- S.B.; Final Approval and Accountability- A.K., S.B.; Material and Technical Support- A.K., S.B.; Supervision- A.K., S.B.
Conflict of Interest: The authors have no conflict of interest to declare.
Grant Support: The authors declared that this study has received no financial support.

REFERENCES

Barad, R. K., Sripathi, S., & England, S. L. (2022). Multi-Instrument Observations of the Ionospheric Response to the 26 December 2019 Solar Eclipse Over Indian and Southeast Asian Longitudes. Journal of Geophysical Research: Space Physics, 127(9), e2022JA030330.

Boral, G. C., Mishra, D. C., Pal, S. K., & Ghosh, K. K. (1981). Effects Of Total Solar Eclipse On Mental Patients A Clinicobiochemical Correlation. Indian Journal of Psychiatry, 23(2), 160.

Chernogor, L. F., Garmash, K. P., Guo, Q., Luo, Y., Rozumenko, V. T., & Zheng, Y. (2022). Some features of the ionospheric radio wave characteristics over China observed during the solar eclipse of 21 June 2020. Radio Science, 57(10), 1-21.

Cofield, C. (2017). Eclipse superstitions are a thing of the past and the present. Retrieved from https://wwww.space.com

Das, B., Barman, K., Pal, S., & Haldar, P. K. (2022). Impact of Three Solar Eclipses of 2019–2020 on the D-Region Ionosphere Observed from a Subtropical Low-Latitude VLF Radio Station. Journal of Geophysical Research: Space Physics, 127(8), e2022JA030353.

Doré, B., Ort, L., Braverman, O., & Ochsner, K. N. (2015). Sadness shifts to anxiety over time and distance from the national tragedy in Newtown, Connecticut. *Psychological science*, *26*(4), 363-373.

Espenak, F., & Anderson, J. (2008). Total Solar Eclipse of 2008 August 01. NASA/TP-2007-214149. Retrieved from https://eclipse.gsfc.nasa.gov/SEpubs/20080801/rp.html

Garcia, D., & Rimé, B. (2019). Collective emotions and social resilience in the digital traces after a terrorist attack. *Psychological science*, *30*(4), 617-628.

Goldy, S. P., Jones, N. M., & Piff, P. K. (2022). The social effects of an awesome solar eclipse. *Psychological Science*, *33*(9), 1452–1462.

Izzuddin, A., Imroni, M. A., Imron, A., & Mahsun, M. (2022). Cultural myth of eclipse in a Central Javanese village: Between Islamic identity and local tradition. HTS Theological Studies, 78(4), 1-9.

Java, A., Song, X., Finin, T., & Tseng, B. (2007, August). Why we Twitter: understanding microblogging usage and communities. In Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis (pp. 56–65).

Javid, M., Tavaziani, Z., & Shayanfar, S. (2022). Good and Bad Luck of Days and Their Avicennian Explanation by Nasir al-Din al-Tusi. Philosophical Meditations, 12(28), 15-40.

Jones, N. M., & Silver, R. C. (2020). This is not a drill: Anxiety on Twitter following the 2018 Hawaii false missile alert. *American Psychologist*, *75*(5), 683.

Krupp, E. C. (1997). Skywatchers, Shamans & Kings: Astronomy and the Archaeology of Power. New York: Wiley.

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1–167.

Liu, B., & Hu, M. (2004). Opinion mining, sentiment analysis, and opinion spam detection. Dosegljivo: https://www. cs. uic. edu/ liub/FBS/sentiment-analysis. html# lexicon.[Dostopano 15. 2. 2016].

Mendon, S., Dutta, P., Behl, A., & Lessmann, S. (2021). A hybrid approach of machine learning and lexicons to sentiment analysis: Enhanced insights from twitter data of natural disasters. Information Systems Frontiers, 23, 1145-1168.

Meng, L., & Dong, Z. S. (2020). Natural Hazards Twitter Dataset. arXiv preprint arXiv:2004.14456.

Mohammad, S. M., and Turney, P. D. (2013) ‘‘NRC Emotion Lexicon’’, National Research Council of Canada: 234, https://doi.org/10.4224/21270984.

Murdin, P. (2001). Effects of the 2001 total solar eclipse on African wildlife. Astronomy & Geophysics, 42(4), 4–4.

Nasa. (2022). Fıve Mıllennıum Catalog Of Solar Eclıpses.https://eclipse.gsfc.nasa.gov/SEcat5/catalog.html

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. arXiv preprint arXiv:1103.2903.

Norris D. (2017). Mind-Body Influences of a Solar Eclipse. https://www.mindbodyjournal.com/mindfulness/mind-body-influences-of-a-solar-eclipse/

Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In LREc (Vol. 10, No. 2010, pp. 1320-1326).

Swamy, V.K. (2020). Solar eclipse: Myths and superstitions from around the world. https://www.readersdigest.in/oddsand-eclipse-story-solar-myths-andsuperstion.

TimeandDate. (2022). Total Solar Eclipses Worldwide – Next 10 Years. https://www.timeanddate.com/eclipse/list-total-solar.html

TimeandDate. (2022).25 Ekim 2022 Partial Solar Eclipse. https://www.timeanddate.com/eclipse/solar/2022-october-25

Types of Solar Eclipses. Nasa. (2022). https://solarsystem.nasa.gov/eclipses/about-eclipses/types/

Wang, J., Sun, Y. Y., Yu, T., Wang, Y., Mao, T., Yang, H., ... & Huang, G. (2022). Convergence effects on the ionosphere during and after the annular solar eclipse on 21 June 2020. *Journal of Geophysical Research: Space Physics*, e2022JA030471.

Winstersteen J. (2022). How The Lunar Eclipse Will Affect You https://www.yogajournal.com/lifestyle/astrology/how-the-lunar-eclipse-will-affect-you/

ACTA INFOLOGICA 2023;7(1):197-207
dergipark.org.tr/acin

DOI: 10.26650/acin.869786

RESEARCH ARTICLE

LISS'nAP: A Peer-Advising Collaborative Learning Approach for Object-Oriented Programming Learning

LISS'nAP: Nesne Yönelimli Programlama için Üç Aşamalı Akran Danışmanlığı İşbirliğine Dayalı Öğrenme Yaklaşımı

Rochdi Boudjehem[1], Ali Benyounes[2], Yacine Lafifi[3]

[1](Dr.) LabSTIC laboratory, University 8 Mai 1945, Guelma, Algeria
[2](MSc.) Computer Science Department, University 8 Mai 1945, Guelma, Algeria
[3](Prof. Dr.) Computer Science Department, LabSTIC laboratory, University 8 Mai 1945, Guelma, Algeria

ORCID: R.B. 0000-0002-9998-4833;
A.B. 0000-0002-5821-1025;
Y.L. 0000-0001-8232-4196

Corresponding author:
Rochdi BOUDJEHEM
University May 8 1945 Guelma, Department of Computer Science
E-mail address: rochdiboudjehem@gmail.com, boudjehem.rochdi@univ-guelma.dz

Submitted: 28.01.2021
Revision Requested: 04.08.2021
Last Revision Received: 02.06.2023
Accepted: 02.06.2023
Published Online: 26.06.2023

Citation: Boudjehem, R., Benyounes, A., Lafifi, Y. (2023). LISS'nAP: a peer-advising collaborative learning approach for object-oriented programming learning. *Acta Infologica, 7*(1), 197-207.
https://doi.org/10.26650/acin.869786

ABSTRACT
Among the existing programming paradigms, two of them are the most familiar and the most used by the majority of programers: procedural programming and object-oriented programming. To teach students programming, most existing curricula, especially academic ones, begin by teaching the basics of procedural programming, before moving on to introduce the concepts of object-oriented programming. Unfortunately, this transition is not always smooth for students, and the concepts of object-oriented programming can be perceived as problematic and confusing for some students who have trouble adjusting to the new programming mode and fail to find a decent object-oriented model for a given problem. Some scientists believe that the problem lies in the transition of vision when modeling problems from one vision that is familiar to students to a new one that is object-oriented. Some scientists think that the problem lies in the transition itself when modeling problems from a familiar vision to a new object-oriented one. Other scientists claim that the problem may not reside in the object-oriented paradigm itself, but rather in the existing tools available to teach it, such as the languages and the environments. In this article, we propose a new collaborative learning method for learning object-oriented programming, which can be also used to teach any other course that shares a similar pedagogical structure.
Keywords: Problem-based learning, collaborative learning, programming learning, e-learning

ÖZ
Mevcut programlama paradigmaları arasında, programcıların çoğu tarafından en tanıdık ve en çok kullanılan iki tanesi prosedürel programlama ile nesne yönelimli programlamadır. Öğrencilere programlamanın nasıl yapılacağını öğretmek için, özellikle akademik olan çoğu mevcut müfredatta, nesne yönelimli programlama kavramlarına geçmeden önce, prosedürel programlamanın temellerini öğreterek başlanır. Ne yazık ki, bu geçiş öğrenciler için her zaman "pürüzsüz" değildir. Nesne yönelimli programlama kavramları, yeni programlama moduna uyum sağlamakta güçlük çeken ve bir nesne için düzgün bir nesne yönelimli model bulamayan bazı öğrenciler için sorunlu ve kafa karıştırıcı olarak algılanabilir. Bazı bilim insanları, öğrencilerin problemleri bilinen bir vizyondan nesne yönelimli yeni bir vizyona modellerken, asıl sorunun vizyon geçişinde yattığına inanmakta. Diğer bazı bilim insanları ise, sorunun nesneye yönelik paradigmanın kendisinde değil, onu öğretmek için kullanılan diller ve ortamlar gibi araçlarda olabileceğini iddia etmekte. Bu makalede, nesne yönelimli programlamayı öğrenmek için yeni bir işbirliğine dayalı öğrenme yöntemi önermekteyiz. Bu yöntem, benzer pedagojik yapıya sahip başka derslerde de kullanılabilir.
Anahtar Kelimeler: Probleme dayalı öğrenme, işbirlikçi öğrenme, programlama öğrenimi, e-öğrenme

1. INTRODUCTION

E-Learning has been widely used in several fields and scientific disciplines recently. It has proven its effectiveness in making the learning process more flexible and manageable, especially during difficult times and emergencies that the world is witnessing more than ever, such as during wars or the spread of epidemics such as the COVID-19 pandemic. As in many fields, education has undergone substantial transformations, especially with the spread of Information and Communication Technologies (ICT). These changes can be seen through the colossal breakthrough achieved in the way the remote education field is presented, where the content of the course is increasingly adapted to the needs of each distinct learner.

Furthermore, specific characteristics of each taught subject are taken into account. Computer Science was among the first disciplines involved in e-learning, considering the ability of its learners to assimilate the use of this new learning form comfortably. Several e-learning environments and systems have been developed to cover different computer science topics, from logic and algorithmics to programming, including its different paradigms, especially the Object-Oriented Programming (OOP) paradigm. Many students often see OOP concepts as problematic, as they fail to develop good object-oriented designs for a given problem (Dillenbourg, 1999; Henri & Lundgren-Cayrol, 1998, 2001; Laal & Ghodsi, 2012; Rogalski, 1998). One commonly used technique to teach programming languages is Problem-Based Learning (PBL), which has proven effective in many educative fields like medicine and computer science (Koh et al., 2008; Walldén & Mäkinen, 2014; Wood, 2008).

Furthermore, it is always tricky for students to understand object-oriented concepts, as it is equally difficult for teachers to transmit them to students. Indeed, the difficulty is mainly related to the transition from the usual procedural modeling of problems to an object-oriented one. Research, as stated in the work of Stroustrup (1994), indicates that the programmer takes about an average of 6 to 18 months to change his mindset from a procedural to an object-oriented view of the world, which implies that the change is difficult.

For Kölling (1999), it is not object-oriented learning that causes problems, but the tools (languages and environments) available to teach it. For this author, the programming languages used are too complex, the programming environments (if they exist) are too confusing, and some of the systems and languages used for teaching were developed for professionals, which makes the situation more complex for students, who fall into the difficulty of using them in learning object-oriented concepts.

In the same sense, other authors see that for beginning programmers, as is the case for university students, identifying objects and creating an object-oriented program design is very difficult (Robins et al., 2003).

Unfortunately, most developed systems focus on the learner as an individual and have no interaction with his peers, making the learner feel lonely and increasing his sense of isolation. Learning with peers can positively influence the learner, which is well known in traditional or distance learning environments where many other skills can be developed, such as soft skills (Tadjer et al., 2020). That is why many works have been conducted on the benefits of using collaboration between learners inside distance learning environments, even though this kind of learning requires more effort and time than individual learning (Bremgartner & de Magalhaes Netto, 2012; Hank & Chikh, 2013; Hu & Xu, 2013; Lafifi et al., 2010; Troussas et al., 2014).

From a constructivist viewpoint, "Peer Interaction" or "Peer Dialogue" is a critical part of the "Peer Advising" course. Vygotsky's (1978) "Zone of Proximal Development" (ZPD) is one of the theoretical bases for "Peer Interaction." It states that social interaction between individuals helps them build their cognitive development, where individuals improve their skills through guidance provided by more experienced ones. Based on this assumption, "Peer Interaction" can provide a platform for peers to expand their knowledge by working together in socially constructed activities, as it allows peers to be "engaged in solving problems and building knowledge" (Merrill Swain & Lapkin, 2000). According to Villamil & De Guerrero (1996), peer-to-peer activities are the social pillar of cognitive process development. Indeed, the exchange of ideas during interaction allows peers to offer and receive help, which allows them to improve their knowledge. Furthermore, because dialogue occurs as part of a socially constructed knowledge-building activity "directly indicative of mental processes" (Merill. Swain, 1995), it becomes a primary source to understand how peers construct knowledge by rigorously analyzing the involved dialogue.

Peer advising involves "students helping students." It highlights the "reciprocal interaction" of peers helping each other through collaboration during the learning process, taking a less formal, less-structured, and non-authoritarian approach. A peer advisor is a learner with more experience in a particular subject area who assists a peer student or a group of peers (less experienced learners) by guiding learning in that area. Although advisors and advisees learn together in the same course, the advisor assumes greater responsibility for providing guidance based on the needs of the advisee. According to Mynard and Carson (2014), there are at least three potential advantages that could be derived from peer advising:

1. Peer responses could be friendlier and more supportive than the teacher's (Rollinson, 2005), making the learning atmosphere anxiety-free and comfortable and helping learners concentrate on learning (Ellis, 1994).
2. By embracing collaborative learning in peer advising, the advisor and the advisee could learn from each other in groups and work together to reach a higher level of performance (Donato, 1994). Furthermore, peer advising helps peers engage in a healthy relationship (Barman & Benson, 1981), where the advisor and the advisee share ideas and develop their knowledge together (Villamil & De Guerrero, 1996).
3. The Reciprocal Interaction of Peer Advising could benefit both the advisor and the advisee in developing their autonomy (Assinder, 1991).

This paper presents LISS'nAP (Listen-Up), a new approach that relies on Peer Advising and collaboration between learners to learn OOP concepts –or any other concepts with a similar pedagogical structure. It addresses these research questions:

- Is there an effective way to group learners to benefit more significantly from the learning system?
- If so, how many learners to assign to each group, and how to assign and assess them?
- Finally, How can a collaborative learning strategy improve OOP learning?

With those questions in mind, we propose to define a new approach that allows learners to deal with many difficulties experienced when learning this paradigm by helping each other. We propose a remote learning system that embodies and helps test the proposed approach.

This paper is organized as follows: Section 2 presents a literature review on programming languages and collaborative learning. In section 3, we give our proposed approach, and in section 4, we offer our conclusion and future works.

2. MATERIALS AND METHODS:

2.1. Online Object-oriented Programming learning

Object-Oriented Programming (OOP) is a fascinating and attractive yet crucial course for the curriculum of a university's computer science students. It is one of the most popular programming paradigms and is highly demanded by technological development giants. However, learning the principles and concepts of this programming paradigm has proven to be tricky and sometimes complicated, leading researchers in this field to propose new approaches and methods to improve the quality of teaching this paradigm. In this sense, several approaches and techniques have been implemented in computer systems designed to solve the challenges and difficulties encountered when learning OOP. In Table. 1, we gathered and compared all found OOP's existing learning systems.

After analyzing the collected works, we have concluded that most of the reviewed papers focused on individual learning techniques that treat each learner separately and omit the fact that the human being is a social creature who can learn by observing his peers and can acquire knowledge from the experiences of others. As for the "Peer Advising" perspective, most of the works we found in the literature used this technique in teaching human languages, and we found no work that used it to teach programming language concepts or paradigms. Indeed, these reasons motivated us to use "Peer Advising" in a new collaborative approach to help students learn OOP programming concepts.

Table 1
Comparative study of OOP's existing learning systems

Authors	Taught Course	Studied Programming Language	Used Technique	Implemented System Type	Implemented System Name	Online / Offline	Assessment?
(Cypher & Halbert, 1993)	*Procedural Programming*	/	*Visual & Example Based*	*Environment*	Pygmalion	*Offline*	*No*
(Benabbou & Hanoune, 2006)	*Algorithmic*	/	*Example-Based & Recommendation*	*Environment*	EasyAlgo	*Offline*	*No*
(Guibert et al., 2006)	*Procedural programming*	/	*Example Based*	*Environment*	MELBA	*Offline*	*Yes*
(Sims & Bubinski, 2011)	*General programming*	*Python, PHP, CSS, JavaScript, HTML*	*Online Learning*	*Platform*	CodeAcademy	*Online*	*No*
(Rose, 2003)	*OOP*	*Java, C++, Visual Basic*	*Visualization & UML Graphic Modeling*	*Tool*	Rational Rose	*Offline*	*No*
(Kölling, 1999a)	*OOP*	*Java*	*Visualization & UML Graphic Modeling*	*Environment*	BlueJ	*Offline*	*No*
(Esteves & Mendes, 2004)	*OOP*	*Java*	*Visualization & Simulation*	*Environment*	OOP-Anim	*Offline*	*No*
(Cheung, 2006)	*OOP*	*Java*	*Visualization & UML Graphic Modeling*	*Environment*	WEBLOOP	*Online*	*No*
(Yan, 2009)	*OOP*	*Java*	*Serious Game*	*Environment*	Greenfoot	*Offline*	*No*
(Djelil et al., 2015)	*OOP*	*C++*	*Serious Game*	*Virtual 3D Game*	PrOgO	*Offline*	*Yes*
(Seng et al., 2018)	*OOP*	/	*Serious Game*	*Mobile 2D Game*	Odyssey of Phoenix OOP	*Offline*	*Yes*

2.2. Participants and Data Collection

This paper proposes a new approach based on collaborative learning to assist students while learning the OOP concepts. In addition to assisting students in understanding and practicing OOP, it can help them overcome difficulties while learning this paradigm.

To reach the desired goals, we first conducted a survey study to uncover everyday problems and identify difficulties faced by students while learning the OOP concepts. Then, we proposed a new approach to assist students in overcoming difficulties while learning OOP using the survey results. After that, we implemented a new prototype system. Finally, we experimented with a group of students using the developed prototype.

The survey conducted with Computer Science students regarding OOP learning at Guelma University (Algeria) comprised 91 participants who answered 28 questions. The survey questions covered different topics about their experience with OOP learning to gather the required data. It was made available to students using two methods: online via a Google form and print-outs handed to students manually. The survey was prepared according to the Likert Scale guidelines for most questions (23 out of 28), using precise and well-crafted options to obtain the most reliable answers possible. The remaining five questions were either multiple-choice or with free answers.

3. RESULTS

3.1. Survey reliability and validity:

To measure the validity of the Survey, we used Pearson's Correlation Coefficient with N=48. In Pearson's Critical Table, with a Degree of Freedom- = N-2 = 46 and α=0.05, the critical value was 0.3044. Applying SPSS Bivariate Correlation Analysis, the obtained correlation table shows that 16 questions out of 23 were valid with Pearson's value > 0.3044 and a p-value<0.05 (Table. 2).

To measure the reliability of the Survey, we used the Cronbach's Alpha test, which gave us a Cronbach's Alpha value of 0.525, which is close to 0.6, indicating that the questionnaire is reliable (Table. 3).

Table 2
Case Processing Summary

		N	%
Cases	**Valid**	48	100.0
	Excluded[a]	0	.0
	Total	48	100.0

Table 3
Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.525	.528	23

3.2. Analysis of the Survey Results and the Proposed Solution

We first analyzed the survey answers to identify the problems encountered faced by the students. After that, we used the survey results to propose a new approach to meet the research objective. Fig. 1 shows gathered statistics in some questions posed within the Survey.

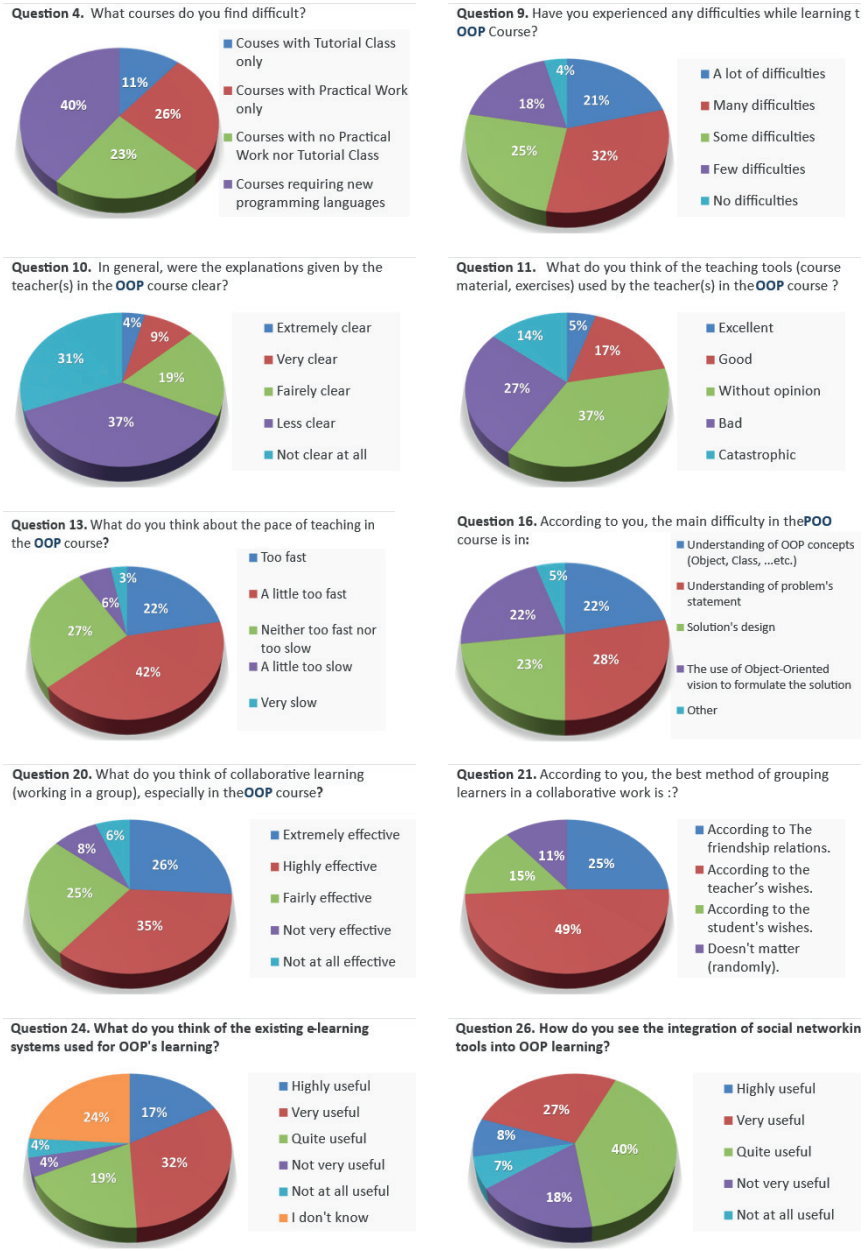


Figure 1. *Samples of the statistics gathered from the Survey's Answers*

By analyzing the obtained answers, we got the following facts:

- Most students have difficulties mastering courses that require new paradigms and programming languages.
- In particular, most students have difficulties studying the Object-Oriented Programming (OOP) paradigm.
- 68% of the students did not understand the explanations given by the teachers in the OOP course.
- 41% of the students believe there is a problem selecting teaching tools for the OOP course.
- Regarding the pace of teaching in the OOP course, 64% of the students confirm it is faster than their own pace.
- As for the main problem in learning OOP, 51% of the students see that the problem is not learning the basic concepts of OOP but understanding statements of the problems posed in this module and arriving at an object-oriented design for these problems.
- As for using a collaborative solution (teamwork), most students (61%) think it can be effective in the OOP module. In addition, 49% saw that teacher-determined grouping is the proper method of grouping learners.
- Concerning the existing OOP's e-learning platforms' effectiveness, 49% of the students think they are helpful.

3.3. Proposition of LISS'nAP Approach (Learn, Imaging, Solve, Share, and APprove)

The proposed approach uses Object-Oriented Analysis and Design (OOAD) to teach OOP instead of a specific object-oriented implementation built on a particular programming language. The proposed LISS'nAP approach is a collaborative approach that relies on peer advising. However, before we go further into explaining the approach, we have to define the course structure and the grouping method.

3.3.1. Course Pedagogical Structure

The course to be studied should have two main stages:

1. The Learning Stage: the basic notions of the course are taught. The course is organized in the form of Chapters made up of a set of Concepts.

2. The practice and application stage: the acquired knowledge during the first stage is practiced, and Final Projects are implemented (Fig. 2).

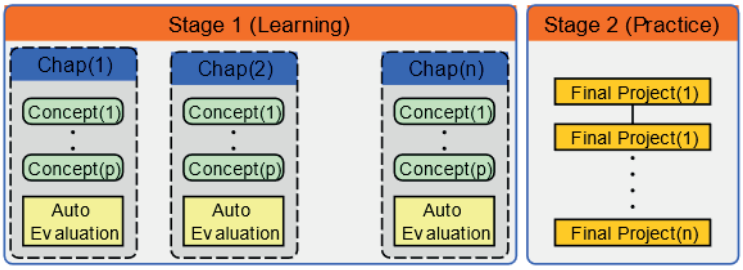


Figure 2. *Pedagogical Structure of the Course*

3.3.2. Group Formation method

When enrolling in the course, the learners have to answer a survey. The Survey is used to assess the learners' level of knowledge. Learners are then classified into five categories, "Excellent," "Good," "Average," "Fair," and "Poor." Learners' groups are formed of a maximum of 4 learners, and each group has to contain at least three different categories of learners. The grouping process takes place after the end of the two first stages.

3.3.3. LISS'nAP Sequence

Fig. 3 shows the two stages of LISS'nAP. Each of the four phases of the two stages is also shown in the figure. The two first phases: Learn and Imagine, are individual and do not require collaboration among learners. These two phases happen during

the Learning stage of the course. The remaining phases: "Solve, Share 'n Approve," and Promote steps, are collaborative and occur in the second stage of practice and application, where the learners are perceived as a group.

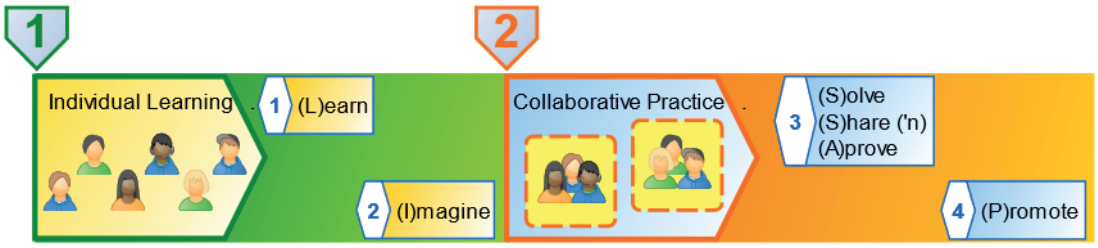


Figure 3. *LISS'nAP Sequence*

LISS'nAP is composed of a total of 6 phases that are the following (Fig. 4):

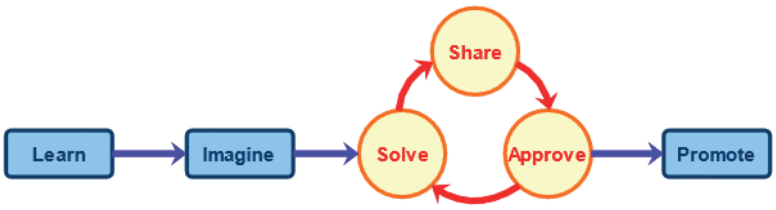


Figure 4. *LISS'nAP approach diagram with the Three Steps Cycle (Solve, Share 'n, Approve)*

1.a. (L)earn:

This first phase is individual. It is mandatory for the majority of educational courses. It is based on teaching the discipline's basic concepts, laws, and rules. It is part of the course's first pedagogical stage (Individual Learning Stage) to be studied, in our case, the OOP. The pedagogical content is organized into chapters and concepts (see Fig. 2). The learner begins to learn the basic concepts of OOP, such as classes, objects, and inheritance, using written content and grouped in well-prepared chapters maintained by the teacher. The latter, who plays the role of the course designer, must meet the following criteria: reliability and simplicity. He must also maintain a rate of progress and a rhythm of difficulty appropriate to the learning abilities of the learners. A self-assessment test quiz is handed out to the learners to evaluate their acquired knowledge at the end of each chapter.

1.b. (I)Imagine:

This phase is also individual and does not require any form of collaboration where each learner practices small tutorials with solutions to learn how to proceed with Analysis and Design step by step with minor solvable problems presenting a straightforward application of one of the concepts studied in the previous phase. Each learner is asked to use their imagination to extract the objects that constitute the solution to a given problem, determine the relationships between them, and then represent them in Unified Modeling Language (UML) diagrams.

Fig. 5 presents the four steps the learner follows to master the imagination.

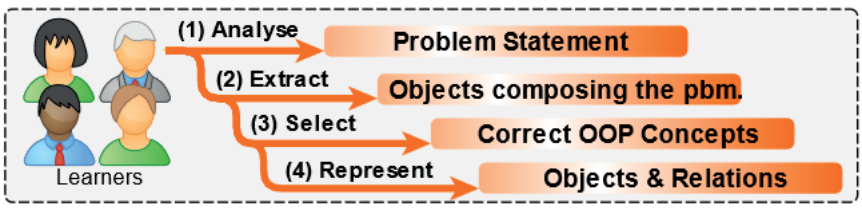


Figure 5. *Imagine Steps*

In this phase, the teacher prepares a set of solved problems using problem statements alongside their illustrative object-oriented designs detailing all the solution steps. The teacher explains how to extract the hidden objects in a problem statement, imagine their relationships, and correctly choose the right OOP concepts to present them.

Example:

The student can imagine the concepts and the relations among them from the following description: "*In a department, a teacher is responsible for a set of courses. The latter is composed of a collection of concepts and exercises. A group of students follows these courses".*

3. (S)olve, (S)hare (and) (A)pprove:

This phase is the first collaborative phase of our approach. This phase runs in a cycle for each learner, who is continually repeating the three-cycle peer advising steps: Solve, Share, and Approve, until he has the approval of his peers or he reaches the deadline of this phase. These three steps are as follows:

3.a. (S)olve:

In this phase, each group member tries to apply what they have learned in phase ***(I)magine*** by individually constructing object-oriented designs for the problem posed as a final project in progress. A real-time graphic design tool is made available to the learners, allowing them to understand better and simultaneously explain their designs to their peers in the next phase.

3.b. (S)hare:

After completing their designs, each learner shares his design with the other group members and awaits his peers' advice or approval. All the other group members have to visualize his solution and either approve his solution or advise him how to correct his design if this member judges the solution as erroneous. A third option is also available when the peer group member cannot give a solution so that he can abstain from voting. In that case, this member's vote is not accounted for, but he can always see the solution of his peer and can have an idea of the solution.

3.c. (A)pprove:

Each learner must advise other group members on their proposed solutions by approving or rejecting them while specifying the reasons behind the rejection. The solution is approved if the number of "approvals" exceeds the number of "rejections." Otherwise, the learner must repeat the cycle and continue improving his solution until getting "approved" or the exercise's time is over.

4. (P)romote :

After each group member finishes their specific design, it is time to choose the best design, promoted as the group's unique solution. Each member has to evaluate his co-members' solutions by ranking them using a star rating system. The voting process may follow these rules :

- The voting process begins after the solution is flagged as completed by the learner who owns the solution.
- The learners cannot rate their solutions.
- The voting operation must be completed before the deadline for the final project submission.
- If the project submission deadline is reached, and all the group members have not voted on any solution, the best solution is determined by a random draw between all the solutions marked as completed.
- The range of the score is from 1 to 5.
- The final solution score equals the sum of the voters' scores divided by the number of voters (the final score equals the average score).

- The solution with the best final (average) score is considered "the group's solution."

- In the case of equality between the average scores of two or more solutions, the best solution is determined by a draw.

Finally, the group solution will be sent to the lead teacher, who will assign a score regarded as "the group's score."

3.3.4. Collaborative and Individual Assessment

In addition to the collaborative assessment, which consists of a unique score for all the group members, another "individual score" is computed separately for each member, representing the learner's engagement with his peers. It is measured using the frequency and the quality of his involvement in developing the collaborative solution. We must mention here that individual solutions, even though approved by the other members, may not have any value in individual and collaborative assessment. Doing so will discourage plagiarism or information hiding and promote the collaborative spirit against the "self-preservation" tendency.

3.4. Results of experimenting with the prototype:

We conducted a pilot experiment to validate and fine-tune the proposed approach. A set of four students at the Computer science department of Guelma University were evaluated before and after using the developed prototype using pre and post-tests (The experiment was conducted remotely during the first days of the Covid-19 containment period in Algeria).

The students had to take a pretest exam before using the developed prototype and another post-test exam after using the prototype. Table 4 presents the obtained results.

Table 4
Pilot Experiment results

	Student 1	Student 2	Student 3	Student 4
Pretest scores (/20)	**8**	**11**	**6**	**10**
Post-test scores (/20)	**12**	**15**	**9**	**13**

As presented in Table. 4, the post-test scores of the participants were better than their pretest scores, supporting that using the proposed approach helped them learn OOP concepts better, which was confirmed by the participants, who expressed their satisfaction with using this system.

4. DISCUSSION:

The main contribution of this study concerns the proposition of a new approach for enhancing the learning of Object-Oriented Programming (OPP) concepts. An online learning prototype adopted this approach. The pilot experiment with some students helped us enhance the system. Unfortunately, the proposed system was only made available to a limited number of students while they followed the OOP course online (during the Coronavirus confinement period, May-June 2020).

During this reduced experiment, we encountered many difficulties and faced many challenges with the students as they were challenging to understand and unpredictable. Some refused to work with others and declined any request for collaboration, while others who were too solicited felt exploited by their teammates, leading them to decrease their contribution or even stop collaborating. Also, the students' grouping process proved tricky and presented a big problem because there were always cases where some students were highly requested while others were not.

The distribution of the tasks among the group members was also not very convenient and was often not even and not fair. Therefore, we sometimes found that the brilliant members did most of the work on behalf of the group while the weakest did not provide the required effort, which is not the intention of collaborative learning.

5. CONCLUSION AND FUTURE WORK

Our empirical study in the form of an online survey shed some light on some of the existing learning difficulties learners face when studying the Object-Oriented Programming paradigm among university students. Furthermore, this empirical

study helped us design and implement a new method composed of two stages called LISS'nAP. The first one is individual, while the second is collaborative; each one includes a set of phases that enable the teaching of the basic concepts of Object-Oriented Programming and helps the learners practice what they have learned by encouraging them to solve real problems.

In the same perspective, and to implement our LISS'nAP method, we designed a system adopting this new method based on one of the Unified Modeling Language (UML) diagrams. The system in question is in the form of an e-learning system called LISS'nAP-POO. This platform provides learners with all the flexible and easy-to-use tools necessary to carry out the LISS'nAP method steps.

To check the effectiveness of the proposed method (LISS'nAP), we have scheduled a full-scale experimentation phase of the LISS'nAP-POO system with computer science students from a higher education institution. Furthermore, in future work, we propose to use artificial intelligence techniques for group formation. Using these techniques, we can get homogeneous groups meeting learners' expectations.

Peer-review: Externally peer-reviewed.
Author Contributions: Conception/Design of Study- R.B., A.B., Y.L.; Data Acquisition- R.B., A.B., Y.L.; Data Analysis/Interpretation- R.B., A.B., Y.L.; Drafting Manuscript- R.B., A.B., Y.L.; Critical Revision of Manuscript- R.B., A.B., Y.L.; Final Approval and Accountability- R.B., A.B., Y.L.; Material and Technical Support- R.B., A.B., Y.L.; Supervision- R.B., A.B., Y.L.
Conflict of Interest: The authors have no conflict of interest to declare.
Grant Support: The authors declared that this study has received no financial support.

REFERENCES

Assinder, W. (1991). Peer teaching, peer learning: One model. *ELT Journal*, *45*(3), 218–229. https://doi.org/10.1093/elt/45.3.218

Barman, C., & Benson, P. (1981). Peer Advising: A Working Model. *NACADA Journal*, *17*, 46–51. https://eric.ed.gov/?id=EJ253991

Benabbou, F., & Hanoune, M. (2006). EasyAlgo : Un environnement d'apprentissage et d'autoévaluation de l'algorithmique. In *admiroutes.asso.fr*. http://admiroutes.asso.fr/larevue/2006/76/Article_Apprentissage.pdf

Bremgartner, V., & de Magalhaes Netto, J. F. (2012). Improving collaborative learning by personalization in Virtual Learning Environments using agents and competency-based ontology. *2012 Frontiers in Education Conference Proceedings*, 1–6. https://doi.org/10.1109/FIE.2012.6462345

Cheung, R. (2006). A web-based learning environment for object-oriented programming. *International Journal of Information and Operations Management Education*, *1*(2), 140. https://doi.org/10.1504/ijiome.2006.009172

Cypher, A., & Halbert, D. C. (1993). What What I Do: Programming by Demonstration. *Cambridge, ….* http://regexp.bjoern.org/archives/000121.html

Dillenbourg, P. (1999). What do you mean by collaborative learning? chapter 1. *Collaborative-Learning : Cognitive and Computational Approaches*.

Djelil, F., Albouy-Kißi, B., Albouy-Kißi, A., Sanchez, E., & Lavest, J. M. (2015). Towards a 3D virtual game for learning object-oriented programming fundamentals and C++ language: Theoretical considerations and empirical results. *CSEDU 2015 - 7th International Conference on Computer Supported Education, Proceedings*, *2*, 289–294. https://doi.org/10.5220/0005475802890294

Donato, R. (1994). Collective Scaffolding in Second Language Learning. In P. lantolf James & A. Gabriela (Eds.), *Vygotskian Approaches to Second Language Research*. Ablex Publishing Corporation. https://books.google.com/books?hl=en&lr=&id=QbB-CGkx4hwC&oi=fnd&pg=PA33&dq=donato+1994+scaffolding&ots=FTssNtoMiN&sig=VZxAqXOq0jpPj8gcKWq_cSBRzBw

Ellis, R. (1994). *The study of second language acquisition*. Oxford University Press.

Esteves, M., & Mendes, A. J. (2004). A simulation tool to help learning of object oriented programming basics. *Proceedings - Frontiers in Education Conference, FIE*, *2*. https://doi.org/10.1109/fie.2004.1408649

Guibert, N., Girard, P., & Guittet, L. (2006). Performances et usages d'un environnement d'apprentisage de la programmation «basé sur exemple». *ErgoIA2006*. https://www.researchgate.net/publication/241141580

Hank, S., & Chikh, A. (2013). Towards a collaborative learning scenario adapted to the multiple learner's intelligence. *2013 3rd International Symposium ISKO-Maghreb*. https://doi.org/10.1109/ISKO-Maghreb.2013.6728135

Henri, F., & Lundgren-Cayrol, K. (1998). Apprentissage collaboratif et nouvelles technologies. In *Bureau des technologies d'apprentissage*.

Henri, F., & Lundgren-Cayrol, K. (2001). *Apprentissage collaboratif à distance*. PUQ.

Hu, L., & Xu, H. (2013). Collaborative M-learning based on multi-agent system. *Proceedings of 2013 IEEE International Conference on Service Operations and Logistics, and Informatics*, 59–62. https://doi.org/10.1109/SOLI.2013.6611382

Koh, G. C. H., Khoo, H. E., Wong, M. L., & Koh, D. (2008). The effects of problem-based learning during medical school on physician competency: A systematic review. *CMAJ*, *178*(1), 34–41. https://doi.org/10.1503/cmaj.070565

Kölling, M. (1999a). Teaching object orientation with the blue environment. *JOOP - Journal of Object-Oriented Programming*, *12*(2), 14–23. https://kar.kent.ac.uk/21734/

Kölling, M. (1999b). The problem of teaching object-oriented programming. *Engineering*, *11*(9), 6–12. https://kar.kent.ac.uk/21879/2/the_problem_of_teaching_object-oriented_kolling_1.pdf

Laal, M., & Ghodsi, S. M. (2012). Benefits of collaborative learning. *Procedia - Social and Behavioral Sciences*, *31*, 486–490. https://doi.org/10.1016/j.sbspro.2011.12.091

Lafifi, Y., Gouasmi, N., Halimi, K., Herkas, W., Salhi, N., & Ghodbani, A. (2010). Trace-based collaborative learning system. *Journal of Computing and Information Technology*, *18*(3), 207–219. https://doi.org/10.2498/cit.1001410

Mynard, J., & Carson, L. (2014). *Advising in language learning: Dialogue, tools and context*. Routledge. https://doi.org/10.4324/9781315833040

Robins, A., Rountree, J., & Rountree, N. (2003). Learning and teaching programming: A review and discussion. *International Journal of Phytoremediation*, *21*(1), 137–172. https://doi.org/10.1076/csed.13.2.137.14200

Rogalski, J. (1998). Concepts et méthodes d'analyse des processus de coopération dans la gestion collective d'environnements dynamiques. *Communications Interactives Dans Les Groupes de Travail*, 27–58.

Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT Journal*, *59*(1), 23–30. https://doi.org/10.1093/elt/cci003

Rose, I. R. (2003). *Visual modelling with Rational Rose*.

Seng, W. Y., Hayati, M., Yatim, M., & Hoe, T. W. (2018). Learning object-oriented programming paradigm via game-based learning game – pilot study. *The International Journal of Multimedia & Its Applications (IJMA)*, *10*(6). https://doi.org/10.5121/ijma.2018.10615

Sims, Z., & Bubinski, R. (2011). *Codecademy.(2011)*.

Stroustrup, B. (1994). *The design and evolution of C++*. Addison-Wesley.

Swain, Merill. (1995). Three functions of output in second language learning. In *Principle and practice in applied linguistics: Studies in honor of H.G. Widdowson* (pp. 125–144). Oxford University Press. https://ci.nii.ac.jp/naid/10025500767

Swain, Merrill, & Lapkin, S. (2000). Task-based second language learning: The uses of the first language. *Language Teaching Research*, *4*(3), 251–274. https://doi.org/10.1177/136216880000400304

Tadjer, H., Lafifi, Y., Seridi-Bouchelaghem, H., & Gülseçen, S. (2020). Improving soft skills based on students' traces in problem-based learning environments. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2020.1753215

Troussas, C., Virvou, M., & Alepis, E. (2014). Collaborative learning: Group interaction in an intelligent mobile-assisted multiple language learning system. *Informatics in Education*, *13*(2), 279–292. https://doi.org/10.15388/infedu.2014.08

Villamil, O. S., & De Guerrero, M. C. M. (1996). Peer Revision in the L2 Classroom: Social-Cognitive Activities, Mediating Strategies, and Aspects of Social Behavior. *Journal of Second Language Writing*, *5*(1), 51–75. https://doi.org/10.1016/S1060-3743(96)90015-6

Walldén, S., & Mäkinen, E. (2014). Educational data mining and problem-based learning. *Informatics in Education*, *13*(1), 141–156. http://www.scopus.com/inward/record.url?eid=2-s2.0-84903887180&partnerID=40&md5=81c98948c4908d535123ec2d7139cc0d

Wood, D. F. (2008, May 3). Problem based learning. *BMJ*, *336*(7651), 971. https://doi.org/10.1136/bmj.39546.716053.80

Yan, L. (2009). Teaching object-oriented programming with games. *2009 Sixth International Conference on Information Technology: New Generations*, 969–974.

ACTA INFOLOGICA 2023;7(1):209-228
dergipark.org.tr/acin

DOI: 10.26650/acin.1198925

RESEARCH ARTICLE

A Modular Efficiency Determination Formula for Information Retrieval Evaluations and Optimizations

Bilgi Erişim Değerlendirmeleri ve Optimizasyonları İçin Modüler Bir Verimlilik Belirleme Formülü

Veli Özcan Budak[1]

ABSTRACT

The notion of efficiency has typically been associated with the efficiency of systems rather than users in Information Retrieval (IR) literature. In the usability literature, on the other hand, this notion is defined from a user-based perspective, corresponding to how long a user accomplishes a task. Despite this, the common aim for both has to do with the efficient use of time. This study examines the efficiency notion in the IR literature from a user-based efficiency window in the usability literature. In the present study, a modular efficiency determination formula (MEDEF) to create different efficiency indicators by focusing on IR system evaluations and optimizations from the usability perspective is proposed. The MEDEF can be thought of as an efficiency indicator creator based on both effectiveness metrics and efficiency indicators already used in IR studies. In the scope of this study, eight MEDEF-based efficiency indicators were created and compared to several baseline efficiency indicators already used in IR studies. While the study's first aim is to reveal how consistent the MEDEF-based indicators are and whether these indicators are more successful/reliable than the baselines, the second is to set an example of the usage of efficiency indicators in evaluations of IR systems from a usability perspective. General findings from interactive user behaviour for one month show that the MEDEF-based indicators outperform the baseline indicators and further strengthen the reflections in the baseline indicators. Several usage scenarios regarding the potential of the MEDEF are also shared and discussed in the scope of the study.

Keywords: Human-Computer Interaction, Efficiency Indicator, Session Abandonment, Interactive Information Retrieval

ÖZ

Verimlilik kavramı, Bilgi Erişim (BE) literatüründe temel olarak kullanıcılardan ziyade sistemlerin verimliliği ile ilişkilendirilmiştir. Kullanılabilirlik literatüründe ise bu kavram, bir kullanıcının bir görevi ne kadar sürede tamamladığına karşılık gelen kullanıcı tabanlı bir bakış açısıyla tanımlanır. Yine de, ortak amaç her iki literatür için de zamanı verimli kullanmaktır. Bu çalışma, BE literatüründeki etkinlik kavramını, kullanılabilirlik literatüründeki kullanıcı tabanlı etkinlik penceresinden incelemektedir. Bu çalışmada, kullanılabilirlik perspektifinden BE sistem değerlendirmelerine ve optimizasyonlarına odaklanarak farklı verimlilik göstergeleri oluşturmak için modüler bir verimlilik belirleme formülü (MEDEF) önerilmiştir. MEDEF, BE çalışmalarında hâlihazırda kullanılan etkililik metriklerine ve verimlilik göstergelerine dayalı bir verimlilik göstergesi üreticisi şeklinde düşünülebilir. Bu çalışma kapsamında, sekiz MEDEF tabanlı verimlilik göstergesi oluşturulmuş ve hâlihazırda BE çalışmalarında kullanılan birkaç temel verimlilik göstergesiyle karşılaştırılmıştır. Çalışmanın ilk amacı, MEDEF temelli göstergelerin ne kadar tutarlı olduğunu ve bu göstergelerin mevcut temel göstergelere göre daha başarılı/güvenilir olup olmadığını ortaya koymak iken, ikincisi, kullanılabilirlik açısından BE sistemlerinin değerlendirmelerinde verimlilik göstergelerinin kullanımına bir örnek oluşturmaktır. Bir aylık etkileşimli kullanıcı davranışlarından elde edilen genel bulgular, MEDEF tabanlı göstergelerin temel göstergelerden daha iyi performans gösterdiğini ve temel göstergelerdeki yansımaları daha da güçlendirdiğini göstermiştir. MEDEF'in potansiyeline ilişkin çeşitli kullanım senaryoları da çalışma kapsamında paylaşılmakta ve tartışılmaktadır.

Anahtar Kelimeler: İnsan-Bilgisayar Etkileşimi, Verimlilik Göstergesi, Oturum Terk Etme, Etkileşimli Bilgi Erişim

[1](Assist. Prof.), Bandırma Onyedi Eylül University, Gonen Vocational School, Balikesir, Turkiye

ORCID: V.Ö.B. 0000-0002-0960-0542

Corresponding author:
Veli Özcan BUDAK
Bandırma Onyedi Eylül University, Gonen Vocational School, Balikesir, Turkiye
E-mail address: veliozcanbudak@gmail.com

Submitted: 03.11.2022
Revision Requested: 25.05.2023
Last Revision Received: 26.05.2023
Accepted: 26.05.2023
Published Online: 19.06.2023

Citation: Budak, V.Ö. (2023). A modular efficiency determination formula for information retrieval evaluations and optimizations. *Acta Infologica, 7*(1), 209-228. https://doi.org/10.26650/acin.1198925

1. INTRODUCTION

Evaluations in Information Retrieval (IR) studies are made mostly by following the Cranfield paradigm and focusing on the effectiveness of systems/models. Two elements underlie the basis of these kinds of evaluations. While reference collections with predetermined information needs correspond to the first element, several evaluation metrics constitute the other. This evaluation method has become indispensable for IR studies and even for those defined as interactive. On the other hand, the efficiency notion can be shown as the other type of evaluation. This notion is mainly associated with the efficiency of IR systems rather than that of users. Two performance elements, throughput and latency, constitute the center of efficiency investigations (Croft, Metzler, and Strohman, 2009; Büttcher, Clarke, and Cormack, 2010), and the main focus is on how quickly a system responds and the sufficiency of system resources (Zhai and Massung, 2016). In other words, efficiency is predominantly associated with the subject of "time." Regardless of the evaluation type, both evaluations intrinsically consist of system-based investigations rather than user-based ones.

When focusing on another research area, usability, it can be seen that the notion of "efficiency" also exists in this area but with a different definition. This notion is defined from a user-based perspective, corresponding to how long it takes for a user to accomplish a task (Nielsen, 1993; Frøkjær, Hertzum, and Hornbæk, 2000; Hornbæk, 2006; Rubin and Chisnell, 2008; Rosenzweig, 2015). Nevertheless, the main subject does not change when compared to the IR perspective: "time." In other words, both research areas focus on the efficient use of time. This study examines the concept of efficiency in the IR literature from a user-based efficiency window in the usability literature, meaning that the focus is on user efficiency rather than on system efficiency.

Logs of search interactions have been intensively utilized for different purposes in IR studies because they consist of natural user behaviors. While Jiang, Leung, Yang, and Ng (2015) and Vidinli and Ozcan (2016) utilized logs for query suggestion purposes, Joachims (2002) and Agichtein, Brill, and Dumais (2006) used them for search result improvement. Liu, Liu, and Belkin (2014) and Kim, Hassan, White, and Zitouni (2014), on the other hand, used logs for the exploration of user behavior patterns. Users who try to satisfy their information needs leave several types of traces (implicit data) in the background while searching. These traces also hold some valuable indicators regarding the time notion. "Dwell time," "time to first click," and "time to last click" can be given as examples of these types of indicators. These indicators have been considered with different definitions/namings in IR studies, such as "user engagement" by Singla and White (2010), "search satisfaction" by Hassan (2012), Kim, Hassan, White, and Zitouni (2014), and Liu et al. (2015), "choice overload" by Beierle, Aizawa, Collins, and Beel (2020), "implicit measures of user interest" by Fox et al. (2005), and "interest detectors" by Claypool, Brown, Le, and Waseda (2001). When considering them from the usability perspective, these indicators carry signs regarding user efficiency even though the namings are different. On the other hand, even if we put this differentiation aside, it is evident that these indicators have not been utilized individually in evaluating IR systems, as explained in Section 2. At this point, two questions appear that need to be answered: (1) How reliable or useful are these indicators when they are not used for the purpose of evaluation? (2) Is it possible to evaluate IR systems only through these indicators? While the first question regards the usage of these indicators already in IR studies, the second regards whether these indicators can be used to evaluate IR systems from the usability perspective. This study seeks answers to these questions and, in addition to the mentioned efficiency indicators, proposes a modular efficiency determination formula (MEDEF) to create different types of user-based efficiency indicators.

In the scope of the present study, users' search interactions were first recorded through the search modules of different department websites of a university in Turkey. Secondly, these websites (23 in total) were separated into six groups based on users' session abandonment behaviors, and fifteen propositions were created regarding the IR performances of these groups. Lastly, eight efficiency indicators created by following the MEDEF and four baseline indicators already used in IR studies were compared statistically to investigate how they successfully reflect group performances based on the created propositions. Apart from the statistical comparisons, several examinations were also made with Machine Learning (ML)-based approaches. The study's first aim is to reveal how consistent the MEDEF-based indicators are and whether these indicators are more successful/reliable than the baselines. The second is to set an example of the usage of efficiency indicators in evaluations of IR systems from a usability perspective.

The rest of the paper is organized as follows. Studies made with efficiency indicators are shared in Section 2. Section 3 explains the MEDEF's features and the method followed in the present study. The study's findings are shared in Section 4. While the results are discussed in Section 5, the contributions of the present study and several usage scenarios of the MEDEF for future work are shared in Section 6.

2. RELATED WORK

In this section studies which have benefited from time-based efficiency indicators are shared.

Hassan, Song, and He (2011) presented a prediction model based on the ML concept to determine user satisfaction in the information-seeking process. While creating the model, the authors benefited from the dwell time indicator, and the model's success in estimating user satisfaction and improving search results was revealed. Lee, Teevan, and de la Chica (2014) conducted a study to characterize user search behavior by utilizing the time to first click and dwell time indicators. The authors made several suggestions for search result improvement in the study results. Based on the ML concept, Arguello (2014) created a model by using the dwell time indicator and several implicit data types to predict search task difficulty. In the study's result, several suggestions regarding utilizing different types of data/indicators in creating a predictive model were made by the author. Similarly, Liu, Liu, and Belkin (2014) also investigated users' behavioral differences regarding search task difficulties in the information-seeking process. The authors created a prediction model utilizing several types of dwell time indicators to determine search task difficulties. In their study, Kim, Hassan, White, and Zitouni (2014) created a model based on the ML concept using the dwell time indicator and tried to predict search satisfaction. In another study using the dwell time indicator an attempt was made to categorize search sessions using the ML concept to show whether they reflected struggling or exploring (Hassan, White, Dumais, and Wang, 2014). Balakrishnan and Zhang (2014) proposed a model benefiting from the dwell time indicator to improve search results. The authors stated that the proposed model revealed significant improvements. A model using the time-to-first click, dwell time, and several different indicators was created by Arkhipova, Grauer, Kuralenok, and Serdyukov (2015) to predict unsuccessful search sessions. The authors investigated users' search engine switching behaviors in the study in which an evaluation metric was also presented for A/B testing studies.

In a study by Borisov, Markov, de Rijke, and Serdyukov (2016), the authors tried to explore behavioral patterns on times between user actions using the time to first and last click and dwell time indicators and proposed a prediction model to be used for personalized IR improvements. The study results indicated that successful findings were reached. Alhabashneh, Iqbal, Doctor, and James (2017) proposed a fuzzy-based mechanism using the dwell time indicator and different implicit data types to estimate document relevance. The authors indicated that the proposed mechanism showed successful prediction performance regarding document relevance. In another study focused on system-based efficiency, a framework proposal was made by Makkar and Kumar (2018) to improve search results by utilizing the dwell time indicator. The authors achieved successful results with the proposed framework. Another study focusing on difficulties in making decisions while searching research articles was made by Beierle, Aizawa, Collins, and Beel (2020). In the study, in which one of the indicators was the time to first click indicator, the authors revealed findings regarding users' behavior patterns. Sarkar, Mitsui, Liu, and Shah (2020) conducted a study focusing on determining users who needed help finding information. The authors created a successful prediction model with the logistic regression classifier, utilizing the dwell time indicator.

When we consider the studies above, it can be seen that two main types of investigation exist: revealing users' behavioral patterns or using efficiency indicators for optimization purposes. As evaluations were made following the Cranfield paradigm regardless of the study types, these studies neither focus on how reliable the efficiency indicators are among each other nor consist of evaluations from users' perspectives. The present study sets an example for evaluating IR systems from a user-based approach. Moreover, it sheds light on the trustworthiness of efficiency indicators and how these indicators can be utilized more successfully through the MEDEF for both IR system evaluations and optimizations.

3. METHOD

This section first explains the components of the data collection process. Secondly, the efficiency indicators that have already been used in IR studies, and the efficiency indicators created by following the proposed formula in the present study, are

explained by being combined with the data preprocessing step. After this, the research questions and how the evaluations were carried out are shared. Lastly, the section concludes by explaining the present study's limitations.

3.1. Data Collection Process

User search-visit interactions, gathered over the period of one month from the textual search modules of different department websites of a university in Turkey , underlie the center of this study. The search modules serve with the boolean model (the same infrastructure) and work integrated with the websites. As each website addresses different information needs, the contents announced differ, meaning each search module can be considered a tool that helps users satisfy different information needs. The interaction data were collected from 66 departments. A relational database with two tables (*queries* and *visits*) was used to record the interactions with these modules during the data collection. The structure of the tables is given in Table 1.

Table 1
The used tables in recording search interactions

	The *queries* table	**The *visits* table**
Attributes	id	id
	department_id	query_id
	query	page_id
	query_length	rank
	dwell_on_SERP*	visit_time
	query_time	
	anonymized_ip	

* The dwell time on the Search Engine Result Page (SERP). This value was recorded in the form of seconds.

As seen in Table 1, both tables were created with a simple design to collect basis interaction data. A record was added to the *queries* table whenever a user searched with a query for information in the data collection process. After users were shown relevant pages and their snippets found by modules, the users had two options: clicking a suggestion on the SERP or abandoning the session/query. While a record regarding the visit behavior was added to the *visits* table for users who preferred the first option, no record was added to the table for the other option, and session states were regarded as being abandoned.

Two different values, query length (QL) and dwell time on the SERP (DwSERP), were also recorded in addition to query time (QT) in each search session (Table 1; the *queries* table) to be utilized in the created indicators. While QL is how many unique words a query term has, DwSERP corresponds to how much time a user spends on the SERP in a search session. DwSERP was recorded at 0.1-second intervals after starting a search session by making a query (a Javascript code was utilized for this process). For the *visits* table, the main values utilized in the indicators are rank and visit time (VT). While the rank value corresponds to what order a clicked web page (document) suggestion is on the SERP, VT indicates the first contact timestamp between users and the clicked documents.

Throughout the data collection process, while the total number of search interactions for all websites was detected as 29,545, the total number of unique queries was observed as 3,402. How these collected data were preprocessed is explained in the next section.

3.2. Data Preprocessing and Efficiency Indicators

In the preprocessing phase, each query was first considered as a session of 30 minutes. Let us assume that a user made a query and started a session. If that user sent another query with the same terms before the opened session ended, that user's visit interactions from the second query were treated as if the user had not sent the second query, meaning that only visit interactions of the user continued to be recorded under the opened session instead of a new session being opened. A new session would be started if that user had sent the second query with different terms, even though the intent was identical. On the other hand, during preprocessing, which session belongs to which user was determined according to users' anonymized_IP

values (Table 1), and a session in which no visit interaction was observed was considered abandoned. While tracking users' visit interactions on the same session, if users clicked the same suggestion for the second/third/fourth time, only the first visit was taken into account (this action was excluded for only one indicator, as explained in Section 3.2.2). In the direction of these procedures, the preprocessing of collected data was carried out in three steps incrementally. These steps are explained below:

Step 1: Users who had made a query in the past six months were excluded from the collected data based on their anonymized IP addresses. The reason for this is to avoid fallacious data that might arise from users and that could possibly have prior information gathered from a past search interaction with the search modules regarding their continuing information needs. In other words, the situation of a user searching with a query with search modules in the past indicates that that user has an experience from the past regarding responses from the query sent in the past. In this case, if the user prefers to use the same query terms for the second time and departments shared no content regarding the user intent until this second query, the used search module will probably suggest the same documents as in the first query. In this direction, the user's decision time will quite likely be very short because of past experience. As this case can misguide the analysis, users from the past six months were excluded, meaning that users who belonged to the month of the data collection process were considered users who were interacting with the system for the first time.

The second action in this step was to exclude queries with no result from the collected data.

Step 2: Except for the search interactions whose DwSERP value was between one and 60 seconds, all interactions were treated as outliers. This action was taken in order to focus only on search interactions made with a high concentration. For the same purpose, the second action was to eliminate the interactions where the first click time (time to first click - TTFC) was higher than 30 seconds in sessions that were not abandoned. The last process in this step was to exclude websites with search sessions of less than 100.

Step 3: The number of cases of session abandonment pertaining to each website was examined after the second step. Websites in which all sessions had been abandoned were excluded from the dataset. Lastly, interactions made by the same users in different departments were removed based on users' anonymized IP addresses in order that statistical analyses be properly carried out. In other words, if a user had searched on both A department and B, only one of these interactions was kept, the other being removed. Thus, interactions in all departments were isolated from each other.

The collected data were preprocessed based on the above steps. The search interaction data, which totaled 29,545 from 66 department websites before preprocessing, were reduced to 11,228 from 23 department websites with 1691 unique query terms. Three different indicator classes were created. While the first two classes correspond to the indicators that reflect natural user behavior and have already been used in IR studies, the third consists of the indicators, which again reflect natural user behavior, created for this study. In contrast to studies that mostly focused on IR system performances/optimizations, in this study, all indicators are considered from the usability perspective, and natural user search behaviors constitute the center of investigations. From this point, the first indicator class is named "Guidance Indicator"; the second class is defined as "Baseline Efficiency Indicators"; the third class is named "MEDEF-Based Efficiency Indicators."

3.2.1. Guidance Indicator

Users' Session Abandonment (SAb) behaviors are a sign of unsuccessful sessions (Liu, Gwizdka, and Liu, 2010), and this type of user behavior can be utilized to evaluate IR systems (Diriye, White, Buscher, and Dumais, 2012). In this study, because the SAb behavior allows detection of how well IR systems serve users in satisfying their information needs from a generic perspective, it is selected as the guidance indicator. To this end, firstly, the IR performance of each website was characterized by the users' SAb behavior. Secondly, which website outperformed the other was determined by examining the percentage of sessions with no abandonment on each website. Afterward, to evaluate the reflectiveness performance of efficiency indicators, the websites that performed similarly to each other in satisfying user information needs were grouped based on the percentage of sessions with no abandonment. In this direction, six groups were created from 23 departments (Table 2).

Table 2
The groups of departments based on their performances

Name of the groups	The number of departments in the groups	Total interaction	Rule* (>X and <Y)	% of sessions with no abandonment
A	7	1296	>0 and <10	3,8
B	4	1317	>=10 and <20	15,1
C	5	7210	>=20 and <30	26,3
D	3	451	>=30 and <40	32,6
E	3	692	>=40 and <50	42,9
F	1	262	>=50	53,1

***** The percentage of sessions with no abandonment is higher than X and less than Y.

Table 2 indicates that group F is more successful than group E, that group E is more successful than group D, and so on. Fifteen propositions (6x5/2; 6 is the total number of the groups) were created to investigate the indicators' reflectiveness performance based on Table 2 (the statistical proof of the propositions is explained in Section 4). The statistical examinations were carried out by comparing the efficiency indicators between each other and focusing on how many propositions an indicator can reflect. Different ML-based examinations were also carried out using the created groups and the efficiency indicators together.

3.2.2. Baseline Efficiency Indicators

In the present study, four different efficiency indicators that have already been used in IR studies were chosen as baselines and explained below:

DwSERP

As described before, the length of time a user spends on the SERP in a search session corresponds to DwSERP. This indicator is inversely proportional to user satisfaction in the information-seeking process. High DwSERP values, which could arise from useless information sources in results or ambiguous query terms that users express, could point out negative experiences (Aula, Khan, and Guan, 2010; Sarkar, Mitsui, Liu, and Shah, 2020). Similarly, it has been stated that having difficulties in satisfying information needs causes more cognitive effort and longer DwSERP (Kuhar and Merčun, 2022). As DwSERP can reflect user satisfaction and, accordingly, how well IR systems serve users, it is chosen as the first baseline indicator. The last point that needs to be clarified for this indicator is that the DwSERP values were separated based on whether a session was abandoned or not in the analyses. The first indicator consists only of the DwSERP values of sessions with no abandonment.

DwSERP of abandoned sessions (DwSERP_Ab)

In addition to DwSERP, another baseline indicator was created from abandoned sessions' DwSERP values. As in DwSERP, we believe that DwSERP_Ab can also offer clues as to how close an IR system is to meeting information needs. In this direction, the tested assumption in the analyses was that the shorter DwSERP_Ab is, the more likely IR systems are close to satisfying information needs.

TTFC

TTFC is the length of time that elapses between the point at which a session starts upon a user sending a query and the point at which that user clicks on one of the pages on the SERP for the first time. In the study by Radlinski, Kurup, and Joachims (2008), it was found that TTFC was correlated with search success, meaning that the quality of results on the SERP decreases when TTFC increases. The faster users find relevant information sources the shorter TTFC is. From this perspective, it can be stated that as users do not have many difficulties when making the first decision, search interactions result in a positive experience. TTFC was chosen as the third indicator to investigate whether it can reflect user behavior in the propositions created through the SAb behaviors. While calculating this value, the QT value was subtracted from the first VT value observed for each session (Table 1).

Time to last click (TTLC)

TTLC is the time that elapses from the moment a session starts upon a user sending a query and the moment the user clicks one of the pages on the SERP for the last time in that active session. When users clicked the same suggestion on the SERP more than once, it was stated that only the first visit was considered except for one indicator. While the first click was recorded for the TTFC indicator, the others were utilized for the TTLC indicator regardless of being the same or different suggestions. The assumption for this indicator is the same as the TTFC indicator. In other words, shorter TTLCs reflect that users are able to meet their information needs in a short time, which indicates a more positive experience (Radlinski, Kurup, and Joachims, 2008).

3.2.3. MEDEF-Based Efficiency Indicators

This section is organized into two parts. While the formula used in creating efficiency indicators is explained in the first part, the second shares eight different indicators created by following the proposed formula.

The MEDEF

The proposed formula is used to give a value to each unique session that is not abandoned. It is calculated with the combination of three metrics: "Ambiguity Reward," "Punishment," and "Effectiveness," and can be thought of as an indicator that consists of both an efficiency and effectiveness metric together (Equation 1).

$$MEDEF = \underbrace{\frac{1}{QL}}_{\text{Ambiquity Reward}} \cdot \underbrace{\frac{1}{baseline\_indicator}}_{\text{Punishment}} \cdot effectiveness \qquad (1)$$

Ambiguity Reward: The Query Expansion (QE) technique has been intensively used in IR studies (Colace, De Santo, Greco, and Napoletano, 2015; Nie et al., 2016; Singh and Sharan, 2017; Nasir, Varlamis, and Ishfaq, 2019) to expand users' initial queries and thus alleviate the burden of finding relevant sources associated with user queries on these systems. Adding similar but differently expressed terms to initial terms to increase the possibility of finding relevant sources by IR systems underlies the QE technique. As QL supports IR systems (Belkin et al., 2003; Kelly and Fu, 2007), reciprocal QL is used as a reward in the MEDEF formula. In other words, when QL increases, the ambiguity decreases, and this increase helps IR systems understand search intents more easily. That is why the more words a query term has, the less reward it gets per session. This approach especially rewards sessions that start with ambiguous queries. In addition, it should be mentioned that no QE technique was used in the present study; rewards were given based on users' natural term selections.

Punishment: When we consider the baseline indicators in the previous section, it can be seen that all of them have a common logic, which is that the increase in an indicator points out dissatisfaction in terms of IR experience. To this end, the reciprocal value of efficiency indicators is utilized to penalize sessions.

Effectiveness: IR studies in the literature focus on whether IR systems/models can satisfy information needs successfully, and for this purpose, evaluations based on effectiveness metrics and reference collections mainly constitute the basis of these kinds of studies. Several metrics have been proposed/created from the past to the present. Gain-based metrics (CG, NCG, DCG, NDCG) by Järvelin and Kekäläinen (2000) and Järvelin and Kekäläinen (2002); Expected Reciprocal Rank (ERR) by Chapelle, Metlzer, Zhang, and Grinspan (2009); Rank-Biased Precision (RBP) by Moffat and Zobel (2008); Binary Preference (BPref) by Buckley and Voorhees (2004) and Mean Reciprocal Rank (MRR) by Voorhees (1999) can be shown as examples for effectiveness metrics. These and homologous metrics focus on determining the success of IR systems/models based on how many relevant/useful information sources are found using reference collections. However, two disadvantages exist in using effectiveness metrics for interactive IR studies. One is the nonexistence of reference collections in interactive environments, as users' information needs constantly change. The other is biased user behavior (Joachims, 2002; Joachims et al., 2005; Agichtein, Brill, and Dumais, 2006). Nonetheless, it was decided to integrate effectiveness metrics into the

formula because users' natural behaviour can take the place of explicit relevance judgments (Croft, Metzler, and Strohman, 2009; Zhai and Massung, 2016), and collecting these natural behavior data in large quantities is easy (Manning, Raghavan, and Schütze, 2008). Effectiveness metrics were not integrated into the formula to reward/penalize sessions. However, whether sessions are rewarded/penalized is left to user actions. Let us assume that the MRR metric is used, and the multiplication of the first two metrics in the formula is calculated as 0.3 for a session. If the user of that session clicks only on the second suggestion on the SERP, the MRR value will be calculated as 1/2 = 0.5, and the MEDEF value of the session will be 0.15. In the other scenario, if the user clicks two suggestions at the ranks of first and second, the MRR value will be calculated as (1/1 + 1/2) / 2 = 0.75, and the MEDEF value of the session will be 0.225. In short, user preference will decide whether sessions are rewarded/penalized.

In light of the above explanations, the last point that needs to be explained is that the increase in the MEDEF value indicates more positive user experiences.

The Created Indicators

Eight indicators with different variations were created using the combination of three baseline indicators and two effectiveness metrics (MRR and Average Precision - AP) to investigate how successful the MEDEF-based indicators perform in reflecting user behavior regarding the propositions created with the guidance indicator (Table 3).

Table 3
The created indicators based on the MEDEF

Description	Punishment	Effectiveness
MRR_DwSERP	$\frac{1}{DwSERP}$	$MRR = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{rank_i}$
MRR_TTFC	$\frac{1}{TTFC}$	
MRR_TTLC	$\frac{1}{TTLC}$	
MRR_ALL	$\frac{1}{DwSERP \cdot TTFC \cdot TTLC}$	
AP_DwSERP	$\frac{1}{DwSERP}$	$AP(L) = \frac{1}{\lvert Rel \rvert}\sum_{i=1}^{n} P(i)$
AP_TTFC	$\frac{1}{TTFC}$	
AP_TTLC	$\frac{1}{TTLC}$	
AP_ALL	$\frac{1}{DwSERP \cdot TTFC \cdot TTLC}$	

Because of the nature of interactive environments, it is impossible to determine which information sources are relevant to a query. Nevertheless, users' clickthrough behaviors can shed light on this ambiguity. Clicks can be utilized in determining whether a document is relevant or not. While using the MRR metric in the indicators, as exemplified before, the MRR value of a session can be calculated considering all rank values of visits of that session. The same process is also followed for calculating each session's AP value. We can define AP for a session with a ranked list L (on the SERP) where $|L|=n$; $P(i)$ donates the precision value of a clicked document at rank i in L; and *Rel* is all relevant documents in the collection. Even though *Rel* is ambiguous in interactive environments, the calculations can be carried out by accepting that *Rel* corresponds to all documents visited by a user in a session. If we exemplify the AP calculation through the same two scenarios given for MRR, the first scenario gives us the same result again because the user only visits the second document in L: 1/2 = 0,5. As for the second, as the user visits documents at the first and second ranks, the AP value will be calculated as (1/1 + 2/2) / 2 = 1. Let us assume another scenario where that user clicks three suggestions at the second, third, and fifth ranks in the same session. The AP value will be calculated as (1/2 + 2/3 + 3/5) / 3 = 0,59.

The MRR and AP values were calculated by following the above method for each session. Afterward, the MEDEF values were determined by multiplying these values with the other two metrics for each session.

3.3. Research Questions and Evaluation Procedure

In the scope of the present study, four research questions were prepared. This section explains these questions and how the evaluations were carried out respectively.

RQ1. What are the differences between the SAb behaviors in the groups?

The reason for creating this question is first to characterize websites that performed similarly based on users' SAb behaviors. The second is to group these websites according to their performance, as in Table 2. The last is to create a set of propositions being followed as guidance regarding the IR performances of these groups of websites. The findings revealed from this question guided the investigations into the other questions.

RQ2. How do baseline efficiency indicators perform in reflecting users' SAb behavior?

This question attempts to answer how reliable the most preferred efficiency indicators in IR studies are in reflecting the service success of IR systems from a user-based window. In addition, each baseline indicator was ranked from the best to the worst on its trustworthiness.

RQ3. How do MEDEF-based efficiency indicators perform in reflecting users' SAb behavior?

The investigations into this question are in the same line as for the second question. Besides, the findings from this question were also used in comparing MEDEF-based efficiency indicators with the baselines to clarify which indicators are more reliable.

To sum up, while examining all indicators statistically, the focus was on how reliable each indicator was in consistently reflecting behavioral patterns in the propositions. The dataset utilized in the examinations consists of a total of 11,228 rows with 16 different attributes:

- session_id
- department_id
- group (refers to which group a department belongs to)
- the baseline indicators (four kinds in total)
- the MEDEF-based indicators (eight kinds in total)
- abandonment (if the session is abandoned, then this value is *true*; otherwise it is *false*)

Each row in the dataset corresponds to a unique session together with its attributes. As the prepared dataset did not fit the normal distribution for each indicator, nonparametric test methods were utilized during the examinations.

RQ4. Which prediction model created based on the type of indicators shows more successful performance?

All examinations for the other questions were carried out from a statistical perspective. Three more datasets were also created from the same dataset for this question to determine the success of indicators with an ML-based perspective. To ensure that all datasets consisted of sessions with no abandonment, the first dataset (baselines) has three baseline indicators (as independent variables) on each row with its group (as the dependent variable); the second (medef_MRR) also has three MEDEF-based indicators (independent variables; MRR_ALL was excluded) created using the MRR metric as the punishment on each row with its group (dependent variables); the third (medef_AP) has three MEDEF-based indicators (independent variables; AP_ALL was excluded) created using the AP metric as the punishment on each row with its group (dependent variables). Two supervised ML algorithms were utilized to create prediction models based on these prepared datasets: Random Forest Classifier (RF) and Decision Tree Classifier (DT). The holdout sampling was used with three different train-test separation rates (shared in Section 4.4). In addition, as the data on each dataset were unbalanced (Table 2), the stratification process

was made while sampling. Precision (P), F1 score, and Accuracy (ACC) metrics were used to assess the created models by the ML algorithms. The scikit-learn module by Pedregosa et al. (2011) was used for all examinations.

What is expected from the models is not to predict the groups perfectly but to reveal several signs that can be interpreted as directly proportional to findings from statistical examinations.

3.4. Limitations

The possibility of users' IP addresses changing after a while is the first limitation of this study because, while user interactions were organized in the preprocessing phase, these addresses were utilized. The second limitation can be stated as arising from the Javascript code used to record dwell times on the SERP. The possibility of users' browsers not supporting the code properly might cause faulty data recording. Technical problems that might arise from the server infrastructure that hosts the websites can be listed as the last limitation. The situation of users encountering this limitation in information-seeking processes might result in unfinished interactions.

4. FINDINGS

In this section, the question of whether users' search habits have changed from the past to the present is first examined, after which the descriptive statistics of indicators are shared. The section concludes with the relationship among the indicators being described and the research questions being answered.

4.1. Users' Search Habits

While the users conducted searches over the period of one month, they used query terms consisting of different numbers of words. A consideration of all sessions led to five groups being created based on how many words the query terms had (Figure 1).

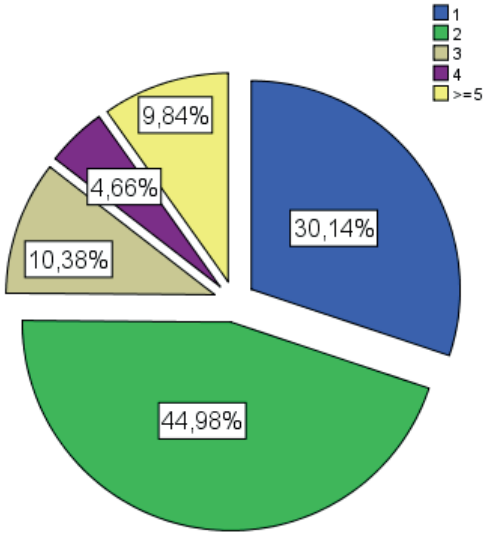


Figure 1. The groups of query terms based on the number of words

It seems that the findings from the studies by Jansen, Spink, Bateman, and Saracevic (1998), Jansen, Spink, and Saracevic (2000), and Jansen and Spink (2005) have maintained their validity in that users still mostly prefer to write a query using terms that consist of two, one, or three word(s). Again, five groups were created to investigate how many visit interactions were made per session based on the number of visits. The findings are given in Figure 2.

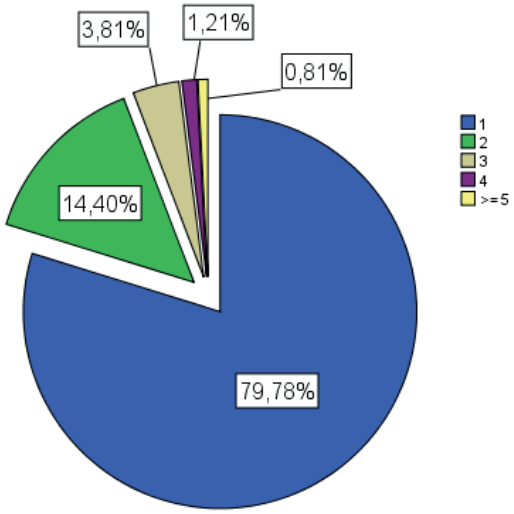


Figure 2. The groups of visit interactions based on the number of visits per session

Even though Jansen and Spink (2005) stated that 30.3% of users examine only one document per session, users' visit interactions in Figure 2 show that almost 80% of users view only one document per session. This can be interpreted as an indication that users have become more impatient while searching. In addition, these findings point out that almost 80% of the TTFC and TTLC values are the same for each session in the prepared datasets. Based on the sessions with no abandonment, the statistics regarding the ranks of documents the users clicked are shared in Figure 3.

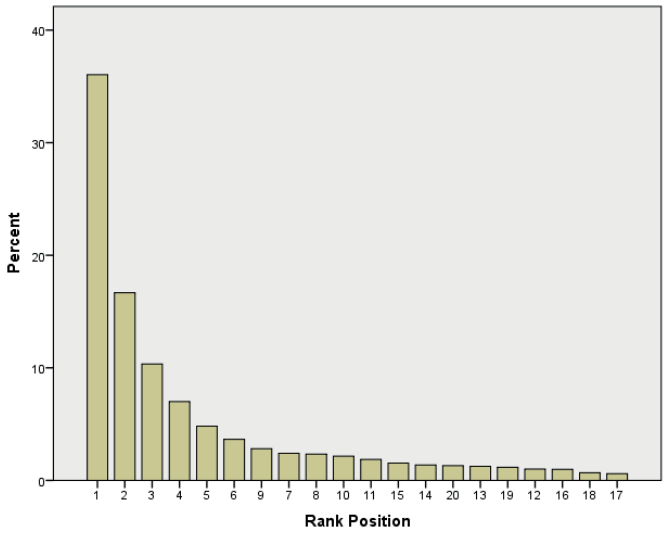


Figure 3. The ranks of the visited documents

Similar to the findings from the studies by Joachims et al. (2005), Agichtein, Brill, Dumais, and Ragno (2006), and Cen et al. (2009), our results show that users mostly prefer to view documents at ranks lower than 10 (Figure 3).

4.2. Descriptive Findings

The descriptive findings that belong to the baseline efficiency indicators for each website group are given in Table 4.

Table 4
The descriptive findings of baseline efficiency indicators

GROUP	A				B				C			
Indicators	**DwSERP**	**TTFC***	**TTLC***	**DwSERP_Ab**	**DwSERP**	**TTFC***	**TTLC***	**DwSERP_Ab**	**DwSERP**	**TTFC***	**TTLC***	**DwSERP_Ab**
Mean	15,52	11,45	60,39	43,24	15,44	12,61	64,87	27,27	14,99	12	60,98	18,03
Median	10,50	9	17	51,10	12,10	11	17	20,20	11,70	10	17	12,20
Std. Dev.	14,38	7,20	144,55	18,12	11,81	7,13	189,66	20,69	11,82	6,87	165,80	16,12
Min.	1	3	3	1	1,50	2	2	1	1	1	1	1
Max.	59,60	28	862	59,60	57,90	30	1785	57,20	59	30	1782	59,90
n		49		1247		199		1118		1899		5311
GROUP	**D**				**E**				**F**			
Indicators	**DwSERP**	**TTFC***	**TTLC***	**DwSERP_Ab**	**DwSERP**	**TTFC***	**TTLC***	**DwSERP_Ab**	**DwSERP**	**TTFC***	**TTLC***	**DwSERP_Ab**
Mean	13,52	11,45	56,86	10,94	11,12	9,73	66,40	11,88	8,95	7,76	129,46	14,97
Median	10,20	10	13	7,45	7,70	8	12	7,90	6,40	6	12	8,80
Std. Dev.	10,92	6,38	138,51	10,28	9,83	6,47	207,12	11,46	8,19	5,23	330,37	15,53
Min.	1	2	2	1	1	1	1	1	1,10	2	2	1
Max.	51,40	29	962	57	49,50	30	1443	59,60	48,70	30	1677	57,20
n		147		304		297		395		139		123

***** These values are in the form of seconds.

In Section 3.2.1, we stated that the website group that managed to keep users most successfully was F, followed by E, D, C, B, and A, respectively. This order also indicates which group performed better than the others. According to the mean values in Table 4, only the DwSERP values carry signs regarding this assumption (here, the mentioned propositions are called "assumptions" because they have not yet been statistically proved). In other words, it can be seen that the mean values decrease from the worst-performing group to the best-performing group. When we focus on the mean TTFC values, this pattern only seems clear for groups D, E, and F. For the mean TTLC values, no consistent pattern was observed. As for the DwSERP_Ab values, only groups A, B, C, and D support the assumptions. Although it is proved in the next section and discussed in Section 5, as a preliminary interpretation, it is better to mention that the DwSERP_Ab indicator shows different characteristics in contrast to the general consensus followed in IR studies. Lastly, as the data of each indicator did not fit the normal distribution (p<0,05), nonparametric test methods were utilized in the analyses. The descriptive findings from the MEDEF-based efficiency indicators are seen in Table 5.

Table 5
The descriptive findings of the MEDEF-based efficiency indicators

GROUP	A							
Indicators	**MRR_DwSERP**	**MRR_TTFC**	**MRR_TTLC**	**MRR_ALL**	**AP_DwSERP**	**AP_TTFC**	**AP_TTLC**	**AP_ALL**
Mean	0,034	0,032	0,021	0,001	0,035	0,035	0,022	0,001
Median	0,020	0,028	0,011	0,00010	0,021	0,028	0,011	0,00010
Std. Dev.	0,038	0,027	0,024	0,004	0,038	0,029	0,024	0,004
Min.	0,00045	0,00101	0,00011	0,0000003	0,00045	0,00101	0,00011	0,0000003
Max.	0,20	0,11	0,08	0,02	0,20	0,11	0,08	0,02
n				49				
GROUP	**B**							
Indicators	**MRR_DwSERP**	**MRR_TTFC**	**MRR_TTLC**	**MRR_ALL**	**AP_DwSERP**	**AP_TTFC**	**AP_TTLC**	**AP_ALL**
Mean	0,046	0,044	0,030	0,001	0,048	0,047	0,031	0,001
Median	0,023	0,028	0,012	0,00008	0,024	0,028	0,013	0,00008
Std. Dev.	0,070	0,051	0,042	0,004	0,071	0,054	0,042	0,004
Min.	0,00086	0,00104	0,00007	0,0000001	0,00115	0,00104	0,00007	0,0000002

Max.	0,53	0,33	0,25	0,04	0,53	0,33	0,25	0,04
n				199				
GROUP				**C**				
Indicators	**MRR_DwSERP**	**MRR_TTFC**	**MRR_TTLC**	**MRR_ALL**	**AP_DwSERP**	**AP_TTFC**	**AP_TTLC**	**AP_ALL**
Mean	0,051	0,044	0,033	0,002	0,054	0,048	0,033	0,002
Median	0,023	0,028	0,013	0,00010	0,026	0,031	0,014	0,00011
Std. Dev.	0,080	0,054	0,052	0,008	0,082	0,056	0,052	0,008
Min.	0,00063	0,00064	0,00003	0,0000001	0,00063	0,00064	0,00003	0,0000001
Max.	0,91	1	1	0,208	0,91	1	1	0,208
n				1899				
GROUP				**D**				
Indicators	**MRR_DwSERP**	**MRR_TTFC**	**MRR_TTLC**	**MRR_ALL**	**AP_DwSERP**	**AP_TTFC**	**AP_TTLC**	**AP_ALL**
Mean	0,053	0,050	0,038	0,002	0,053	0,51	0,038	0,002
Median	0,033	0,033	0,023	0,00020	0,033	0,033	0,023	0,00020
Std. Dev.	0,058	0,051	0,048	0,006	0,060	0,052	0,048	0,006
Min.	0,00123	0,00197	0,00030	0,0000008	0,00178	0,00216	0,00030	0,0000011
Max.	0,36	0,33	0,33	0,042	0,36	0,33	0,33	0,042
n				147				
GROUP				**E**				
Indicators	**MRR_DwSERP**	**MRR_TTFC**	**MRR_TTLC**	**MRR_ALL**	**AP_DwSERP**	**AP_TTFC**	**AP_TTLC**	**AP_ALL**
Mean	0,084	0,070	0,054	0,005	0,085	0,072	0,054	0,005
Median	0,048	0,048	0,031	0,00045	0,050	0,050	0,031	0,00045
Std. Dev.	0,097	0,066	0,066	0,017	0,097	0,067	0,066	0,017
Min.	0,00118	0,00114	0,00010	0,0000003	0,00119	0,00114	0,00010	0,0000003
Max.	0,59	0,5	0,5	0,227	0,59	0,5	0,5	0,227
n				297				
GROUP				**F**				
Indicators	**MRR_DwSERP**	**MRR_TTFC**	**MRR_TTLC**	**MRR_ALL**	**AP_DwSERP**	**AP_TTFC**	**AP_TTLC**	**AP_ALL**
Mean	0,142	0,112	0,076	0,008	0,147	0,119	0,076	0,008
Median	0,077	0,083	0,042	0,00061	0,085	0,089	0,042	0,00067
Std. Dev.	0,173	0,096	0,094	0,025	0,172	0,096	0,094	0,025
Min.	0,00321	0,00529	0,00016	0,0000011	0,00522	0,00529	0,00016	0,0000022
Max.	0,91	0,5	0,5	0,208	0,91	0,5	0,5	0,208
n				139				

With the exception of the MRR_ALL and AP_ALL indicators, a consideration of the mean value of each indicator reveals that all indicators support the assumptions, meaning that the increase in the mean values is directly proportional to the groups' performances. This pattern is not clear for the MRR_ALL and AP_ALL indicators. Even though the TTLC indicator in Table 4 is seen not to be consistent with the assumptions, it was found that the MEDEF positively affected this indicator, making it more consistent, as seen in Table 5. When we focus on each conjugate indicator of each group in Table 5 (such as MRR_DwSERP and AP_DwSERP or MRR_ALL and AP_ALL), it can also be seen that there are slight differences between the indicator values. It is thought that user actions caused this similarity. In other words, as the users mostly viewed only one document in their sessions (Figure 2), the calculation of MRR and AP values was equal for most sessions of these groups. As in the data that belong to the baseline efficiency indicators, the data of the MEDEF-based efficiency indicators did not fit the normal distribution (p<0,05). This is why nonparametric test methods were utilized in the analyses.

4.3. Relationship Among the Indicators

Spearman correlation tests were made on the indicators considering their types. To this end, the findings from the baseline indicators, MEDEF-based indicators created using the MRR metric, and MEDEF-based indicators created using the AP metric are explained, respectively.

Table 6
The findings from the baseline efficiency indicators

INDICATORS	DwSERP	TTFC	TTLC
DwSERP	1,000	,600*	,499*
TTFC	,600*	1,000	,477*
TTLC	,499*	,477*	1,000

* correlation coefficients; $p < 0,01$

According to Table 6, both the DwSERP and TTFC indicators have a positive relationship with the TTLC indicator (r_s: 0,499, 0,477, respectively). As for the relationship between the DwSERP and TTFC indicators, a strong positive relationship was observed (r_s: 0,600, Table 6), which indicates that DwSERP can be foreseen through users' TTFC behaviors and vice versa.

The findings gathered from the MEDEF-based indicators created using the MRR metric are given in Table 7.

Table 7
The findings from the MEDEF-based indicators - MRR

INDICATORS	MRR_DwSERP	MRR_TTFC	MRR_TTLC	MRR_ALL
MRR_DwSERP	1,000	,842*	,714*	,848*
MRR_TTFC	,842*	1,000	,706*	,783*
MRR_TTLC	,714*	,706*	1,000	,905*
MRR_ALL	,848*	,783*	,905*	1,000

* correlation coefficients; $p < 0,01$

Similar but stronger findings were observed for the MEDEF-based indicators as in the baseline indicators (Table 7). In addition, it was revealed that user behaviors can be estimated more consistently from the MEDEF-based indicators based on the MRR metric. These findings also indicate that the MEDEF concretizes the relationship between different user behaviors (the baseline indicators) more clearly. The other findings that belong to the MEDEF-based indicators created using the AP metric are shared in Table 8.

Table 8
The findings from the MEDEF-based indicators - AP

INDICATORS	AP_DwSERP	AP_TTFC	AP_TTLC	AP_ALL
AP_DwSERP	1,000	,829*	,688*	,830*
AP_TTFC	,829*	1,000	,670*	,748*
AP_TTLC	,688*	,670*	1,000	,899*
AP_ALL	,830*	,748*	,899*	1,000

* correlation coefficients; $p < 0,01$

Even though a slight decrease was observed in Table 8 compared to Table 7, our results clearly reveal that the MEDEF-based indicators are more consistent in being able to foresee user behaviors. All findings show the preliminary signs that the MEDEF has a high potential to read user behaviors. The statistical base of whether these signs carry meaning is examined in the next section.

4.4. Investigations of Research Questions

In the scope of the present study, after preprocessing the collected data, twenty-three websites were separated into six groups based on their performances regarding users' Sab behaviors. Whether this grouping is meaningful statistically was investigated with the first research question.

RQ1. What are the differences between the SAb behaviors in the groups?

According to the findings from the Chi-square test made on six groups, the propositions created to guide the other research questions have been found statistically significant (Table 9; $\chi^2 = 638,165$; $p < 0,01$).

Table 9
The performance findings of groups regarding users' SAb behaviors

			Abandonment		Total
			false	true	
Groups	A	Count	49	1247	1296
		% within Groups	3,8%	96,2%	100,0%
	B	Count	199	1118	1317
		% within Groups	15,1%	84,9%	100,0%
	C	Count	1899	5311	7210
		% within Groups	26,3%	73,7%	100,0%
	D	Count	147	304	451
		% within Groups	32,6%	67,4%	100,0%
	E	Count	297	395	692
		% within Groups	42,9%	57,1%	100,0%
	F	Count	139	123	262
		% within Groups	53,1%	46,9%	100,0%
Total		Count	2730	8498	11228
		% within Groups	24,3%	75,7%	100,0%

This research proves that the propositions created presumptively through Table 2 can be followed as guidance to determine the performance of the efficiency indicators. In addition to Table 9, the groups were also compared as pairs (15 comparisons) with the Chi-square test, and it was observed that the results did not change for all comparisons. While answering the other questions, the reflectiveness performances of each efficiency indicator were evaluated through these fifteen performance propositions, and success was determined according to how many propositions could be detected by each indicator.

RQ2. How do baseline efficiency indicators perform in reflecting users' SAb behavior?

Sixty comparisons, 15 for each indicator, were made with the Mann-Whitney U test between the groups to answer this question, and the revealed findings are given in Table 10.

Table 10
The findings regarding the reflectiveness performances of baseline efficiency indicators

Indicators	DwSERP	TTFC	TTLC	DwSERP_Ab
NUS *	9	8	5	12
NUNOS **	6	7	10	3
Rate of NUS	60%	53,33%	33,33%	80%
Rate of NUNOS	40%	46,67%	66,67%	20%

The significance level is chosen as 0,05; $p<0,05$
* The number of comparisons found as significant
** The number of comparisons found as nonsignificant

A consideration of the sessions with no abandonment reveals that the most reliable indicator is DwSERP, followed by TTFC and TTLC, respectively. In addition, strikingly, DwSERP_Ab was determined to be the most successful indicator in reflecting group performances compared to the other baseline indicators (the reason why it is striking is discussed in Section 5).

RQ3. How do MEDEF-based efficiency indicators perform in reflecting users' SAb behavior?

One hundred twenty comparisons, 15 for each MEDEF-based indicator, were made with the Mann-Whitney U test between the groups to answer this question, and the findings are shared in Table 11.

Table 11
The findings regarding the reflectiveness performances of MEDEF-based efficiency indicators

Indicators	MRR_DwSERP	MRR_TTFC	MRR_TTLC	MRR_ALL	AP_DwSERP	AP_TTFC	AP_TTLC	AP_ALL
NUS	12	12	10	10	12	10	11	10
NUNOS	3	3	5	5	3	5	4	5
Rate of NUS	80%	80%	66,67%	66,67%	80%	66,67%	73,33%	66,67%
Rate of NUNOS	20%	20%	33,33%	33,33%	20%	33,33%	26,67%	33,33%

The significance level is chosen as 0,05; $p<0,05$

According to Table 11, DwSERP was found to be the most reliable punishment element regardless of which effectiveness metric is used in the MEDEF. In addition, even though the NUNOS value is three for the MRR_DwSERP and AP_DwSERP indicators, the analyses with the Mann-Whitney U test showed that these indicators tended to reflect the three propositions. For the MRR metric, it is also clear that using TTFC as a punishment element reveals the same results. Moreover, TTLC is in the second position for the AP metric in reflecting group performances successfully. When we consider all MEDEF-based indicators, it can be stated that they have outperformed the reflectiveness performance of the baseline indicators unambivalently.

RQ4. Which prediction model created based on the type of indicators shows more successful performance?

Before creating the prediction models, the data on each dataset were separated into two sets (training and test) depending on the separation rate used. As the purpose was to reach similar findings to the second and third research questions, no preprocessing was applied on all sets, and the data was utilized without a touch. The findings gathered from two different ML algorithms on each dataset are given in Table 12.

Table 12
The performances of created prediction models

Datasets	Algorithms	Separation Rate (Train: Test)	ACC	P	F1 Score
baselines	DT	0,9: 0,1	0,696	0,484	0,571
		0,8: 0,2	0,696	0,484	0,571
		0,7: 0,3	0,659	0,533	0,576
	RF	0,9: 0,1	0,586	0,533	0,554
		0,8: 0,2	0,590	0,542	0,564
		0,7: 0,3	0,602	0,527	0,559
medef_MRR	DT	0,9: 0,1	0,696	0,504	0,578
		0,8: 0,2	0,698	0,512	0,579
		0,7: 0,3	0,698	0,513	0,579
	RF	0,9: 0,1	0,626	0,570	0,592
		0,8: 0,2	0,612	0,565	0,586
		0,7: 0,3	0,624	0,555	0,585
medef_AP	DT	0,9: 0,1	0,700	0,537	0,579
		0,8: 0,2	0,698	0,512	0,579
		0,7: 0,3	0,698	0,513	0,579
	RF	0,9: 0,1	0,623	0,559	0,587
		0,8: 0,2	0,612	0,559	0,582
		0,7: 0,3	0,629	0,552	0,583

When considering the DT algorithm, although the ACC values on each separation rate do not show the performance differences clearly, the P value and F1 score revealed that the MEDEF supported the algorithm in detecting which session belonged to which website group (Table 12). For the RF algorithm, however, the prediction performance of each model was observed more clearly in the process of proving the success of MEDEF, regardless of which type of evaluation metric was considered (Table 12). These findings correspond to the findings in the third research question, meaning that the MEDEF allows user behavior regarding baseline indicators to be interpreted in a more reliable way.

5. DISCUSSION

In the scope of the present study, users' search interactions were recorded through search modules that were integrated into the department websites of a university in Turkey for one month. After preprocessing the collected data, the websites were separated into six groups based on users' SAb behaviors, and fifteen propositions were created regarding the IR performances of these groups. Afterward, 12 efficiency indicators, consisting of four baselines already used in IR studies, and eight created by following the proposed efficiency formula, were compared to each other. While the statistical investigations focused on how many propositions these indicators can reflect and how reliable they are for evaluations and optimizations, the ML-based examinations were carried out to reveal signs that can be interpreted as directly proportional to findings from the statistical investigations.

A consideration of the baseline efficiency indicators reveals that the most reliable indicator was found to be DwSERP for sessions with no abandonment, followed by TTFC and TTLC. In his study, Arguello (2014) found that dwell time is a useful predictor when determining search task difficulty. Even though the author considered the dwell time on landing pages, the findings in the present study are seen to be in the same line, meaning that the dwell time is a reliable indicator regardless of which type of usage is preferred. In the study by Jung, Herlocker, and Webster (2007), the authors tried to improve the quality of search results and emphasized that the last visited documents were useful. Although the context between the present and their study is different, the TTLC indicator was not seen as a reliable efficiency indicator in this study.

A separate parenthesis is required for the DwSERP_Ab indicator. The expectation of DwSERP_Ab is that it will inversely reflect the success of groups in the propositions because the consensus in IR studies is that the decrease in the DwSERP_Ab value points out negative/bad experiences (Song, Shi, White, and Awadallah, 2014; Borisov, Markov, de Rijke, and Serdyukov, 2016). However, with the exception of the E and F groups, Table 4 shows that the increase in group performances and the mean values of the DwSERP_Ab indicator are inversely proportional to each other. These findings are also supported by the analysis results from the Mann-Whitney U test, meaning that the present study shows opposing results. The reason for this could be that after users saw the first response from the search modules, they reformulated their query in a short time and accordingly reached satisfying information in their consecutive queries. The other possibility, as indicated in the study by Stamou and Efthimiadis (2010), could be that users met their information needs by briefly examining only the result snippets on the SERP without spending much time. Regardless of what type of interaction users performed, this result can be interpreted to show that the more "negative abandonment" IR systems have, the more efficient performance users show.

As for the created indicators, it was proved that all eight indicators outperformed the baseline indicators for the sessions with no abandonment. Moreover, the TTFC and TTLC indicators, which performed less consistently than DwSERP, also showed more consistent performance when used as punishments. As DwSERP is already a reliable indicator, the reflectiveness of both the MEDEF-based indicators created with it also showed the most successful performance. In general terms, the MEDEF further strengthened the reflections in baseline efficiency indicators.

6. CONCLUSION

In the present study, a modular efficiency determination formula, MEDEF, is proposed. Using the MEDEF, eight efficiency indicators were created and compared with the baseline indicators already used in IR studies. The findings revealed that the MEDEF-based indicators outperform the baselines. It is believed that indicators created by following the MEDEF will likely show more reliable findings in evaluations and optimizations of IR systems. Moreover, only three different efficiency indicators (the baselines) were used as punishments in this study to create the MEDEF-based indicators. The common part

of these baselines is that they are implicit data, meaning that other implicit data types can also be integrated into the MEDEF. Apart from this, if implicit data, which will be integrated, are a kind of reward rather than a punishment, then the punishment metric can be used reciprocally.

As in the baseline indicators, two effectiveness metrics used in the MEDEF-based indicators were also based on natural user behaviors. In other words, they also are implicit data (clickthrough). For future studies, it is thought that different metrics (such as NDCG or ERR) based on explicit data, which can be gathered from users while they search, can also be integrated into the MEDEF.

In conclusion, the present study revealed the trustworthiness and consistency of four different efficiency indicators already in use in IR studies for the purpose of optimization. In addition, it was determined that the MEDEF-based indicators were more successful than these indicators. This result can be interpreted to show that the indicators created by following the MEDEF will boost the success of optimization-based IR studies in the future. Moreover, the question of how IR systems can be evaluated using only efficiency indicators was enlightened. While the reflectiveness performance of baseline efficiency indicators (except for DwSERP_Ab) is not adequate for individual usage, the MEDEF-based indicators showed successful performance, meaning that the MEDEF-based indicators can be utilized individually to evaluate IR systems from the usability perspective, as exemplified in the present study.

Peer-review: Externally peer-reviewed.
Conflict of Interest: The author has no conflict of interest to declare.
Grant Support: The author declared that this study has received no financial support.

REFERENCES

Agichtein, E., Brill, E., & Dumais, S. (2006). Improving web search ranking by incorporating user behavior information. *Proceedings of the Twenty-Ninth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2006, 19–26. New York, New York, USA: ACM Press. https://doi.org/10.1145/1148170.1148177

Agichtein, E., Brill, E., Dumais, S., & Ragno, R. (2006). Learning user interaction models for predicting web search result preferences. *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '06*, 3. New York, New York, USA: ACM Press. https://doi.org/10.1145/1148170.1148175

Alhabashneh, O., Iqbal, R., Doctor, F., & James, A. (2017). Fuzzy rule based profiling approach for enterprise information seeking and retrieval. *Information Sciences*, 394–395, 18–37. https://doi.org/10.1016/J.INS.2016.12.040

Arguello, J. (2014). Predicting search task difficulty. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 8416 LNCS (pp. 88–99). https://doi.org/10.1007/978-3-319-06028-6_8

Arkhipova, O., Grauer, L., Kuralenok, I., & Serdyukov, P. (2015). Search Engine Evaluation based on Search Engine Switching Prediction. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 723–726. New York, NY, USA: ACM. https://doi.org/10.1145/2766462.2767786

Aula, A., Khan, R. M., & Guan, Z. (2010). How does search behavior change as search becomes more difficult? *Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI '10*, 35. New York, New York, USA: ACM Press. https://doi.org/10.1145/1753326.1753333

Balakrishnan, V., & Zhang, X. (2014). Implicit user behaviours to improve post-retrieval document relevancy. *Computers in Human Behavior*, 33, 104–112. https://doi.org/10.1016/J.CHB.2014.01.001

Beierle, F., Aizawa, A., Collins, A., & Beel, J. (2020). Choice overload and recommendation effectiveness in related-article recommendations. *International Journal on Digital Libraries*, 21(3), 231–246. https://doi.org/10.1007/s00799-019-00270-7

Belkin, N. J., Kelly, D. F., Kim, G., Kim, J. Y., Lee, H., Muresan, G., Tang, M. C., Yuan, X., Cool, C. (2003). Query Length in Interactive Information Retrieval. SIGIR Forum (ACM Special Interest Group on Information Retrieval), (SPEC. ISS.), 205–212. New York, New York, USA: ACM Press. https://doi.org/10.1145/860472.860474

Borisov, A., Markov, I., de Rijke, M., & Serdyukov, P. (2016). A Context-aware Time Model for Web Search. *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 205–214. New York, NY, USA: ACM. https://doi.org/10.1145/2911451.2911504

Buckley, C., & Voorhees, E. M. (2004). Retrieval evaluation with incomplete information. *Proceedings of the 27th Annual International Conference on Research and Development in Information Retrieval - SIGIR '04*, 25. New York, New York, USA: ACM Press. https://doi.org/10.1145/1008992.1009000

Büttcher, S., Clarke, C. L. A., & Cormack, G. V. (2010). *Information Retrieval: Implementing and Evaluating Search Engines*. The MIT Press.

Cen, R., Liu, Y., Zhang, M., Ru, L., & Ma, S. (2009). Study on the click context of web search users for reliability analysis. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 5839 LNCS (pp. 397–408). https://

doi.org/10.1007/978-3-642-04769-5_35

Chapelle, O., Metlzer, D., Zhang, Y., & Grinspan, P. (2009). Expected reciprocal rank for graded relevance. *Proceeding of the 18th ACM Conference on Information and Knowledge Management - CIKM '09*, 621. New York, New York, USA: ACM Press. https://doi.org/10.1145/1645953.1646033

Claypool, M., Brown, D., Le, P., & Waseda, M. (2001). Inferring user interest. *IEEE Internet Computing*, 5(6), 32–39. https://doi.org/10.1109/4236.968829

Colace, F., De Santo, M., Greco, L., & Napoletano, P. (2015). Improving relevance feedback-based query expansion by the use of a weighted word pairs approach. *Journal of the Association for Information Science and Technology*, 66(11), 2223–2234. https://doi.org/10.1002/asi.23331

Croft, B., Metzler, D., & Strohman, T. (2009). *Search Engines: Information Retrieval in Practice* (1st ed.). Boston: Pearson.

Diriye, A., White, R., Buscher, G., & Dumais, S. (2012). Leaving so soon?: understanding and predicting web search abandonment rationales. *Proceedings of the 21st ACM International Conference on Information and Knowledge Management - CIKM '12*, 1025. New York, New York, USA: ACM Press. https://doi.org/10.1145/2396761.2398399

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., and Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. Retrieved from http://scikit-learn.sourceforge.net.

Fox, S., Karnawat, K., Mydland, M., Dumais, S., & White, T. (2005). Evaluating implicit measures to improve Web search. *ACM Transactions on Information Systems*, 23(2), 147–168. https://doi.org/10.1145/1059981.1059982

Frøkjær, E., Hertzum, M., & Hornbæk, K. (2000). Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '00*, 345–352. New York, New York, USA: ACM Press. https://doi.org/10.1145/332040.332455

Hassan, A. (2012). A semi-supervised approach to modeling web search satisfaction. *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '12*, 275. New York, New York, USA: ACM Press. https://doi.org/10.1145/2348283.2348323

Hassan, A., Song, Y., & He, L. (2011). A task level metric for measuring web search satisfaction and its application on improving relevance estimation. *Proceedings of the 20th ACM International Conference on Information and Knowledge Management - CIKM '11*, 125. New York, New York, USA: ACM Press. https://doi.org/10.1145/2063576.2063599

Hassan, A., White, R. W., Dumais, S. T., & Wang, Y.-M. (2014). Struggling or exploring? Disambiguating Long Search Sessions. *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, 53–62. New York, NY, USA: ACM. https://doi.org/10.1145/2556195.2556221

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies*, 64(2), 79–102. https://doi.org/10.1016/J.IJHCS.2005.06.002

Jansen, B. J., & Spink, A. (2005). An analysis of Web searching by European AlltheWeb.com users. *Information Processing & Management*, 41(2), 361–381. https://doi.org/10.1016/S0306-4573(03)00067-0

Jansen, B. J., Spink, A., Bateman, J., & Saracevic, T. (1998). Real life information retrieval: a study of user queries on the Web. ACM SIGIR Forum, 32(1), 5–17. https://doi.org/10.1145/281250.281253

Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: a study and analysis of user queries on the web. *Information Processing & Management*, 36(2), 207–227. https://doi.org/10.1016/S0306-4573(99)00056-4

Jarvelin, K., & Kekalainen, J. (2000). IR evaluation methods for retrieving highly relevant documents. SIGIR Forum (ACM Special Interest Group on Information Retrieval), 41–48. New York, New York, USA: ACM Press. https://doi.org/10.1145/3130348.3130374

Järvelin, K., & Kekäläinen, J. (2002). Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems*, 20(4), 422–446. https://doi.org/10.1145/582415.582418

Jiang, D., Leung, K. W. T., Yang, L., & Ng, W. (2015). Query suggestion with diversification and personalization. *Knowledge-Based Systems*, 89, 553–568. https://doi.org/10.1016/J.KNOSYS.2015.09.003

Joachims, T. (2002). Optimizing search engines using clickthrough data. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 133–142. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/775047.775067

Joachims, T., Granka, L., Pan, B., Hembrooke, H., & Gay, G. (2005). Accurately interpreting clickthrough data as implicit feedback. *SIGIR 2005 - Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 154–161. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/1076034.1076063

Jung, S., Herlocker, J. L., & Webster, J. (2007). Click data as implicit relevance feedback in web search. *Information Processing & Management*, 43(3), 791–807. https://doi.org/10.1016/J.IPM.2006.07.021

Kelly, D., & Fu, X. (2007). Eliciting better information need descriptions from users of information search systems. *Information Processing & Management*, 43(1), 30–46. https://doi.org/10.1016/J.IPM.2006.03.006

Kim, Y., Hassan, A., White, R. W., & Zitouni, I. (2014). Modeling dwell time to predict click-level satisfaction. *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, 193–202. New York, NY, USA: ACM. https://doi.org/10.1145/2556195.2556220

Kuhar, M., & Merčun, T. (2022). Exploring user experience in digital libraries through questionnaire and eye-tracking data. *Library & Information Science Research*, 44(3), 101175. https://doi.org/10.1016/J.LISR.2022.101175

Lee, C.-J., Teevan, J., & de la Chica, S. (2014). Characterizing multi-click search behavior and the risks and opportunities of changing results during use. *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, 515–524. New York, NY, USA: ACM. https://doi.org/10.1145/2600428.2609588

Liu, Y., Chen, Y., Tang, J., Sun, J., Zhang, M., Ma, S., & Zhu, X. (2015). Different Users, Different Opinions. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 493–502. New York, NY, USA: ACM. https://doi.org/10.1145/2766462.2767721

Liu, C., Gwizdka, J., & Liu, J. (2010). Helping identify when users find useful documents: examination of query reformulation intervals. *Proceeding of the Third Symposium on Information Interaction in Context - IIiX '10*, 215. New York, New York, USA: ACM Press. https://doi.org/10.1145/1840784.1840816

Liu, C., Liu, J., & Belkin, N. J. (2014). Predicting Search Task Difficulty at Different Search Stages. *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, 569–578. New York, NY, USA: ACM. https://doi.org/10.1145/2661829.2661939

Makkar, A., & Kumar, N. (2018). User behavior analysis-based smart energy management for webpage ranking: Learning automata-based solution. *Sustainable Computing: Informatics and Systems*, 20, 174–191. https://doi.org/10.1016/J.SUSCOM.2018.02.003

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. New York: Cambridge University Press.

Moffat, A., & Zobel, J. (2008). Rank-biased precision for measurement of retrieval effectiveness. *ACM Transactions on Information Systems*, 27(1), 1–27. https://doi.org/10.1145/1416950.1416952

Nasir, J. A., Varlamis, I., & Ishfaq, S. (2019). A knowledge-based semantic framework for query expansion. *Information Processing & Management*, 56(5), 1605–1617. https://doi.org/10.1016/j.ipm.2019.04.007

Nie, L., Jiang, H., Ren, Z., Sun, Z., & Li, X. (2016). Query Expansion Based on Crowd Knowledge for Code Search. *IEEE Transactions on Services Computing*, 9(5), 771–783. https://doi.org/10.1109/TSC.2016.2560165

Nielsen, J. (1993). *Usability Engineering* (1st ed.). Morgan Kaufmann.

Radlinski, F., Kurup, M., & Joachims, T. (2008). How does clickthrough data reflect retrieval quality? *Proceeding of the 17th ACM Conference on Information and Knowledge Mining - CIKM '08*, 43. New York, New York, USA: ACM Press. https://doi.org/10.1145/1458082.1458092

Rosenzweig, E. (2015). *Successful User Experience: Strategies and Roadmaps* (1st ed.). Morgan Kaufmann.

Rubin, J., & Chisnell, D. (2008). *Handbook Of Usability Testing How To Plan Design And Conduct Effective Tests*. Wiley Pub.

Sarkar, S., Mitsui, M., Liu, J., & Shah, C. (2020). Implicit information need as explicit problems, help, and behavioral signals. *Information Processing & Management*, 57(2), 102069. https://doi.org/10.1016/J.IPM.2019.102069

Singh, J., & Sharan, A. (2017). A new fuzzy logic-based query expansion model for efficient information retrieval using relevance feedback approach. *Neural Computing and Applications*, 28(9), 2557–2580. https://doi.org/10.1007/s00521-016-2207-x

Singla, A., & White, R. W. (2010). Sampling high-quality clicks from noisy click data. *Proceedings of the 19th International Conference on World Wide Web - WWW '10*, 1187. New York, New York, USA: ACM Press. https://doi.org/10.1145/1772690.1772867

Song, Y., Shi, X., White, R., & Awadallah, A. H. (2014). Context-aware web search abandonment prediction. *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, 93–102. New York, NY, USA: ACM. https://doi.org/10.1145/2600428.2609604

Stamou, S., & Efthimiadis, E. N. (2010). Interpreting user inactivity on search results. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 5993 LNCS (pp. 100–113). https://doi.org/10.1007/978-3-642-12275-0_12

Vidinli, I. B., & Ozcan, R. (2016). New query suggestion framework and algorithms: A case study for an educational search engine. *Information Processing and Management*, 52(5), 733–752. https://doi.org/10.1016/j.ipm.2016.02.001

Voorhees, E. M. (1999). The TREC-8 Question Answering Track Report. *Text REtrieval Conference*, 77–82.

Zhai, C., & Massung, S. (2016). *Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining*. ACM. https://doi.org/10.1145/2915031

INFORMATION FOR AUTHORS

DESCRIPTION

Acta INFOLOGICA (ACIN) is the publication of Informatics Department of the Istanbul University. It is an open access, scholarly, peer-reviewed journal published biannually in June and December. The journal was founded in 2017.

AIM AND SCOPE

ACIN aims to contribute to the scientific community interested in the field of informatics and aims to provide a platform for researchers exploring issues based on the concepts of data-information-knowledge, information and communication technologies and applications. The journal welcomes multidisciplinary studies regarding the field as well.

The areas of study covered in the scope of ACIN are in below;
Intelligent Systems
Information Security and Law
Knowledge Management
Computer Networks
Computer Architecture
Information Systems
Bioinformatics
Geographic Information Systems
E-Applications
Internet Technologies
Decision Support Systems and Business Intelligence
Microcontroller and Applications
Mobile Systems
Modeling and Optimization
Project Management
Social and Digital Media
Data Mining
Database Systems
Artificial Intelligence and Machine Learning
Software Engineering

EDITORIAL POLICIES AND PEER REVIEW PROCESS

Publication Policy

The subjects covered in the manuscripts submitted to the Journal for publication must be in accordance with the aim and scope of the journal. The journal gives priority to original research papers submitted for publication.

General Principles

Only those manuscripts approved by its every individual author and that were not published before in or sent to another journal, are accepted for evaluation.

Submitted manuscripts that pass preliminary control are scanned for plagiarism using iThenticate software. After plagiarism check, the eligible ones are evaluated by editor-in-chief for their originality, methodology, the importance of the subject covered and compliance with the journal scope.

Short presentations that took place in scientific meetings can be referred if indicated in the article. The editor hands over the papers matching the formal rules to at least two national/international referees for evaluation and gives green light for publication upon modification by the authors in accordance with the referees' claims. Changing the name of an author (omission, addition or order) in papers submitted to the Journal requires written permission of all declared authors. Refused manuscripts and graphics are not returned to the author.

INFORMATION FOR AUTHORS

OPEN ACCESS STATEMENT

The journal is an open access journal and all content is freely available without charge to the user or his/her institution. Except for commercial purposes, users are allowed to read, download, copy, print, search, or link to the full texts of the articles in this journal without asking prior permission from the publisher or the author. This is in accordance with the BOAI definition of open access.

The open access articles in the journal are licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license.

Article Processing Charge
All expenses of the journal are covered by the Istanbul University. Processing and publication are free of charge with the journal. There is no article processing charges or submission fees for any submitted or accepted articles.

Peer Review Process

Only those manuscripts approved by its every individual author and that were not published before in or sent to another journal, are accepted for evaluation.

Submitted manuscripts that pass preliminary control are scanned for plagiarism using iThenticate software. After plagiarism check, the eligible ones are evaluated by Editor-in-Chief for their originality, methodology, the importance of the subject covered and compliance with the journal scope. Editor-in-Chief evaluates manuscripts for their scientific content without regard to ethnic origin, gender, sexual orientation, citizenship, religious belief or political philosophy of the authors and ensures a fair double-blind peer review of the selected manuscripts.

The selected manuscripts are sent to at least two national/international referees for evaluation and publication decision is given by Editor-in-Chief upon modification by the authors in accordance with the referees' claims.

Editor-in-Chief does not allow any conflicts of interest between the authors, editors and reviewers and is responsible for final decision for publication of the manuscripts in the Journal.

Reviewers' judgments must be objective. Reviewers' comments on the following aspects are expected while conducting the review.

- Does the manuscript contain new and significant information?
- Does the abstract clearly and accurately describe the content of the manuscript?
- Is the problem significant and concisely stated?
- Are the methods described comprehensively?
- Are the interpretations and consclusions justified by the results?
- Is adequate references made to other Works in the field?
- Is the language acceptable?

Reviewers must ensure that all the information related to submitted manuscripts is kept as confidential and must report to the editor if they are aware of copyright infringement and plagiarism on the author's side.

A reviewer who feels unqualified to review the topic of a manuscript or knows that its prompt review will be impossible should notify the editor and excuse himself from the review process.

The editor informs the reviewers that the manuscripts are confidential information and that this is a privileged interaction. The reviewers and editorial board cannot discuss the manuscripts with other persons. The anonymity of the referees is important.

COPYRIGHT NOTICE
Authors publishing with the journal retain the copyright to their work licensed under the Creative Commons Attribution-NonCommercial 4.0 International license (CC BY-NC 4.0) (https://creativecommons.org/licenses/by-nc/4.0/) and grant the Publisher non-exclusive commercial right to publish the work. CC BY-NC 4.0 license permits unrestricted, non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

INFORMATION FOR AUTHORS

PUBLICATION ETHICS AND PUBLICATION MALPRACTICE STATEMENT

Acta INFOLOGICA (ACIN) is committed to upholding the highest standards of publication ethics and pays regard to Principles of Transparency and Best Practice in Scholarly Publishing published by the Committee on Publication Ethics (COPE), the Directory of Open Access Journals (DOAJ), to access the Open Access Scholarly Publishers Association (OASPA), and the World Association of Medical Editors (WAME) on https://publicationethics.org/resources/guidelines-new/principles-transparency-and-best-practice-scholarly-publishing
All parties involved in the publishing process (Editors, Reviewers, Authors and Publishers) are expected to agree on the following ethical principles.

All submissions must be original, unpublished (including as full text in conference proceedings), and not under the review of any other publication synchronously. Each manuscript is reviewed by one of the editors and at least two referees under double-blind peer review process. Plagiarism, duplication, fraud authorship/denied authorship, research/data fabrication, salami slicing/salami publication, breaching of copyrights, prevailing conflict of interest are unnethical behaviors.

All manuscripts not in accordance with the accepted ethical standards will be removed from the publication. This also contains any possible malpractice discovered after the publication. In accordance with the code of conduct we will report any cases of suspected plagiarism or duplicate publishing.

RESEARCH ETHICS

Acta INFOLOGICA (ACIN) adheres to the highest standards in research ethics and follows the principles of international research ethics as defined below. The authors are responsible for the compliance of the manuscripts with the ethical rules.

- Principles of integrity, quality and transparency should be sustained in designing the research, reviewing the design and conducting the research.
- The research team and participants should be fully informed about the aim, methods, possible uses and requirements of the research and risks of participation in research.
- The confidentiality of the information provided by the research participants and the confidentiality of the respondents should be ensured. The research should be designed to protect the autonomy and dignity of the participants.
- Research participants should participate in the research voluntarily, not under any coercion.
- Any possible harm to participants must be avoided. The research should be planned in such a way that the participants are not at risk.
- The independence of research must be clear; and any conflict of interest or must be disclosed.
- In experimental studies with human subjects, written informed consent of the participants who decide to participate in the research must be obtained. In the case of children and those under wardship or with confirmed insanity, legal custodian's assent must be obtained.
- If the study is to be carried out in any institution or organization, approval must be obtained from this institution or organization.
- In studies with human subject, it must be noted in the method's section of the manuscript that the informed consent of the participants and ethics committee approval from the institution where the study has been conducted have been obtained.

AUTHOR RESPONSIBILITIES

It is authors' responsibility to ensure that the article is in accordance with scientific and ethical standards and rules. And authors must ensure that submitted work is original. They must certify that the manuscript has not previously been published elsewhere or is not currently being considered for publication elsewhere, in any language. Applicable copyright laws and conventions must be followed. Copyright material (e.g. tables, figures or extensive quotations) must be reproduced only with appropriate permission and acknowledgement. Any work or words of other authors, contributors, or sources must be appropriately credited and referenced.

All the authors of a submitted manuscript must have direct scientific and academic contribution to the manuscript. The author(s) of the original research articles is defined as a person who is significantly involved in "conceptualization and design of the study", "collecting the data", "analyzing the data", "writing the manuscript", "reviewing the manuscript with a critical perspective" and "planning/conducting the study of the manuscript and/or revising it". Fund raising, data collection or supervision of the research group are not sufficient roles to be accepted as an author. The author(s) must meet all these criteria described above. The order of names in the author list of an article must be a co-decision and it must be indicated in the Copyright Agreement Form. The individuals who do not meet the authorship criteria but contributed to the study must take place in the acknowledgement section. Individuals providing technical support, assisting writing, providing a general support, providing material or financial support are examples to be indicated in acknowledgement section.

INFORMATION FOR AUTHORS

All authors must disclose all issues concerning financial relationship, conflict of interest, and competing interest that may potentially influence the results of the research or scientific judgment.

When an author discovers a significant error or inaccuracy in his/her own published paper, it is the author's obligation to promptly cooperate with the Editor to provide retractions or corrections of mistakes.

RESPONSIBILITY FOR THE EDITOR AND REVIEWERS

Editor-in-Chief evaluates manuscripts for their scientific content without regard to ethnic origin, gender, sexual orientation, citizenship, religious belief or political philosophy of the authors. He/She provides a fair double-blind peer review of the submitted articles for publication and ensures that all the information related to submitted manuscripts is kept as confidential before publishing.

Editor-in-Chief is responsible for the contents and overall quality of the publication. He/She must publish errata pages or make corrections when needed.

Editor-in-Chief does not allow any conflicts of interest between the authors, editors and reviewers. Only he has the full authority to assign a reviewer and is responsible for final decision for publication of the manuscripts in the Journal.

Reviewers must have no conflict of interest with respect to the research, the authors and/or the research funders. Their judgments must be objective.

Reviewers must ensure that all the information related to submitted manuscripts is kept as confidential and must report to the editor if they are aware of copyright infringement and plagiarism on the author's side.

A reviewer who feels unqualified to review the topic of a manuscript or knows that its prompt review will be impossible should notify the editor and excuse himself from the review process.

The editor informs the reviewers that the manuscripts are confidential information and that this is a privileged interaction. The reviewers and editorial board cannot discuss the manuscripts with other persons. The anonymity of the referees must be ensured. In particular situations, the editor may share the review of one reviewer with other reviewers to clarify a particular point.

MANUSCRIPT ORGANIZATION

LANGUAGE

The publication language of the journal is English.

Manuscript Organization and Submission

All correspondence will be sent to the first-named author unless otherwise specified. Manuscpript is to be submitted online via dergipark.org.tr/login that can be accessed at http://acin.istanbul.edu.tr and it must be accompanied by a title page specifying the article category (i.e. research article, review etc.) and including information about the manuscript (see the Submission Checklist) and cover letter to the editor. Manuscripts should be prepared in Microsoft Word 2003 and upper versions. In addition, Copyright Agreement Form that has to be signed by all authors must be submitted.

1. Use ACIN article document as a template if you are using Microsoft Word 6.0 or upper versions. Otherwise, use this document as an instruction set.
2. The first letters of words in the article title should be written in uppercase; the entire title should not be capitalized. Avoid writing formulas in the title. Do not write "(Invited)" or similar expressions in the title.
3. The abstract must be between 150–250 words and written as one paragraph. It should not contain displayed mathematical equations or tabular material. The abstract should include three to five different keywords or phrases, as this will help readers to find it. It is important to avoid over-repetition of such phrases as this can result in a page being rejected by search engines. Ensure that your abstract reads well and is grammatically correct.
4. Underneath the abstracts, a minimum of 3 and a maximum of 5 keywords that inform the reader about the content of the study should be specified. Keywords must be defined by taking into consideration authorities like "TR Dizin Anahtar Terimler Listesi", "Medical

Subject Headings", "CAB Theasarus", "JISCT, "ERIC", etc.

5. The manuscripts should contain mainly these components: title, abstract and keywords; sections, references, tables and figures.
6. A title page including author information must be submitted together with the manuscript. The title page is to include fully descriptive title of the manuscript and, affiliation, title, e-mail address, ORCID, postal address, phone, mobile phone and fax number of the author(s) (see The Submission Checklist).
7. References should be prepared as APA 6th edition.

REFERENCES

Reference Style and Format

Acta INFOLOGICA (ACIN) complies with APA (American Psychological Association) style 6th Edition for referencing and quoting. For more information:

- American Psychological Association. (2010). Publication manual of the American Psychological Association (6th ed.). Washington, DC: APA.

- http://www.apastyle.org

Accuracy of citation is the author's responsibility. All references should be cited in text. Reference list must be in alphabetical order. Type references in the style shown below.

Citations in the Text

Citations must be indicated with the author surname and publication year within the parenthesis.
If more than one citation is made within the same paranthesis, separate them with (;).

Samples:

More than one citation;
(Esin, et al., 2002; Karasar, 1995)
Citation with one author;
(Akyolcu, 2007)
Citation with two authors;
(Sayıner & Demirci, 2007)
Citation with three, four, five authors;
First citation in the text: (Ailen, Ciambrune, & Welch, 2000) Subsequent citations in the text: (Ailen, et al., 2000)
Citations with more than six authors;
(Çavdar, et al., 2003)

Citations in the Reference

All the citations done in the text should be listed in the References section in alphabetical order of author surname without numbering. Below given examples should be considered in citing the references.

Basic Reference Types

Book

a) Turkish Book

Karasar, N. (1995). *Araştırmalarda rapor hazırlama* (8th ed.) [Preparing research reports]. Ankara, Turkey: 3A Eğitim Danışmanlık Ltd.

b) Book Translated into Turkish

Mucchielli, A. (1991). *Zihniyetler* [Mindsets] (A. Kotil, Trans.). İstanbul, Turkey: İletişim Yayınları.

c) Edited Book

Ören, T., Üney, T., & Çölkesen, R. (Eds.). (2006). *Türkiye bilişim ansiklopedisi* [Turkish Encyclopedia of Informatics]. İstanbul, Turkey: Papatya Yayıncılık.

d) Turkish Book with Multiple Authors

Tonta, Y., Bitirim, Y., & Sever, H. (2002). T*ürkçe arama motorlarında performans değerlendirme* [Performance evaluation in Turkish search engines]. Ankara, Turkey: Total Bilişim.

e) Book in English

Kamien R., & Kamien A. (2014). *Music: An appreciation.* New York, NY: McGraw-Hill Education.

f) Chapter in an Edited Book

Bassett, C. (2006). Cultural studies and new media. In G. Hall & C. Birchall (Eds.), *New cultural studies: Adventures in theory* (pp. 220–237). Edinburgh, UK: Edinburgh University Press.

g) Chapter in an Edited Book in Turkish

Erkmen, T. (2012). Örgüt kültürü: Fonksiyonları, öğeleri, işletme yönetimi ve liderlikteki önemi [Organization culture: Its functions, elements and importance in leadership and business management]. In M. Zencirkıran (Ed.), *Örgüt sosyolojisi* [Organization sociology] (pp. 233–263). Bursa, Turkey: Dora Basım Yayın.

h) Book with the same organization as author and publisher

American Psychological Association. (2009). *Publication manual of the American psychological association* (6th ed.). Washington, DC: Author.

Article

a) Turkish Article

Mutlu, B., & Savaşer, S. (2007). Çocuğu ameliyat sonrası yoğun bakımda olan ebeveynlerde stres nedenleri ve azaltma girişimleri [Source and intervention reduction of stress for parents whose children are in intensive care unit after surgery]. *Istanbul University Florence Nightingale Journal of Nursing, 15*(60), 179–182.

b) English Article

de Cillia, R., Reisigl, M., & Wodak, R. (1999). The discursive construction of national identity. *Discourse and Society, 10*(2), 149–173. http://dx.doi.org/10.1177/0957926599010002002

c) Journal Article with DOI and More Than Seven Authors

Lal, H., Cunningham, A. L., Godeaux, O., Chlibek, R., Diez-Domingo, J., Hwang, S.-J. ... Heineman, T. C. (2015). Efficacy of an adjuvanted herpes zoster subunit vaccine in older adults. *New England Journal of Medicine, 372,* 2087–2096. http://dx.doi.org/10.1056/NEJMoa1501184

d) Journal Article from Web, without DOI

Sidani, S. (2003). Enhancing the evaluation of nursing care effectiveness. *Canadian Journal of Nursing Research, 35*(3), 26–38. Retrieved from http://cjnr.mcgill.ca

e) Journal Article wih DOI

Turner, S. J. (2010). Website statistics 2.0: Using Google Analytics to measure library website effectiveness. *Technical Services Quarterly, 27,* 261–278. http://dx.doi.org/10.1080/07317131003765910

f) Advance Online Publication

Smith, J. A. (2010). Citing advance online publication: A review. *Journal of Psychology.* Advance online publication. http://dx.doi.org/10.1037/a45d7867

g) Article in a Magazine

Henry, W. A., III. (1990, April 9). Making the grade in today's schools. *Time, 135,* 28–31.

Doctoral Dissertation, Master's Thesis, Presentation, Proceeding

a) Dissertation/Thesis from a Commercial Database

Van Brunt, D. (1997). *Networked consumer health information systems* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9943436)

b) Dissertation/Thesis from an Institutional Database

Yaylalı-Yıldız, B. (2014). *University campuses as places of potential publicness: Exploring the politicals, social and cultural practices in Ege University* (Doctoral dissertation). Retrieved from Retrieved from: http://library.iyte.edu.tr/tr/hizli-erisim/iyte-tez-portali

c) Dissertation/Thesis from Web

Tonta, Y. A. (1992). *An analysis of search failures in online library catalogs* (Doctoral dissertation, University of California, Berkeley). Retrieved from http://yunus.hacettepe.edu.tr/~tonta/yayinlar /phd/ickapak.html

d) Dissertation/Thesis abstracted in Dissertations Abstracts International

Appelbaum, L. G. (2005). Three studies of human information processing: Texture amplification, motion representation, and figure-ground segregation. *Dissertation Abstracts International: Section B. Sciences and Engineering, 65*(10), 5428.

e) Symposium Contribution

Krinsky-McHale, S. J., Zigman, W. B., & Silverman, W. (2012, August). Are neuropsychiatric symptoms markers of prodromal Alzheimer's disease in adults with Down syndrome? In W. B. Zigman (Chair), *Predictors of mild cognitive impairment, dementia, and mortality in adults with Down syndrome.* Symposium conducted at the meeting of the American Psychological Association, Orlando, FL.

f) Conference Paper Abstract Retrieved Online

Liu, S. (2005, May). *Defending against business crises with the help of intelligent agent based early warning solutions.* Paper presented at the Seventh International Conference on Enterprise Information Systems, Miami, FL. Abstract retrieved from http://www.iceis.org/iceis2005/abstracts_2005.htm

g) Conference Paper - In Regularly Published Proceedings and Retrieved Online

Herculano-Houzel, S., Collins, C. E., Wong, P., Kaas, J. H., & Lent, R. (2008). The basic nonuniformity of the cerebral cortex. *Proceedings of the National Academy of Sciences, 105,* 12593–12598. http://dx.doi.org/10.1073/pnas.0805417105

h) Proceeding in Book Form

Parsons, O. A., Pryzwansky, W. B., Weinstein, D. J., & Wiens, A. N. (1995). Taxonomy for psychology. In J. N. Reich, H. Sands, & A. N. Wiens (Eds.), *Education and training beyond the doctoral degree: Proceedings of the American Psychological Association National Conference on Postdoctoral Education and Training in Psychology* (pp. 45–50). Washington, DC: American Psychological Association.

i) Paper Presentation

Nguyen, C. A. (2012, August). *Humor and deception in advertising: When laughter may not be the best medicine.* Paper presented at the meeting of the American Psychological Association, Orlando, FL.

Other Sources

a) Newspaper Article

Browne, R. (2010, March 21). This brainless patient is no dummy. *Sydney Morning Herald, 45.*

b) Newspaper Article with no Author

New drug appears to sharply cut risk of death from heart failure. (1993, July 15). *The Washington Post,* p. A12.

c) Web Page/Blog Post

Bordwell, D. (2013, June 18). David Koepp: Making the world movie-sized [Web log post]. Retrieved from http://www.davidbordwell.net/blog/page/27/

d) Online Encyclopedia/Dictionary

Ignition. (1989). In *Oxford English online dictionary* (2nd ed.). Retrieved from http://dictionary.oed.com

Marcoux, A. (2008). Business ethics. In E. N. Zalta (Ed.). *The Stanford encyclopedia of philosophy.* Retrieved from http://plato.stanford.edu/entries/ethics-business/

e) Podcast

Dunning, B. (Producer). (2011, January 12). *inFact: Conspiracy theories* [Video podcast]. Retrieved from http://itunes.apple.com/

f) Single Episode in a Television Series

Egan, D. (Writer), & Alexander, J. (Director). (2005). Failure to communicate. [Television series episode]. In D. Shore (Executive producer), *House;* New York, NY: Fox Broadcasting.

g) Music

Fuchs, G. (2004). Light the menorah. On *Eight nights of Hanukkah* [CD]. Brick, NJ: Kid Kosher.

SUBMISSION CHECKLIST

Ensure that the following items are present:

- Cover letter to the editor
  - The category of the manuscript
  - Confirming that "the paper is not under consideration for publication in another journal".
  - Including disclosure of any commercial or financial involvement.
  - Confirming that last control for fluent English was done.
  - Confirming that journal policies detailed in Information for Authors have been reviewed.
  - Confirming that the references cited in the text and listed in the references section are in
  - line with APA 6.
- Copyright Agreement Form
- Permission of previous published material if used in the present manuscript
- Title page
  - The category of the manuscript
  - The title of the manuscript
  - All authors' names and affiliations (institution, faculty/department, city, country),
  - e-mail addresses
  - Corresponding author's email address, full postal address, telephone and fax number
  - ORCIDs of all authors.
- Main Manuscript Document
  - The title of the manuscript
  - Abstract (150-250 words)
  - Key words: 3-5 words
  - Grant support (if exists)
  - Conflict of interest (if exists)
  - Acknowledgement (if exists)
  - References
  - All tables, illustrations (figures) (including title, explanation, captions)

TELİF HAKKI ANLAŞMASI FORMU / COPYRIGHT AGREEMENT FORM

EK 2-A

İstanbul Üniversitesi
Istanbul University

Dergi Adı: Acta INFOLOGICA (ACIN
Journal name: Acta INFOLOGICA (ACIN)

Telif Hakkı Anlaşması Formu
Copyright Agreement Form

Sorumlu Yazar *Responsible/Corresponding Author*	
Makalenin Başlığı *Title of Manuscript*	
Kabul Tarihi *Acceptance Date*	
Yazarların Listesi *List of Authors*	

Sıra No	**Adı-Soyadı** *Name - Surname*	**E-Posta** *E-Mail*	**İmza** *Signature*	**Tarih** *Date*
1				
2				
3				
4				
5				

Makalenin türü (Araştırma makalesi, Derleme, v.b.)
Manuscript Type (Research Article, Review, etc.)

Sorumlu Yazar:
Responsible/Corresponding Author:

Çalıştığı kurum	*University/company/institution*	
Posta adresi	*Address*	
E-posta	*E-mail*	
Telefon no; GSM no	*Phone; mobile phone*	

Yazar(lar) aşağıdaki hususları kabul eder:
Sunulan makalenin yazar(lar)ın orijinal çalışması olduğunu ve intihal yapmadıklarını,
Tüm yazarların bu çalışmaya asli olarak katılmış olduklarını ve bu çalışma için her türlü sorumluluğu aldıklarını,
Tüm yazarların sunulan makalenin son halini gördüklerini ve onayladıklarını,
Makalenin başka bir yerde basılmadığını veya basılmak için sunulmadığını,
Makalede bulunan metnin, şekillerin ve dokümanların diğer şahıslara ait olan Telif Haklarını ihlal etmediğini kabul ve taahhüt ederler.
İSTANBUL ÜNİVERSİTESİ'nin bu fikri eseri, Creative Commons Atıf-GayrıTicari 4.0 Uluslararası (CC BY-NC 4.0) lisansı ile yayınlamasına izin verirler. Creative Commons Atıf-GayrıTicari 4.0 Uluslararası (CC BY-NC 4.0) lisansı, eserin ticari kullanım dışında her boyut ve formatta paylaşılmasına, kopyalanmasına, çoğaltılmasına ve orijinal esere uygun şekilde atıfta bulunmak kaydıyla yeniden düzenleme, dönüştürme ve eserin üzerine inşa etme dâhil adapte edilmesine izin verir.
Yazar(lar)ın veya varsa yazar(lar)ın işvereninin telif dâhil patent hakları, fikri mülkiyet hakları saklıdır.
Ben/Biz, telif hakkı ihlali nedeniyle üçüncü şahıslarca vuku bulacak hak talebi veya açılacak davalarda İSTANBUL ÜNİVERSİTESİ ve Dergi Editörlerinin hiçbir sorumluluğunun olmadığını, tüm sorumluluğun yazarlara ait olduğunu taahhüt ederim/ederiz.
Ayrıca Ben/Biz makalede hiçbir suç unsuru veya kanuna aykırı ifade bulunmadığını, araştırma yapılırken kanuna aykırı herhangi bir malzeme ve yöntem kullanılmadığını taahhüt ederim/ederiz.
Bu Telif Hakkı Anlaşması Formu tüm yazarlar tarafından imzalanmalıdır/onaylanmalıdır. Form farklı kurumlarda bulunan yazarlar tarafından ayrı kopyalar halinde doldurularak sunulabilir. Ancak, tüm imzaların orijinal veya kanıtlanabilir şekilde onaylı olması gerekir.

The author(s) agrees that:
The manuscript submitted is his/her/their own original work and has not been plagiarized from any prior work,
all authors participated in the work in a substantive way and are prepared to take public responsibility for the work,
all authors have seen and approved the manuscript as submitted,
the manuscript has not been published and is not being submitted or considered for publication elsewhere,
the text, illustrations, and any other materials included in the manuscript do not infringe upon any existing copyright or other rights of anyone.
ISTANBUL UNIVERSITY will publish the content under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license that gives permission to copy and redistribute the material in any medium or format other than commercial purposes as well as remix, transform and build upon the material by providing appropriate credit to the original work.
The Contributor(s) or, if applicable the Contributor's Employer, retain(s) all proprietary rights in addition to copyright, patent rights.
I/We indemnify ISTANBUL UNIVERSITY and the Editors of the Journals, and hold them harmless from any loss, expense or damage occasioned by a claim or suit by a third party for copyright infringement, or any suit arising out of any breach of the foregoing warranties as a result of publication of my/our article. I/We also warrant that the article contains no libelous or unlawful statements and does not contain material or instructions that might cause harm or injury.
This Copyright Agreement Form must be signed/ratified by all authors. Separate copies of the form (completed in full) may be submitted by authors located at different institutions; however, all signatures must be original and authenticated.

Sorumlu Yazar; *Responsible/Corresponding Author*	**İmza / Signature**	**Tarih** / *Date*
		……../……../……