

Yıl:2024

Cilt:8

Sayı:2

Year:2024

Vol:8

No:2

UYBİSBBD

ULUSLARARASI YÖNETİM BİLİŞİM SİSTEMLERİ
VE
BİLGİSAYAR BİLİMLERİ DERGİSİ

ULUSLARARASI INTERNATIONAL JOURNAL OF
YÖNETİM MANAGEMENT
BİLİŞİM SİSTEMLERİ INFORMATION SYSTEMS
VE AND
BİLGİSAYAR BİLİMLERİ DERGİSİ COMPUTER SCIENCE

Cilt: 8 • Sayı: 2 • Aralık 2024
Vol: 8 • No: 2 • December 2024

e-ISSN: 2618 - 5954

**ULUSLARARASI YÖNETİM BİLİŞİM SİSTEMLERİ
VE
BİLGİSAYAR BİLİMLERİ DERGİSİ**

**INTERNATIONAL JOURNAL OF MANAGEMENT INFORMATION SYSTEMS
AND
COMPUTER SCIENCE**

Cilt: 8 • Sayı: 2 • Aralık 2024
Vol: 8 • No: 2 • December 2024

e-ISSN: 2618-5954

E-mail : ybsbb.info@gmail.com

Web : dergipark.gov.tr/uybisbbd

UYBİSBBD'in tarandığı İndeksler (*The indexes that UYBİSBBD is scanned*):

ERIHPLUS, Crossref, BASE, Cite Factor, Index Copernicus, Academic Resource Index
ResearchBib, Scientific Indexing Services, Cosmos Impact Factor, WorldCat, ESJI (Eurasian
Scientific Journal Index), Rootindexing, Google Scholar, SOBIAD

UYBİSBBD, uluslararası hakemli, uluslararası indeksli, açık erişimli bilimsel bir dergidir.
UYBİSBBD is an international peer-reviewed, internationally indexed, open-access scientific journal.



**ULUSLARARASI YÖNETİM BİLİŞİM SİSTEMLERİ VE BİLGİSAYAR BİLİMLERİ
DERGİSİ**
**INTERNATIONAL JOURNAL OF MANAGEMENT INFORMATION SYSTEMS
AND COMPUTER SCIENCE**

Kurucu (Founder)

Doç. Dr. Adem KORKMAZ

Baş Editör (Editor-in-Chief)

Dr. Öğr. Üyesi Selma BULUT

Editörler (Editors)

Prof. Dr. Aysun COŞKUN

Doç. Dr. Tarık TALAN

Doç. Dr. Selahattin KOŞUNALP

Doç. Dr. Ayşe ÇİÇEK KORKMAZ

Doç. Dr. Adem KORKMAZ

Yayın Kurulu (Editorial Board)

Prof. Dr. Florentin SMARANDACHE	(University of New Mexico, USA)
Prof. Dr. Aysun COŞKUN	(Gazi Üniversitesi)
Doç. Dr. Selahattin KOŞUNALP	(Bandırma Onyediy Eylül Üniversitesi)
Doç. Dr. Tarık TALAN	(Gaziantep İslam Bilim ve Teknoloji Üniversitesi)
Doç. Dr. Adem KORKMAZ	(Bandırma Onyediy Eylül Üniversitesi)
Dr. Öğr. Üyesi Mustafa Mikail ÖZÇİLOĞLU	(Kilis 7 Aralık Üniversitesi)
Doç. Dr. Ayşe ÇİÇEK KORKMAZ	(Bandırma Onyediy Eylül Üniversitesi)
Dr. Bogdan PATRUT	(Alexandru Ioan Cuza University of Iasi, Romania)
Dr. Iulian FURDU	(Vasile Alecsandri University of Bacau, Romania)
Dr. Sadiq HUSSAIN	(Dibrugarh University, India)
Dr. Svitlana ILNYTSKA	(National Aviation University, Ukraine)

Danışma Kurulu (Advisory Board)

Prof. Dr. Abdulkadir YILDIZ	(Kahramanmaraş Sütçü İmam Üniversitesi)
Prof. Dr. Aysun COŞKUN	(Gazi Üniversitesi)
Prof. Dr. Erdem UÇAR	(Trakya Üniversitesi)
Prof. Dr. Florentin SMARANDACHE	(University of New Mexico)
Prof. Dr. H. Mustafa PAKSOY	(Gaziantep Üniversitesi)
Prof. Dr. İsmail Rakıp KARAŞ	(Karabük Üniversitesi)
Prof. Dr. Sadettin PAKSOY	(Gaziantep Üniversitesi)
Prof. Dr. Sevinç GÜLSEÇEN	(İstanbul Üniversitesi)
Prof. Dr. Ülkü BAYKAL	(İstanbul Üniversitesi)
Prof. Dr. Yılmaz KILIÇASLAN	(Adnan Menderes Üniversitesi)
Prof. Dr. Mustafa ŞEKKELİ	(Kahramanmaraş Sütçü İmam Üniversitesi)
Prof. Dr. Yusuf Ekrem AKBAŞ	(Adıyaman Üniversitesi)
Doç. Dr. Ercan BULUŞ	(Tekirdağ Namık Kemal Üniversitesi)
Doç. Dr. Erdinç UZUN	(Tekirdağ Namık Kemal Üniversitesi)
Doç. Dr. İlhan UMUT	(Trakya Üniversitesi)

Adres (Address)

Bandırma Onyediy Eylül Üniversitesi, Gönen Meslek Yüksekokulu
10900 Balıkesir / TÜRKİYE

E-mail : ybsbb.info@gmail.com

Web : <https://dergipark.org.tr/tr/pub/uybisbbd>

YAYIN POLİTİKASI

Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi yılda iki kez Haziran ve Aralık aylarında yayınlanan uluslararası hakemli bir dergidir. Dergide yer alan yazılar kaynak gösterilmeksizin kısmen ya da tamamen iktibas edilemez. Bu dergide yayınlanan çalışmaların bilim ve dil sorumluluğu yazarlarına aittir.

Dergimize gönderilen çalışmalar, alanında uzman iki ayrı hakem tarafından incelendikten sonra uygun görülenler yayınlanmaktadır. Yazım kurallarına ilişkin bilgilere dergimizin web adresinde yer verilmiştir. Bu derginin tüm hakları saklıdır. Önceden yazılı izin almaksızın hiçbir iletişim ve kopyalama sistemi kullanılarak yeniden kopyalanamaz, çoğaltılamaz ve satılamaz.

International Journal of Management Information Systems and Computer Science is an international peer-reviewed journal which is published two times a year in June and December. The articles cannot be cited partly or entirely without showing resources. The responsibility about scientific and grammatical issues is belong to authors.

The papers sent to the journal are reviewed by two referees and after their approval, they will be sent to edit before being published. Writing & Publishing Policies can be found in the journal's website. All rights reserved. No part of this publication may be reproduced, stored or introduced into a retrieval system without prior written permission.

Makaleler / Articles

Makine Öğrenimi Algoritması Kullanarak Kişisel Göstergelere Dayalı Çalışan Teşviklerinin Tahmini

Forecasting Employees' Promotion Based on Personal Indicators by Using a Machine Learning Algorithm

Makale Türü: Araştırma Makalesi / Paper Type: Research Paper
Yasmine Aya IBRIR & Mahmut ÇAVUR 75-98

Açık Anahtar Altyapısı ile Dijital İmzalamanın Zararlı Yazılımlar Üzerindeki Etkisi

Impact of Digital Signing on Malware in Public Key Infrastructure

Makale Türü: Araştırma Makalesi / Paper Type: Research Paper
Mehmetcan TOPAL & Zeynep ALTAN 99-109

Android Güvenlik Açıklarının Modellenmesi: İstatistiksel Dağılımlardan Analizler

Modeling Android Security Vulnerabilities: Insights from Statistical Distributions

Makale Türü: Araştırma Makalesi / Paper Type: Research Paper
Kerem GENCER & Fatih BASCİFTCİ 110-126

Süper Yapay Zeka Devrimleriyle Yönetim Bilişim Sistemlerinin Evrimi

Evolution of Management Information Systems by Super Artificial Intelligence Revolutions

Makale Türü: Araştırma Makalesi / Paper Type: Research Paper
Ahmet EFE 127-142

Oyun Geliştirme Konulu Lisansüstü Tezlerin Bibliyometrik Analizi

Bibliometric Analysis of Postgraduate Theses on Game Development

Makale Türü: Derleme Makalesi / Paper Type: Review Paper
Barancan UZUN, Emel GÜVEN & Tamer EREN 143-164

Makine Öğrenimi Algoritması Kullanarak Kişisel Göstergelere Dayalı Çalışan Teşviklerinin Tahmini

Forecasting Employees' Promotion Based on Personal Indicators by Using a Machine Learning Algorithm

Yasmine Aya IBRIR¹ 

Mahmut ÇAVUR² 

DOI:10.33461/uybisbbd.1471499

Öz

Makale Bilgileri

Makale Türü:
Araştırma Makalesi

Geliş Tarihi:
21.04.2024

Kabul Tarihi:
18.07.2024

©2024 UYBISBBD
Tüm hakları saklıdır.



Terfi, çalışanın kendini geliştirmesi ve işin yükünü ve sorumluluğunu, kendisine yüklenen pozisyonla birlikte taşıma isteği için motive etmenin bir aracı olarak hareket eder. Geleneksel yöntemler ile yapılan terfilerin hakkaniyeti ve ölçülebilirliği nicel olarak ölçülemediği için farklı yöntemlere ihtiyaç duyulmaktadır. Son yıllarda şirketlerde bilgi sistemlerinin kullanımın yaygınlaşması ile çalışanlara ait performans bilgileri gibi birçok bilgi dijital ortamda tutulmaya başlandı. Yine ver bilimlerinin gelişmesi ve birçok alana uygulanması ile birlikte çalışanlara ait bu verilerin değerlendirmesinde makine öğrenmesi ve yapay zekâ algoritmalarının kullanımı yaygınlaştı. Bu çalışma, çeşitli özelliklere dayalı olarak bir kuruluş içindeki çalışanların terfilerini tahmin etmek için sağlam bir çerçeve oluşturmayı amaçlamaktadır. Bu özellikler, eğitim sayısını, önceki yıl derecelendirmelerini, hizmet süresini, kazanılan ödülleri ve ortalama eğitim puanını içermekle birlikte bunlarla sınırlı değildir. Çalışmanın amacı, kuruluşların bilinçli terfi kararları almaları için güvenli bir araç sağlamak ve bu çerçevenin diğer tahmin problemlerine genelleştirilebileceğini göstermektir. Deneysel sonuçlar XGBoost modelinin doğruluk açısından en verimli model olduğunu göstermektedir. XGBoost, %94 doğruluk ve ROC AUC, %94 duyarlılık ve %94 hassasiyetle bellek kullanımı verimliliği, doğruluk ve çalışma süresi açısından üstün bir algoritma olarak kabul edilmektedir.

Anahtar Kelimeler: Çalışan Terfisi, Çalışan Terfi Tahmin Çerçevesi, XGBoost, Makine Öğrenimi, Denetimli Öğrenme.

Abstract

Article Info

Paper Type:
Research Paper

Received:
21.04.2024

Accepted:
18.07.2024

©2024 UYBISBBD
All rights reserved.



Promotion is a tool to motivate employees to improve themselves and take on the burden and responsibility of the position assigned to them. Due to the fairness and measurability of promotions conducted by traditional methods needing to be quantifiable, different methods are required. In recent years, with the widespread use of information systems in companies, much information, such as performance data of employees, has started to be stored digitally. Additionally, with the development of data sciences and their application in many fields, machine learning and artificial intelligence algorithms in evaluating this data have become widespread. This study aims to establish a robust framework to predict employee promotions based on various features. These features include but are not limited to the number of training sessions attended, previous year ratings, tenure, awards received, and average training scores. The study aims to provide organizations with a reliable tool to make informed promotion decisions and demonstrate that this framework can be generalized to other prediction problems. Experimental results show that the XGBoost model is the most efficient in terms of accuracy. XGBoost is considered a superior algorithm with 94% accuracy, 94% ROC AUC, 94% sensitivity, and 94% precision, excelling in memory usage efficiency, accuracy, and runtime.

Keywords: Employee Promotion, Prediction, XGBoost, Machine Learning, Supervised Learning.

Atf/ to Cite (APA): Ibrir, Y.A. and Çavur, M. (2024). Forecasting Employees' Promotion Based on Personal Indicators by Using a Machine Learning Algorithm. International Journal of Management Information Systems and Computer Science, 8(2), 75-98. DOI: 10.33461/uybisbbd.1471499

¹ Kadir Has University, Management Information System Department, yasmineaya.ibrir@stu.khas.edu.tr, İstanbul, Türkiye.

² Dr. Kadir Has University, Management Information System Department, mahmut.cavur@khas.edu.tr, İstanbul, Türkiye.

1. INTRODUCTION

Promotion is regarded as one of the most important issues in any organization because it is essential for administrative development and a means of motivating employees to pursue self-development. Promotion has always been an important research point in several areas, including human development. Nowadays, many organizations need help with job promotion and professional stability. The ability of institutions to achieve their goals depends largely on the extent to which the administration succeeds in providing sufficient satisfaction and setting up a promotion program according to objective criteria that permit achieving organizational effectiveness. Businesses must be able to forecast what will happen to their client base and staff so that they can take appropriate actions before the "promotion" process. The term "promotion" is defined as a means of an employee's career advancement and development and is linked to the employee's level of performance.

The issue that this study attempts to solve is the difficulty organizations have in accurately forecasting and overseeing employee promotions. Promotions may cause unhappiness and inefficiency without an impartial and reliable prediction framework, which lowers employee morale and corporate performance. Therefore, we clarify how machine learning (ML) models enhance the ability of enterprises to forecast employee promotions. Not only the type of ML but also the parameters are critical to predicting the promotions. Therefore, defining, deciding and explaining those parameters affecting the promotions is essential. Finally, choosing, adopting, and/or developing ML algorithms for promotion prediction is necessary. Compared to conventional techniques, machine learning models may greatly improve the accuracy of employee promotion forecasts. Reliability, years of service, practical efficiency, and credentials are important indicators of when an individual will be promoted. The models' forecast accuracy will increase with new criteria like total score, work percentage, and years to retirement. Since there can be one or more machine learning models that perform better in terms of prediction accuracy and reliability than the baseline models among the numerous models, we selected and adopted several ML algorithms to compare their performances with several metrics.

In this thesis study, we aimed to set up a robust framework that can be used and generalized to predict problems in the business, not just the problem of predicting employee promotion. Employees are promoted based on their practical efficiency and loyalty in performing their jobs, as well as their years of service and qualifications. The essential idea is to promote the right man to the right place, thus being the path to success for the company. We believe that our framework can be successfully used in choosing the appropriate employee according to the organizational hierarchy without fraud or prior knowledge. The primary goal of the learning models is to anticipate the individual's promotion within a particular period reliably. We will reward workers based on their performance and workplace behavior, utilizing these aspects from the HR Analytics Vidhya data. We create a framework for predicting employee promotion using machine learning algorithms. We chose existing features and added new several features, including total score, work fraction, work start year, years remaining to retire, and performance. We compare the experimental results with different baseline models using evaluation performance metrics.

2. RELATED WORKS

This section outlines the related research on predicting employee advancement and turnover.

2.1. Predicting Employee Promotion

Employee promotion refers to an employee's upward progress within the organization to a new or higher job position, tasks, and responsibilities. Promotion is an important step in the life cycle of both employees and organizations. Choosing the correct candidate for promotion at the right moment is a critical issue. According to many studies, several strategies rely on machine learning to overcome

real-world challenges, particularly in human resource management, and employee promotion is one of them.

Managers spend a great deal of time recruiting capable employees. Promotion is an issue that both businesses and employees are concerned about. On the one hand, promotion is a strategy used by businesses to pick exceptional people and boost their competitiveness. Employee promotion, systems, and organizational performance have a good relationship (Chen, Hsu and Wu, 2012). On the other hand, advancement prospects have a higher impact on employee performance, as do leadership, job promotion, and work environment. These components work together to improve employee performance (Febrina, 2017). It is regarded as an excellent policy to replace gaps in higher-level positions through internal promotions since such advancements give encouragement and motivation to employees while also removing sentiments of stagnation and discontent (Li et al., 2021).

According to certain research, internal promotion in organizations is influenced by a variety of factors, including age (Li et al., 2021; Long et al., 2018; Machado and Portela, 2021) gender, education background (Jantan and Hamdan, 2010; Long et al., 2018), and job experience (De Pater et al., 2009; Long et al., 2018). Categorization is one of the most important tasks in data mining, which is used to extract knowledge from massive amounts of data. This technique is frequently utilized in a variety of sectors, although it has received less attention in HRM. Using an employee's performance data, an experiment was carried out to illustrate the practicality of recommended classification techniques (Jantan and Hamdan, 2010).

Higher compensation is always followed by a job promotion and increased experience in problem-solving, loyalty, honesty, and responsibility at work, according to Febrina's (2017) research. Based on his findings, it is possible to infer that leadership, job advancement, and job environment all positively and substantially influence staff performance at the bank. These components work together to improve employee performance (Febrina, 2017).

Businesses must prioritize human capital in the age of big data and Industry 4.0. Liu and colleagues (2019) emphasize that people should work in various places and divisions to broaden their experiences. Working in jobs where mobility is more reliable, resources are available, or experience is accessible can help the worker advance. Their study used supervised learning to predict staff advancement and build models using logistic regression, random forests, and AdaBoost. In the end, RF performs better and has a reasonable time consumption (Liu et al., 2019).

Long and colleagues (2018) have used machine learning to predict employee advancement using data from a Chinese state-owned firm. Two types of features were created based on five methodologies by extracting personal basic information and position information from this data. Using correlation analysis, they validate the efficiency of attributes in estimating employee advancement. According to the findings of the study, the influence of post features on promotion is stronger than that of personal basic characteristics (Long et al., 2018).

Job classifications are frequently established using the k-means clustering technique, which is a common method of job classification establishment. Sarker (2018) and colleagues used a decision tree algorithm to swiftly identify employees and make suitable decisions. Employees are divided into three groups based on their level of performance. According to the results, support vector machines outperform the other classifiers in terms of accuracy (Sarker et al., 2018).

Although substantial progress has been made using big data analytic technologies in human resource management, research on the mining of promotion characteristics is limited, and further research is needed. Thus, using data from Analytics Vidhya, we build various promotion attributes and predict using machine learning methods.

2.2. Predicting Employee Turnover

Employee turnover is seen as a critical issue for all firms. To address this issue, organizations are now relying on machine learning approaches to forecast employee turnover. Employee turnover

can be viewed as a defacement of the organization's intellectual capital. The literature study focuses on the strategies and techniques provided by various researchers for forecasting employee attrition.

A team of researchers has proposed a new model for forecasting employee attrition based on machine learning using XGBoost. It is recognized as a superior algorithm in terms of memory use efficiency, accuracy, and running time. The model provided in this research has a very low rate of less than 30% and an accuracy of around 90%. A total of 14 factors have a greater effect on the attrition rate than any other component - frozen Promotions and Salary Hikes, Imbalance of Work-Life, Employee Misalignment, Unsuitable Behavior, and Inadequate Professional Skills (Jain and Nayyar, 2018).

In this study¹, various machine learning techniques have been implemented DT, RF, and SVM, it is possible to infer that RF outperforms. In the case of employee attrition, an estimate was made as to whether a person would quit the organization. Using this approach, the business may choose the individuals who have the highest likelihood of leaving the organization and then provide them with specific incentives (Jain, Jain and Pamula, 2020).

In another study, a strategy for selecting features to reduce the dimension of the feature space was described. The recommended feature selection improves the predictor's performance. This paper offers a three-stage method for constructing an accurate employee attrition prediction model. The parameters of the logistic regression model are validated by assessing their fluctuations when trained through several bootstraps. The results suggest that the "max-out" feature selection strategy improves the F1-score performance metric (Najafi-Zangeneh et al., 2021).

A research study insists that employee turnover in the IT industry is significant. Often, their early attrition is the result of company-related or personal concerns. A correlation matrix in the form of a heatmap was developed to determine the essential factors that may affect the attrition rate. The Random Forest classifier was revealed to be the best model for predicting IT staff attrition (Bandyopadhyay and Jadhav, 2021).

Employee promotions are a crucial component of keeping a motivated and skilled workforce. Prior to the "promotion" process, firms must be able to predict what will happen to their client base and workforce. With machine learning, we may identify employees who are most likely to be promoted based on prior data such as degree, experience, age, ratings, and overall score, using primarily the XGBoost model and other models. This framework can be used and generalized to all prediction problems, not just our problem of predicting employee promotion.

We believe that, without fraud or prior information, our approach may be effectively utilized to select the proper individual according to the organizational hierarchy. We consider several elements, including traditional employee traits and work history. Favoritism and family must be excluded from these accounts to prevent the problem from being passed on to someone else.

3. MATERIAL AND METHOD

3.1. Data Description

Analytics Vidhya Data Analysis provided the data (Table 1) used in this study. The dataset has 14 characteristics 54808 records for train data and 23490 records for test data. Not all the 14 features are considered in our work for employees' predictions of promotions. We will choose the relevant aspects and add new ones that impact the employee's promotion as an important indication for the promotion process. Data exploration is a two-step process that involves identifying the data type and category of the variables used to make up an empirical data set.

The dataset contains a target feature identified by the variable that is promoted. "No" denotes an employee who did not receive the promotion, and "Yes" represents an employee who did. If this

training process is repeated over time and conducted on relevant samples, the predictions generated in the output will be more accurate.

In this paper, we explore the Analytics Vidhya dataset step-by-step. The methodologies of variable identification, univariate analysis, bivariate analysis, and multivariable analysis were applied to the Data Exploration phase of the data processing process for the first time.

Table 1: Attributes description and identification

Attributes	Description	Data Type	Variable Category	Type of Variable
Employee ID	Unique ID for the employee	Numeric	Continuous	Predictor
Department	Department of the employee	Character	Categorical	Predictor
Region	Region of employment (unordered)	Character	Categorical	Predictor
Education	Educational level	Character	Categorical	Predictor
Gender	Gender of the employee	Character	Categorical	Predictor
Recruitment channel	Channel of recruitment for the employee	Character	Categorical	Predictor
No. of training	No of other trainings were completed in the previous year on soft skills, technical skills, etc.	Numeric	Categorical	Predictor
Age	Age of the employee	Numeric	Continuous	Predictor
Previous year rating	Employee rating in the prior year	Numeric	Categorical	Predictor
Length of service	Length of service in years	Numeric	Continuous	Predictor
Awards won	If awards were won during the previous year, then 1 else 0	Numeric	Categorical	Predictor
Avg training score	The average score in current training evaluations	Numeric	Continuous	Predictor
KPIs met >80%	If the percentage of KPIs (Key Performance Indicators) >80%, then 1, else 0	Numeric	Categorical	Predictor
Is promoted	(Target) Recommended for promotion	Numeric	Categorical	Target variable

3.2. Descriptive Statistics

3.2.1. Uni-variate analysis

Univariate analysis is the most basic type of statistical analysis. Continuous and categorical variables are investigated in the univariate analysis. We investigated various techniques and statistical measures for categorical and continuous variables individually. Several statistical metrics and visualization approaches describe this type of relationship.

We constructed the descriptive statistics of the dataset. We considered the following variables: count, mean, standard deviation (std), minimum and maximum values (min/max), and 25%/50%/75% 95% percentile. **Hata! Başvuru kaynağı bulunamadı.** is an excerpt from the full dataset.

Count

The count of a dataset is simply the number of observations, denoted as n .

$$Count = n \quad (1)$$

Mean

A dataset's mean (average) is the sum of all observations divided by the number of observations.

$$\text{Mean } (\mu) = \frac{1}{n} \sum_{i=1}^n x_i \quad (2)$$

where x_i represents each observation in the dataset.

Standard Deviation (Std)

The standard deviation measures the amount of variation or dispersion in a dataset.

$$\begin{aligned} \text{Standard Deviation } (\sigma) & \quad (3) \\ & = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2} \end{aligned}$$

where μ is the mean of the dataset.

Minimum and Maximum Values (Min/Max)

The minimum value is the smallest observation in the dataset, and the maximum value is the largest observation in the dataset.

$$\text{Minimum} = \min(x_1, x_2, \dots, x_n) \quad (4)$$

$$\text{Maximum} = \max(x_1, x_2, \dots, x_n) \quad (5)$$

Percentiles (25%, 50%, 75%, 95%)

Percentiles are values below which a certain percentage of observations in a dataset fall.

- 25th Percentile (First Quartile, Q1): The value below 25% of the observations falls.

$$Q1 = P_{25} \quad (6)$$

- 50th Percentile (Median, Q2): The value below 50% of the observations falls.

$$Q2 = P_{50} \quad (7)$$

- 75th Percentile (Third Quartile, Q3): The value below 75% of the observations falls.

$$Q3 = P_{75} \quad (8)$$

Figure 1 represents the bar chart of the variables; these are the value ranges of all features. Every feature has a different distribution of values. We looked at each one independently to get a better and deeper understanding of the characteristics.

Using machine learning models with imbalanced classes often leads to very poor results that are completely biased towards the class having a higher distribution. Clearly, the data needs to be balanced; a 91% and 9% ratio between promoted and non-promoted employees is very unbalanced.

Figure 1: Numeric features distribution

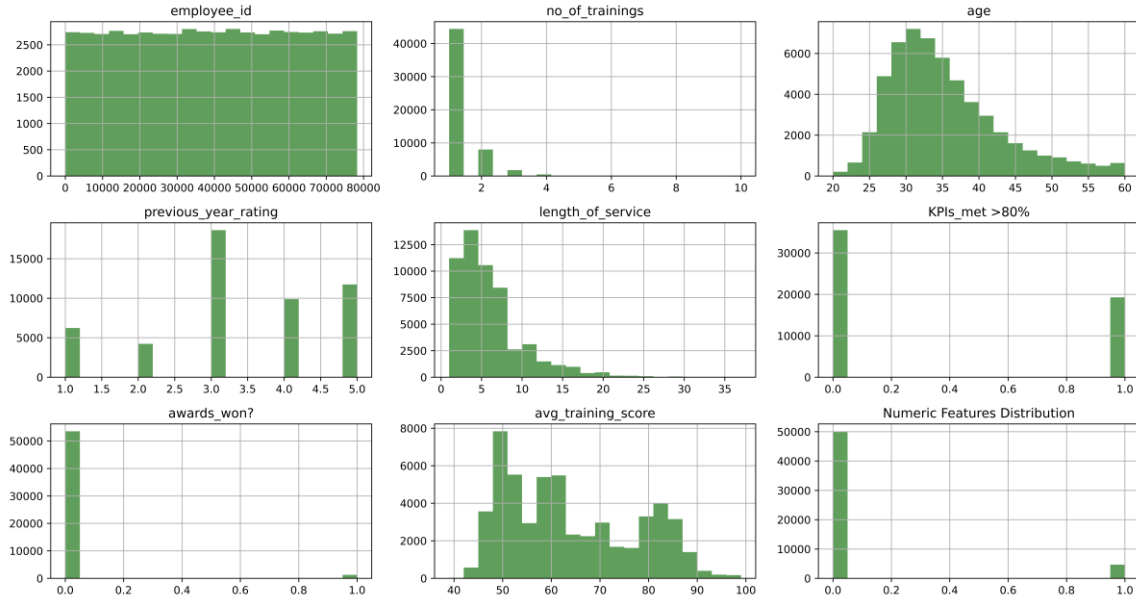
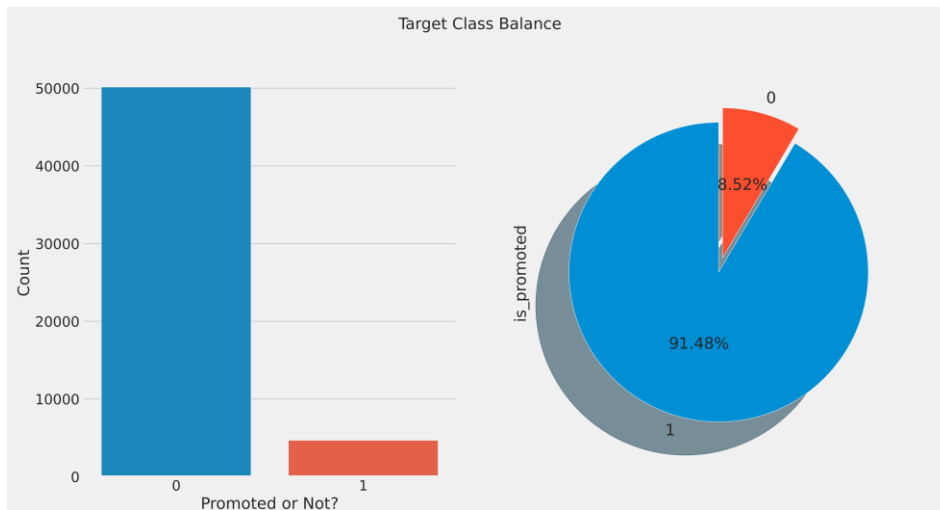


Figure 2: Is Promoted Distribution



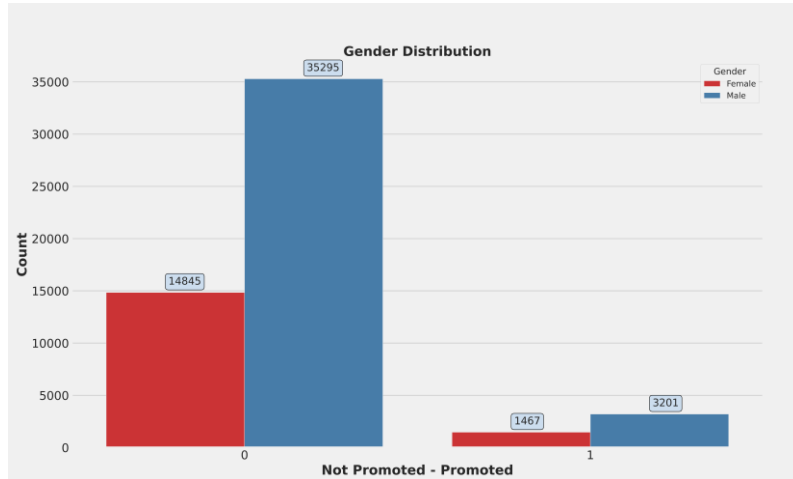
3.2.2. Bi-variate analysis

In this section, we examine variables at a predetermined significance threshold. Bivariate analysis may be used for any grouping of absolute, categorical, and continuous variables. During the analytic process, many strategies are utilized to manage these groups. The following are some examples of the specific combinations that are possible.

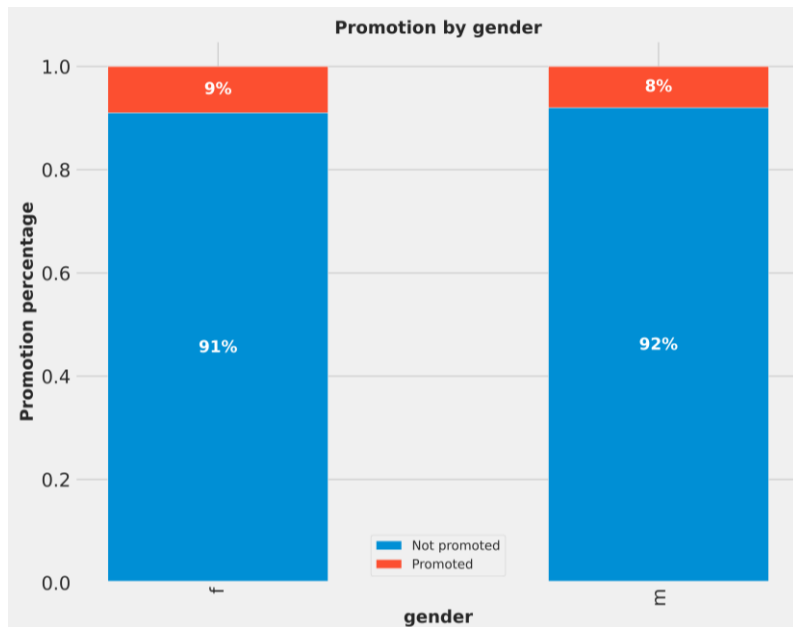
- **Gender versus Employee Promoted**

According to **Hata! Başvuru kaynağı bulunamadı.**, male employees are promoted at a higher rate than female employees. Furthermore, male employees continue to be promoted more than female ones. As previously stated, women are in the minority, but when it comes to promotion, they compete head-to-head with their male counterparts.

Figure 3: Gender versus Employee Promoted a) in number, b) in percentage



a)

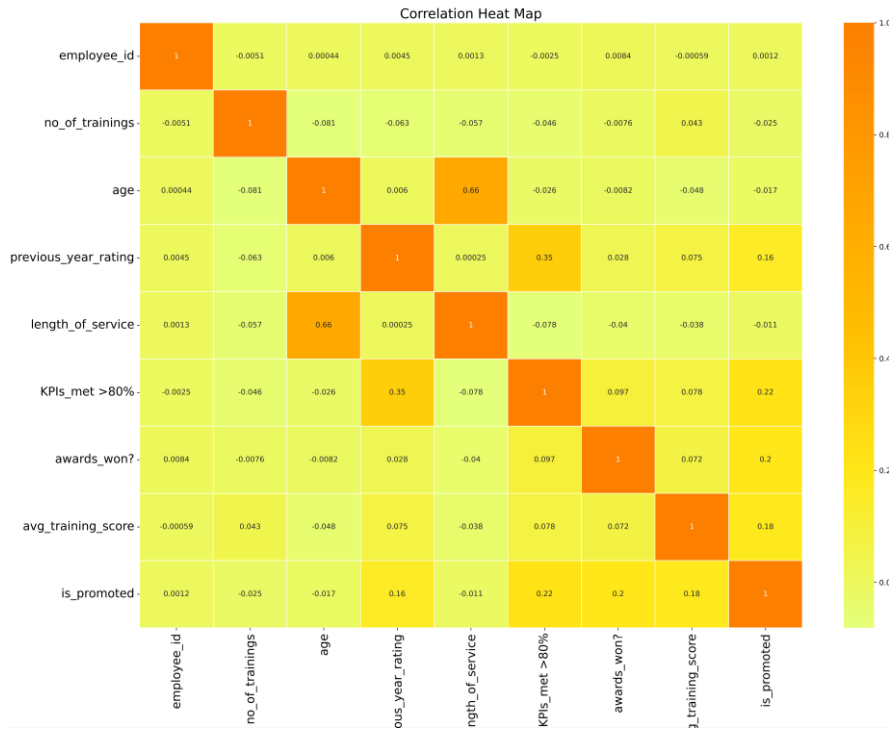


b)

3.2.3. Multivariate analysis

Multivariate analysis is based on multivariate statistics concepts, simultaneously involving observing and analyzing several statistical result variables. First, we will examine the association between the numerical columns using the Correlation Heatmap.

Figure 4: Correlation Heat map



KPIs and the previous year's ratings are correlated to some degree, implying some relationship. However, we will do feature engineering before modeling to avoid multicollinearity. This heatmap shows the correlation between the columns, which is highly beneficial for regression issues.

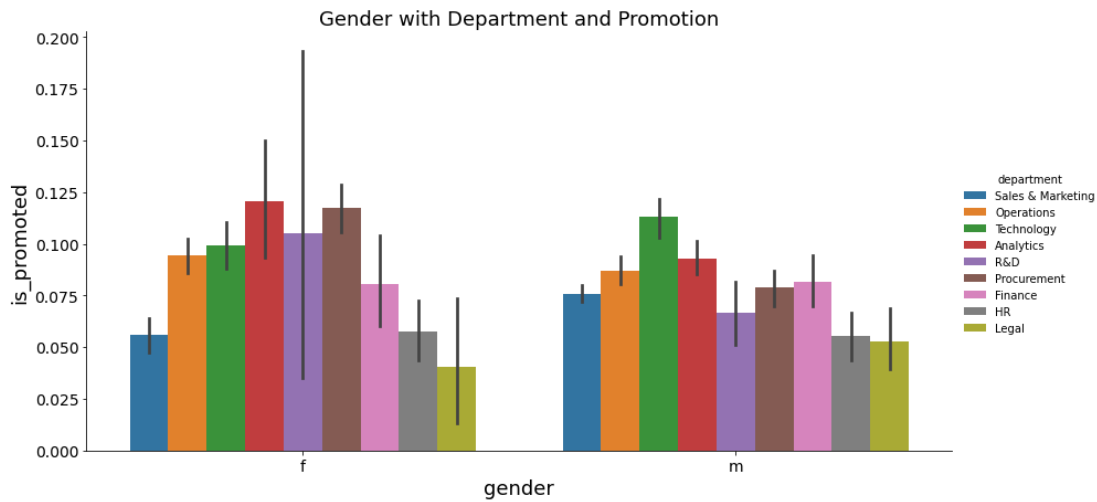
3.3. Data Visualizations

In this part, data visualization is conducted on continuous and categorical variables to understand the link between these variables and our objective variables. Graphs are the most effective way to comprehend the behavior of characteristics and the relationships between them. Here is one example:

- **Gender with Department and Promotion**

Hata! Başvuru kaynağı bulunamadı. shows the gender breakdown by department and promotion. This figure represents the distribution of females and males in the department section. There are varying percentages in the departments, and for females, there are higher promotions in the two departments of analytics and procurement. Unlike males, technology and analytics are the two sections with the highest percentages.

Figure 5: Gender with Department and Promotion



3.4. Machine Learning Algorithms

Machine learning is a branch of computer science with significant links to statistics and optimization. In this thesis, we use supervised learning approaches for binary classification. The experiment is carried out within the Scikit-learn library, and the code is written in Python using supervised algorithms.

3.4.1. XGBoost

XGBoost is a boosted tree approach based on the gradient boosting principle. It employs more precise approximations by applying second-order gradients and enhanced regularization. It is a fast approach based on parallel tree creation that is designed to be fault-resistant in a distributed situation (Jain and Nayyar, 2018). The classifier accepts data in the form of DMatrix (Saradhi and Palshikar, 2011). During the research, the following characteristics were investigated and incorporated(9):

$$\text{Obj}(\theta) = \sum_{i=1}^n L(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (9)$$

where:

- θ represents the parameters of the model.
- L is the loss function (e.g., mean squared error for regression, log loss for classification).
- y_i is the true value of the i -th instance.
- \hat{y}_i is the predicted value of the i -th instance.
- $\Omega(f_k)$ is the regularization term for the k -th tree.
- n is the number of instances.
- K is the number of trees

Regularization: This is the primary advantage of XGBoost. It also aids in reducing overfitting. This technique is used in linear and tree-based models to prevent overfitting (Aarshay, 2020).

Parallel Processing: XGBoost uses parallel processing and is much quicker than GBM. XGBoost now supports the Hadoop implementation (Aarshay, 2020).

High Flexibility: Custom optimization targets and assessment criteria can be defined by users in XGBoost. This gives the model a new dimension, and there are no restrictions on what we may accomplish (Aarshay, 2020).

XGBoost has a procedure for dealing with missing values. The user must offer a value that differs from the other observations and pass it as a parameter. It tries different things and learns which path to follow for missing values in the future (Aarshay, 2020).

Tree Pruning: When a GBM encounters a negative loss in the split, it will cease dividing the node. On the other hand, XGBoost divides up to the max depth set before pruning the tree backwards and removing splits beyond which no positive benefit is obtained (Aarshay, 2020).

Cross-validation is supported by XGBoost at each iteration of the boosting process, making it straightforward to acquire the precise optimal number of boosting rounds in a single run. In contrast to GBM, we must execute a grid search, and only a restricted number of parameters may be examined (Aarshay, 2020).

Sklearn allows users to begin training an XGBoost model from the previous run's last iteration - this can be a substantial benefit in some situations. This capability is also available in the GBM implementation of sklearn, so they are on the same page (Aarshay, 2020).

3.4.2. Random Forest (RF)

Random Forest is a classifier that uses several decision trees on different subgroups of a given dataset and averages them to enhance the predicted accuracy. The larger the number of trees in the forest, the higher the accuracy and the lower the risk of overfitting (Jaiswal, 2022).

where p_i is the proportion of instances of class i at node t .

$$\text{Gini}(t) = 1 - \sum_{i=1}^c p_i^2 \quad (10)$$

3.4.3. Decision Tree (DT)

A decision tree chart may help us examine alternatives and their outcomes before committing to a solution. It provides a stylized universe where we may play out a sequence of actions and see where they go without devoting unnecessary real-world time and resources (Jaiswal, 2022).

where p_i is the proportion of instances of class i at node t .

$$\text{Gini}(t) = 1 - \sum_{i=1}^c p_i^2 \quad (11)$$

3.4.4. Logistic Regression (LR)

Logistic Regression is an important machine learning technique because it can offer probabilities and classify new data using continuous and discrete datasets. It seeks to calculate the likelihood that the output variable belongs to a certain class. Logistic Regression may be used to categorize observations using many forms of data and can quickly discover the most efficient factors for classification (Jaiswal, 2022).

The sigmoid function transforms the linear combination of inputs into a probability value between 0 and 1.

$$\sigma(\mathbf{z}) = \frac{1}{1 + e^{-z}} \quad (12)$$

where:

- $\sigma(z)$ is the sigmoid function.
- z is the linear combination of inputs.

3.4.5. AdaBoost

AdaBoost is a classifier that uses ensemble boosting to improve classifier accuracy. The AdaBoost classifier creates a strong classifier by merging numerous low-performing classifiers. AdaBoost's main assumption is to build classifier weights and train the data sample in each iteration (Navlani, 2022).

$$W_i(1) = \frac{1}{m} \quad (13)$$

where:

- $w_i^{(1)}$ is the initial weight for the i -th instance.
- m is the total number of training instances.

3.4.6. Gradient Boosting

One of the most successful machine learning algorithms is the gradient boosting strategy. Gradient Boosting's distinguishing feature is that it instead fits a new predictor to the residual errors created by the preceding prediction (Tarbani, 2021), rather than fitting an algorithm to the data at each iteration.

$$F(x) = \operatorname{argmin}_{\gamma} \sum_{i=1}^m L(y_i, \gamma) \quad (14)$$

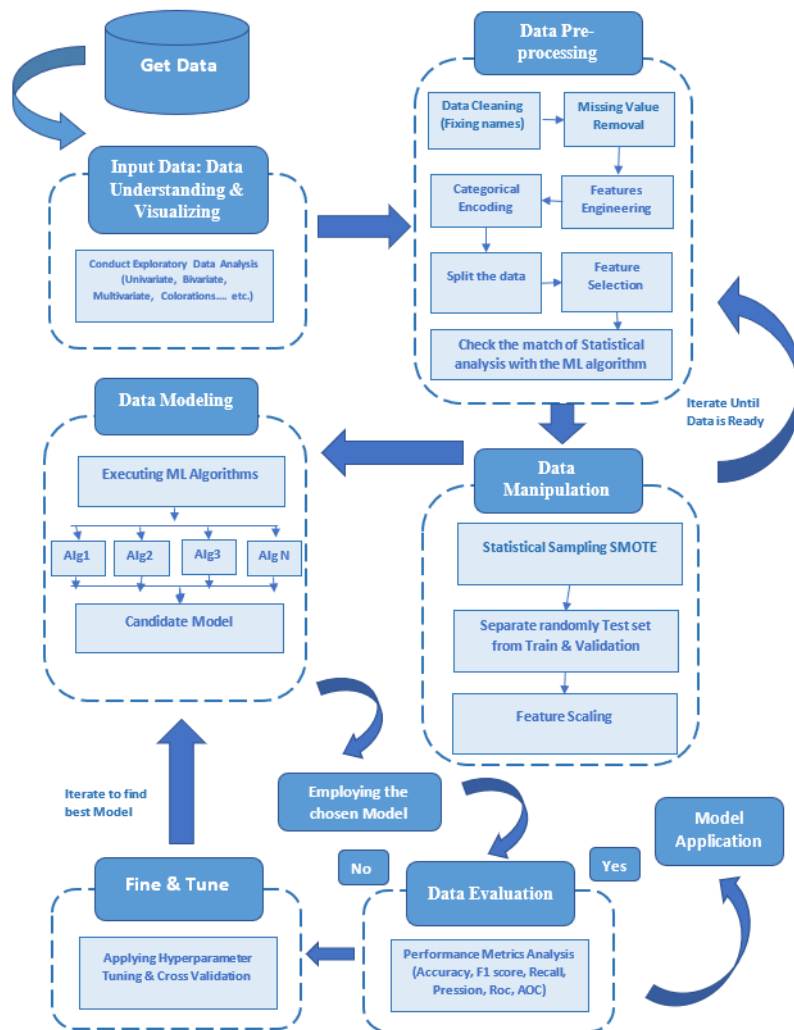
where:

- $F_0(x)$ is the initial prediction.
- $L(y_i, \gamma)$ is the loss function for the i -th instance.
- y_i is the true label of the i -th instance.

3.5. Methodology

Our methodology is mainly composed of five phases: Input Data (which includes Data Understanding & Visualizing), Data Pre-processing (which provides for Data Cleaning, Data Preparing, & Data Splitting), Data Manipulation (which provides for Data Preprocessing and Manipulation), Data Modeling, and Data Evaluation (Fine & Tune). **Hata! Başvuru kaynağı bulunamadı.** depicts a high-level overview. The following sections go through the specifics of each step. At the outset, the dataset will be described.

Figure 6: The general structure of the proposed employee promotion prediction framework



To give some brief explanations about the implementation of our study, algorithms are used through the Scikit-learn library, and the experiment is carried out within Python. These phases are described as follows:

Input Data: Data Understanding & Visualizing

We use publicly available data provided by Analytics Vidhya Data Analysis. This dataset has 14 characteristics 54808 records for train data and 23490 for test data. We will choose the relevant aspects and add new ones that impact the employees’ promotion as an important indication of the promotion process. More details of the dataset were explained in the Background section.

Data Pre-processing: Data Cleaning, Data Preparing & Data Splitting

The data preparation stage is critical for our investigation to acquire clean and usable data. There were instances in the raw data that were not appropriate. This was due to mistakes and abnormalities that had to be removed. Data cleaning and filling-in of missing values in the dataset were conducted. Analysis of the dataset is critical for our investigation to acquire clean and usable data. Treating missing values is important in any machine learning model's creation. Various reasons, such as incomplete forms, unavailable values, data entry errors, and data loss, can cause missing values. We do not have to delete any missing values; we can impute the values using mean, median, and mode.

Extracting features from raw data using domain expertise and data mining tools is known as "feature engineering." Feature engineering may be thought of as applied machine learning. Many in the industry believe it to be the most crucial step in improving model performance. It involves removing unnecessary columns, binning the numerical and categorical features, and aggregating some features.

Aggregating Multiple Features:

The variables need to be categorized so that the impact of making groupings can be seen more clearly. Many of the variables are either continuous in nature or have many discrete values that peak at specific places. New features—such as Metric of sum, Total score, Work fraction, Work start year, Years remaining to retire, and Performance—are calculated because of the following assumptions:

The metric of sum: this feature is the sum of awards won, KPIs met, and the previous year's rating.

Total score: training and workshops are essential for employees since they are conducted to help employees grow their skills. These training ratings assist the organization in determining whether staff are progressing. The columns "number of trainings" and "average training score" describe the number of company-organized workshops and training the employee attended and the average training score. The total score field is numeric and on the ratio scale. The goal is to see whether there are any correlations between the total score and the "is promoted" column. The total score column is separated into three bins (categories) for this purpose: Low (65 or below), Mediocre (65 to 145 points), and High (145 or higher).

Work fraction: this new feature was created to represent the fraction of work done with their age.

Work start year: this was another feature that represents the starting age of the employee.

Years remaining to retire: this is another new feature representing the remaining years for the employee until retirement.

Performance: For ease of analysis, KPIs_met and awards_won are combined into a single-column performance using the any() function. Any employee who has either won an award or has met KPIs has shown good performance. Most employees who were promoted demonstrated excellent performance, but those who were not promoted had a high rate of non-performance. Many employees who have worked hard yet have not been promoted. This might be related to a variety of different factors and provides a solid reason to investigate the other aspects as well.

Binning the Numerical and Categorical Features:

To perform well in this phase, we combine the levels of 'no_of_trainings', which has fewer observations in train and test data. For the age feature, we bin 'age' data into groups (every five years as a bin).

Removing Unnecessary Features:

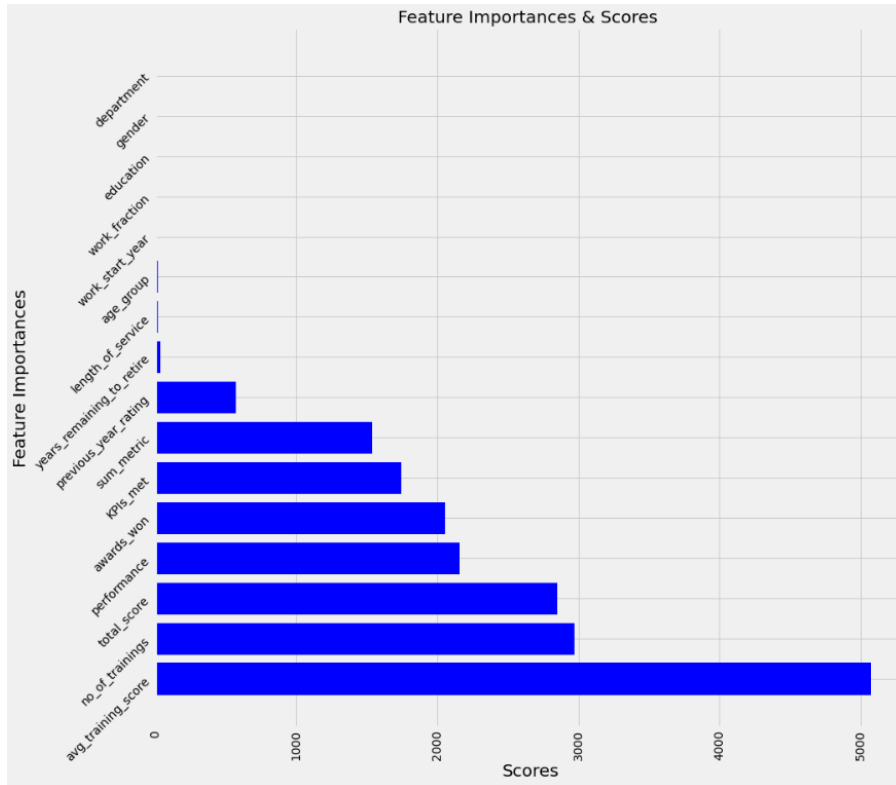
In the next stage of training, we will delete some of the columns that are not relevant for forecasting promotions. In addition, we will encrypt our object data and convert it to numeric form for the machine learning model to accept it. Categorical variables are well-known for hiding and masking important information in data collection. It is critical to understand how to cope in such situations. Otherwise, we lose out on discovering the most essential variables in a model. The train-test split is a method of assessing the performance of a machine learning system. It may be used for supervised learning and classification or regression tasks.

By eliminating the target column from the data, we store the target variable in y and the remainder of the columns in x. The Label Encoder will then encode the Department, Gender, and Number of Training Columns.

When creating a predictive model, the first and most critical phases in constructing our model should be feature selection and data cleansing. To conduct research, the dataset must contain many characteristics that impact employee advancement directly or indirectly. Irrelevant attributes in our data can reduce model accuracy and cause our model to train based on irrelevant information.

Machine learning feature selection strategies may be roughly categorized into the following. Wrapper, filter, and intrinsic supervised techniques are the three types of supervised methods. Using the model’s feature importance attribute, we can determine the feature importance of each feature in our dataset. We will use SelectKBest to extract the top features from the dataset.

Figure 7: Feature importance and scores



We next wrap the model in a SelectFromModel instance using the feature importance derived from our training dataset. This is used to pick features on our training dataset, train a model using the XGBoost classifier using the selected subset of features, and then assess the model on the test set using the same feature selection strategy. We may test several thresholds for picking features based on feature relevance for interest. The feature importance of each input variable allows us to rank each subset of features in order of significance, starting with all features and ending with the most significant feature (Brownlee, 2022). According to the results of this procedure, ‘department,’ ‘education,’ and ‘gender’ are omitted. These are necessarily the least important features for promotion prediction in our model. From the above conclusions, it can be stated with confidence that no factor alone is responsible for the promotion of an employee.

Table 2: Factors considered for predictive modeling

Factors considered for predictive modeling
no_of_trainings
previous_year_rating
length_of_service
KPIs_met
awards_won
avg_training_scor

sum_metric
total_score
work_fraction
work_start_year
years_remaining_to_retire
Performance
age_group

Data Manipulation: Preprocessing and Manipulate Data

Machine learning algorithms perform best when the number of samples in each class is about equal. This is because most algorithms are intended to enhance accuracy while minimizing mistakes. A class imbalance develops when observation in one class exceeds observation in other courses. They are often divided into two classes: the majority (negative) class and the minority (positive) class.

Resampling is a nonparametric statistical inference approach. Several statistical methods are available for resampling data, including oversampling and undersampling. The Synthetic Minority Oversampling Technique (SMOTE) oversamples the minority class by manufacturing "synthetic" cases rather than oversampling using replacement. After balancing the data, we are separating/splitting the entire dataset into training and testing data. Feature scaling is a strategy for lowering the values of all the independent characteristics of the dataset on the same scale. It is also known as data normalization and is performed during the data preparation step (SagarDhandare, 2022).

There must be a clear rule to determine when to normalize or standardize our data. We, therefore, have decent performance using standardization rather than normalization. It is best to fit the scaler to the training data before using it to change the testing data. This prevents data leaks during the model testing procedure.

Data Modeling

Classification has two separate implications in machine learning. Multiple predictive models such as XGBoost, Random Forest, Decision Tree, Logistic Regression, AdaBoost, and Gradient Boosting were applied in this scenario. The objective is to find the best classifier for the problem under consideration. Each classifier must be trained on the feature set, and the classifier with the best classification results is used to forecast. The original dataset was partitioned into two portions with an 80:20 ratio—one for training and another for testing.

Data Evaluation (fine & tune)

The data mentioned above set includes features such as "performance", "no_of_training", "previous_year_rating", and so on. Based on these values, the learning algorithm will anticipate whether the employee will be promoted to the organization. A typical confusion matrix may determine the number of cases properly categorized by a model. The confusion matrix visualizes a classifier's performance, providing a complete analysis with data on the number of true positives, false positives, true negatives, and false negatives. A classification report would show the model's accuracy, recall, Roc, AOC, and F1-Score. Precision and recall are based on the measure of relevance, with precision being the proportion of relevant samples found among the retrieved samples. When the datasets are divided, it is critical to maintain the same distribution of target variables across both the training and test datasets.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

$$\text{Precision} = \frac{TP}{TP + FP} \quad (16)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (17)$$

$$\text{F1 - Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (18)$$

The goal "is_promoted" attribute is a binary variable with 91% "No" and 9% "Yes", with both datasets retaining the same percentage after splitting. The classifiers are assessed using the evaluation criteria in order to choose the best model for the issue (Jain, Jain and Pamula, 2020).

Cross-validation is a strategy for preventing over-fitting and simplifying the model. This research utilized a five-fold CV. The training-set was randomly divided into five parts (k), with one serving as a validation-set and the other k-1s serving as training-sets. Each iteration used a different section as the validation set, and the average prediction error was calculated by averaging the average errors from both sets.

XGBoost is a powerful machine learning method, particularly in terms of speed and accuracy. A grid search must be performed for all relevant model parameters before hyper-parameter adjustment is required. There are numerous settings, especially in the case of XGBoost, and it may be CPU-expensive at times (Aarshay, 2020). XGBoost settings have been classified into three groups by the creators of XGBoost (Aarshay, 2020):

General parameters: Direct the overall operation.

Booster parameters: Direct the specific booster (tree/regression) at each stage.

Learning Task Parameters: Direct the optimization process.

The next step is using a general approach for parameter tuning. The various steps to be performed are:

Selecting a fast-learning rate: Generally, a learning rate of 0.1 is adequate, although values ranging from 0.05 to 0.3 should suffice for various problems. Determine the best number of trees to use for this learning rate. XGBoost has a very handy function called "cv" that does cross-validation at each boosting iteration, delivering the optimum number of trees needed. In our study, the learning_rate is fixed to 0.1 and cv to 5.

Tree-specific parameters (max depth, min child weight, gamma, subsample, colsample by tree) should be fine-tuned for the chosen learning rate and number of trees. Here, after many iterations and tuning with changing in different values and looking at the performance, we finally fixed these values: max_depth=4, min_child_weight=6, gamma=0.1, subsample=0.8, colsample_bytree=0.8.

Regularization settings (lambda, reg_alpha=0.01) for XGBoost can be adjusted to minimize model complexity and improve performance.

'scale_pos_weight' is one of the most critical factors people frequently overlook when dealing with an unbalanced dataset. This parameter should be fine-tuned with caution since it may result in overfitting the data.

4. RESULTS AND DISCUSSIONS

This phase assessed the suitability of the models used. But first and foremost, we had to select the appropriate variables for our work. Hence, as we mentioned earlier in section 3 about the importance of choosing the feature, we will display the results and emphasize using the XGBoost classifier in selecting those features because we deemed it more important. In our scenario, we train and test an XGBoost model on the whole training and test datasets.

Table 3. Evaluation procedure with several selected features

Threshold	Features Number	Accuracy
0.414	n =1	77.28%
0.150	n =2	77.28%
0.073	n =3	77.63%
0.073	n =4	78.66%
0.070	n =5	80.11%
0.040	n=6	80.08%
0.037	n=7	83.68%
0.026	n=8	84.04%
0.024	n=9	83.87%
0.018	n=10	83.92%
0.018	n=11	84.10%
0.014	n=12	83.93%
0.013	n=13	83.96%
0.012	n=14	84.16%
0.012	n =15	84.19%
0.006	n =16	84.21%

Table 3 demonstrates that the model's performance typically improves as the number of selected characteristics increases, starting with feature number seven, which has an accuracy of 83.68%. This problem has a trade-off between features and test set accuracy. We could decide to take a complex model (larger attributes such as n = 13) and accept a modest decrease in estimated accuracy from 84.21 percent to 83.96%, which is likely to be more useful based on the importance of the variables used, and, of course, the accuracy will improve more using grid search as the model evaluation scheme.

Following that, after selecting the 13 features to be used in the model, but this is insufficient for us, the next phase is to run the model and compare the results without the features that correlate, such as length of service and age being highly associated, as we discovered earlier in section 3. It can also be noticed that KPIs and previous year's ratings are correlated to some extent, signaling some linkage. Thus, we eliminate those two characteristics of age and KPIs to avoid multicollinearity. As a result, the following ten characteristics will be used in running the six models: 'the number of trainings', 'previous year rating', 'length of service', 'awards won', 'avg training score', 'sum metric', 'total score', 'work fraction', 'work start year', and 'years remaining'.

Comparing the results, clearly seen that having high accuracy with the 13 features rather than ten features. Therefore, we decided to choose 13 features to be employed in the model. Hence, the outcomes of the prediction phase judgments were initially gathered in the relative "confusion matrix," without using a grid search and then a grid search for each method. This is a matrix in which the classifier's predicted values are given in the columns, and the actual values of each instance of the test set are shown in the rows. To start with the performance evaluation, we used the confusion matrix to generate a set of essential metrics to quantify the efficiency of each algorithm: accuracy, precision, recall, and F1-score.

Table 4 summarizes these measures, which are based on the number of mistakes and accurate responses generated by the classifier.

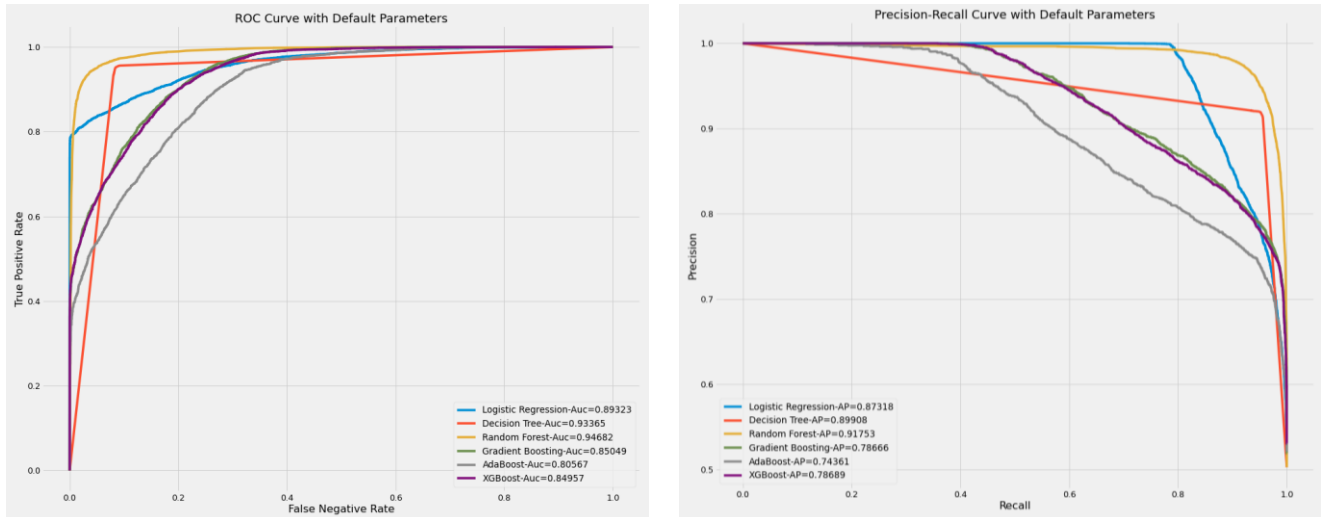
Table 4. Evaluation metrics with default parameters of different classifiers

Classifiers	Accuracy	Precision	Recall	F1-score	ROC AUC
Logistic Regression	0.88%	0.88%	0.88%	0.87%	0.87%
Decision Tree	0.92%	0.91%	0.91%	0.91%	0.91%
Random Forest	0.93%	0.92%	0.92%	0.92%	0.92%
AdaBoost	0.80%	0.79%	0.79%	0.79%	0.79%
Gradient Boosting	0.82%	0.82%	0.82%	0.82%	0.82%
XGBoost	0.82%	0.77%	0.77%	0.77%	0.77%

The results of this experiment revealed that all the classifiers had acceptable accuracy, which is greater than 80%. In many situations, this level of accuracy is seen as adequate. The dataset yielded acceptable models for this experiment's specified classification techniques. The model's accuracy is used to select the most acceptable classifier for the dataset to choose the appropriate classifier. As demonstrated in

Table 4, the Random Forest classifier has the best accuracy, with 93%, the highest among the chosen classifiers. Similarly, compared to other classifiers, Random Forest scores well on metrics such as precision or recall, F-Measure, and ROC AUC. However, the AdaBoost Classifier model is less accurate than others, with just over 80% accuracy. With 92% accuracy, the decision tree also performed well. **Hata! Başvuru kaynağı bulunamadı.** shows the same conclusion in terms of the ROC AUC graph. According to **Hata! Başvuru kaynağı bulunamadı.**, the random forest has the highest average precision, i.e., true positive rate.

Figure 8: Default parameters of different classifiers for a) ROC b) Precision-Recall Curve



a)

b)

The accuracy of DT and RF is substantially higher, and these classifiers may be used to forecast whether an individual will be promoted within the organization. However, in our research, we will focus more on the XGBoost classifier because it is the uniqueness of our study and the first time using this classifier to forecast such a problem; thus, we will use grid search to improve the accuracy of XGBoost beyond 82%.

Considering model hyper-parameters impact performance, we alter the parameters of five models using grid search and cross-validation, with a heavy emphasis on the XGBoost classifier. The basic concept behind this approach is to select numerous parameter combinations in advance and run

cross-validation for each set of parameters to discover the ideal parameter combination for XGBoost using five-fold cross-validation.

Classifiers	Accuracy	Precision	Recall	F1-score	ROC AUC
Logistic Regression	0.88%	0.883%	0.8832%	0.88%	0.8835%
Decision Tree	0.92%	0.91%	0.91%	0.91%	0.91%
Random Forest	0.93%	0.933%	0.9328%	0.93%	0.9327%
Gradient Boosting	0.86%	0.858%	0.8582%	0.86%	0.8582%
XGBoost	0.94%	0.94%	0.94%	0.9398%	0.9397%

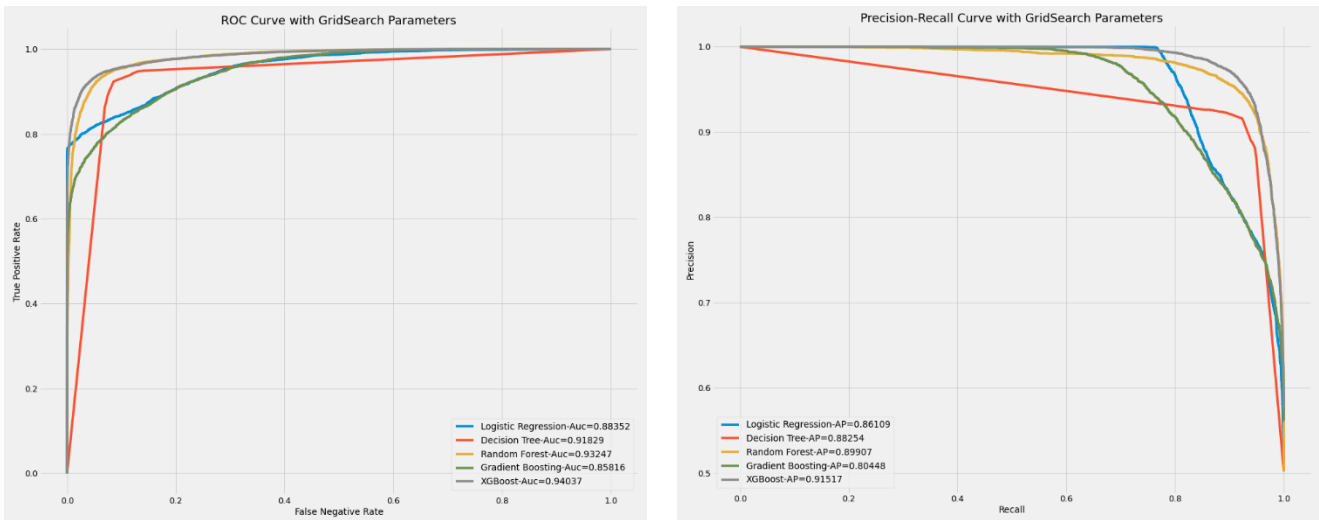
and

Table 5 Hata! Başvuru kaynağı bulunamadı. Hata! Başvuru kaynağı bulunamadı. show that the XGBoost model outperforms other models in terms of decile performance. It also consistently outperforms a random estimate, with XGBoost outperforming Random Forest. In terms of accuracy, memory consumption, and time-consuming, the XGBoost classifier surpasses the other classifiers.

Table 5. Evaluation metrics with GridSearch parameters of different classifiers

Classifiers	Accuracy	Precision	Recall	F1-score	ROC AUC
Logistic Regression	0.88%	0.883%	0.8832%	0.88%	0.8835%
Decision Tree	0.92%	0.91%	0.91%	0.91%	0.91%
Random Forest	0.93%	0.933%	0.9328%	0.93%	0.9327%
Gradient Boosting	0.86%	0.858%	0.8582%	0.86%	0.8582%
XGBoost	0.94%	0.94%	0.94%	0.9398%	0.9397%

Figure 9: Result with GridSerach parameters of different classifiers for a) ROC b) Precision-Recall Curve



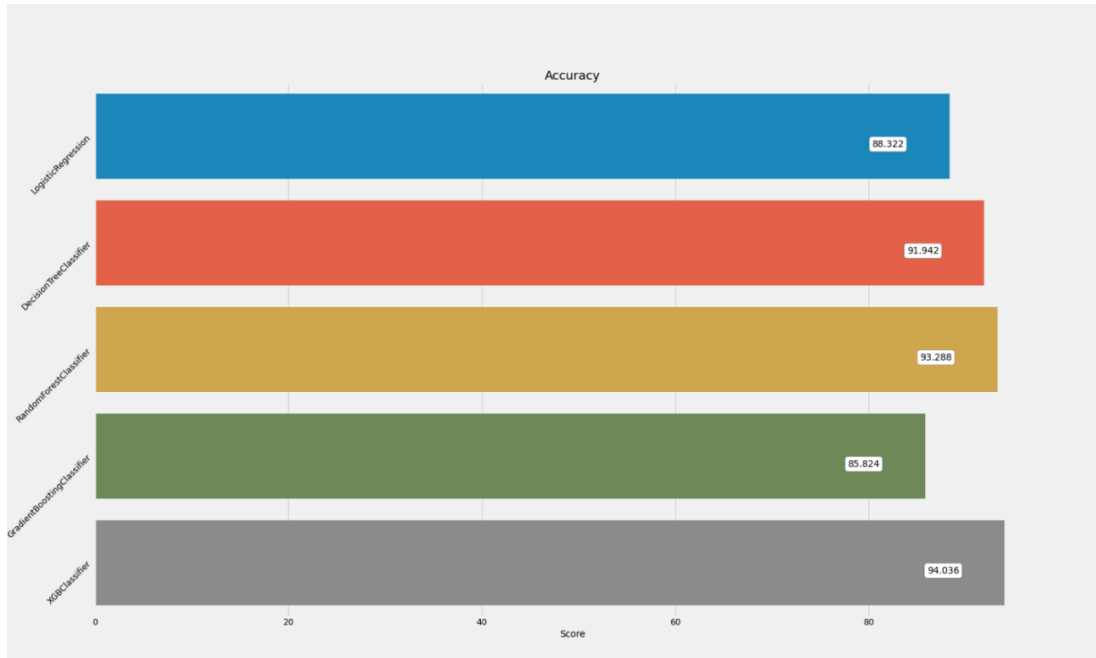
a)

b)

As the results show, while random forests rely on their randomization steps to help them achieve higher generalization, more is needed to prevent over-fitting in this scenario. On the other hand, XGBoost attempts to create new trees that complement the existing ones. Boosting improves training for difficult-to-classify data points. Another significant thing to consider is the over-fitting experienced by classifiers other than XGBoost, notwithstanding regularization or the addition of randomness. Because of its superior intrinsic regularization, XGBoost solves this issue and works wonderfully in our scenario.

The XGBoost classifier is also designed to be fault-tolerant in a distributed setting and is optimized for fast, parallel tree construction. The XGBoost classifier accepts DMatrix data. DMatrix is an XGBoost internal data structure designed for memory economy and training speed. DMatrixes were created here by combining numpy arrays containing features and classes. Due to those reasons, the XGBoost classifier was selected as the best classifier for the dataset.

Figure 10: Final accuracy of the classifiers



Furthermore, XGBoost surpasses the competition, and its time consumption is reasonable. As a result, the XGBoost classifier, based on 13 features, is chosen as the final prediction model, with accuracy and AUC of 94.036%, a recall of 94%, and a precision of 94%. It outperformed the baseline model in terms of accuracy, increasing it by up to 94%, as shown in

Classifiers	Accuracy	Precision	Recall	F1-score	ROC AUC
Logistic Regression	0.88%	0.883%	0.8832%	0.88%	0.8835%
Decision Tree	0.92%	0.91%	0.91%	0.91%	0.91%
Random Forest	0.93%	0.933%	0.9328%	0.93%	0.9327%
Gradient Boosting	0.86%	0.858%	0.8582%	0.86%	0.8582%
XGBoost	0.94%	0.94%	0.94%	0.9398%	0.9397%

In a comparable study conducted by Analytics Vydha, the participant rated first (faizankshaikh, 2022) received an accuracy of 92.88%, a recall of 42.13%, and a precision of 63.13% for promotion prediction, with a synthesized value (F1 score) of roughly 50.54% (faizankshaikh, 2022) whereas that of our experiment was 93.98%. Similarly, in the realm of employee turnover, papers (Punnoose and Ajit, 2016; Jain and Nayyar, 2018; Yedida et al., 2018) used the XGboost classifier to obtain accuracy values of 88%, 89%, and 90%, respectively. As a result, the suggested framework and technique perform well and are equal to the other published methods.

5. CONCLUSION

In this thesis, we attempt to forecast whether an employee will be promoted at their present company. To solve this categorization problem, we employ supervised machine learning methods.

Our approach's primary contributions are the application of machine learning algorithms and a framework for forecasting employee promotion.

XGBoost is recognized as a superior algorithm in terms of memory use efficiency, accuracy, and running time, with an accuracy and ROC AUC of 0.94036%, a recall of 0.94%, and a precision of 0.94%. It is a robust and scalable approach for handling all types of noise from large data sets. The suggested automated predictor's results show that the important promotion factors are average training score, number of training, total score, and performance.

The data analysis results constitute a beginning point for creating increasingly efficient employee promotion classifiers. Using extra datasets or simply updating them regularly, feature engineering to uncover new relevant qualities in the dataset and the availability of more information. Management may use this project to estimate the likelihood of promotion, allowing managers to choose the best conditions for someone to be promoted.

We want to deploy the suggested model in real-world firms soon so that organizations may learn about employee promotion variables. The study can be expanded by integrating features such as Scope of Development, Views on Workload Distribution, Career Goal Discussion, and Issues of Unhealthy Work Ethics. For example, we can estimate promotion speed or investigate whether a promoted person is qualified for a higher-level role and then provide appropriate management recommendations. It is advised to examine the use of deep learning models for forecasting promotion. A well-designed network with enough hidden layers may enhance accuracy, but scalability and practical implementation must also be considered. Instead of eliminating the region column, we may divide the 32 columns into two sections. We may also eliminate the gender column because there is only a small difference in the likelihood of males and females receiving a promotion..

REFERENCES

- Aarshay (2020). XGBOOST parameters: XGBoost parameter tuning. Analytics Vidhya. Available at: <https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-tuning-xgboost-with-codes-python/>. (Accessed April 11, 2022).
- Bandyopadhyay, N. and Jadhav, A. (2021) 'Churn Prediction of Employees Using Machine Learning Techniques.', Technical Journal / Tehnicki Glasnik, 15(1), pp. 51–59. Available at: <http://icproxy.khas.edu.tr/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edb&AN=149158643&site=eds-live>.
- Brownlee, J. (2022). Feature Importance and Feature Selection With XGBoost in Python. [online] Machine Learning Mastery. Available at: <https://machinelearningmastery.com/feature-importance-and-feature-selection-with-xgboost-in-python/> (Accessed 14 April 2022).
- Chen, K.-Y., Hsu, Y.-L. and Wu, C.-C. (2012) Num 2 Fall 2012 1 The International Journal Of Organizational Innovation Volume 5 Number 2 Fall 2012 Information Regarding The International Journal Of Organizational Innovation 4 IJOI, The International Journal of Organizational Innovation. Available at: <http://www.iaoiusa.org> (Accessed: 1 March 2022).
- Chen, K.-Y., Hsu, Y.-L. and Wu, C.-C. (2012) Num 2 Fall 2012 1 The International Journal Of Organizational Innovation Volume 5 Number 2 Fall 2012 Information Regarding The International Journal Of Organizational Innovation 4 IJOI, The International Journal of Organizational Innovation. Available at: <http://www.iaoiusa.org> (Accessed: 1 March 2022).
- De Pater, I. E. et al. (2009) 'Employees' Challenging Job Experiences And Supervisors' Evaluations Of Promotability', Personnel Psychology, 62(2), pp. 297–325. doi: 10.1111/j.1744-6570.2009.01139.x.
- Faizankshaikh (2022). wns-analytics-wizard-2018/Rank 1: Siddharth3977 at master · analyticsvidhya/wns-analytics-wizard-2018. [online] GitHub. Available at:

- <https://github.com/analyticsvidhya/wns-analytics-wizard-2018/tree/master/Rank%201:%20Siddharth3977> (Accessed 13 March 2022).
- Febrina, S. C. (2017) 'Predicting Employee Performance by Leadership, Job Promotion, and Job Environmental in Banking Industry', *Jurnal Keuangan dan Perbankan*, 21(4), pp. 641–649. doi: 10.26905/jkdp.v21i4.1630.
- Jain, P. K., Jain, M. and Pamula, R. (2020) 'Explaining and predicting employees' attrition: a machine learning approach', *SN Applied Sciences*, 2(4). doi: 10.1007/s42452-020-2519-4.
- Jain, R. and Nayyar, A. (2018) 'Predicting employee attrition using xgboost machine learning approach', in *Proceedings of the 2018 International Conference on System Modeling and Advancement in Research Trends, SMART 2018*. (1)Department of Computer Science and Engineering (CSE), Bharati Vidyapeeth's College of Engineering: Institute of Electrical and Electronics Engineers Inc., pp. 113–120. doi: 10.1109/SYSMART.2018.8746940.
- Jaiswal, Logistic regression in machine learning - javatpoint. www.javatpoint.com. Available at: <https://www.javatpoint.com/logistic-regression-in-machine-learning> (Accessed April 11, 2022).
- Jaiswal, Machine learning decision tree classification algorithm - javatpoint. www.javatpoint.com. Available at: <https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm> (Accessed April 11, 2022).
- Jaiswal, Machine learning random forest algorithm - javatpoint. www.javatpoint.com. Available at: <https://www.javatpoint.com/machine-learning-random-forest-algorithm> (Accessed April 11, 2022).
- Jantan, H. and Hamdan, A. (2010) 'Applying Data Mining Classification Techniques for Employee's Performance Prediction', *Knowledge ...*, pp. 601–607. Available at: <http://www.kmice.cms.net.my/ProcKMICe/KMICe2010/Paper/PG601-607.pdf> (Accessed: 29 November 2021).
- Li, M. G. T. et al. (2021) 'Employee performance prediction using different supervised classifiers', in *Proceedings of the International Conference on Industrial Engineering and Operations Management*, pp. 6870–6876.
- Liu, J. et al. (2019) 'A data-driven analysis of employee promotion: The role of the position of organization', in *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*. National University of Defense Technology, College of Systems Engineering: Institute of Electrical and Electronics Engineers Inc., pp. 4056–4062. doi: 10.1109/SMC.2019.8914449.
- Long, Y. et al. (2018) 'Prediction of employee promotion based on personal basic features and post features', in *ACM International Conference Proceeding Series*. Association for Computing Machinery, pp. 5–10. doi: 10.1145/3224207.3224210.
- Machado, C. S. and Portela, M. (2021) 'Age and Opportunities for Promotion', *SSRN Electronic Journal*. doi: 10.2139/ssrn.2367639.
- Najafi-Zangeneh, S. et al. (2021) 'An improved machine learning-based employees attrition prediction framework with emphasis on feature selection', *Mathematics*, 9(11). doi: 10.3390/math9111226.
- Navlani, A., AdaBoost classifier algorithms using python Sklearn tutorial. DataCamp. Available at: <https://www.datacamp.com/tutorial/adaboost-classifier-python> (Accessed April 11, 2022).
- Punnoose, R. and Ajit, P. (2016) 'Prediction of Employee Turnover in Organizations using Machine Learning Algorithms', *International Journal of Advanced Research in Artificial Intelligence*, 5(9). doi: 10.14569/ijarai.2016.050904.

- SagarDhandare (2022). Feature Scaling In Data Science!. [online] Medium. Available at: <https://medium.datadriveninvestor.com/feature-scaling-in-data-science-5b1e82492727> (Accessed 13 April 2022).
- Saradhi, V. V. and Palshikar, G. K. (2011) 'Employee churn prediction', *Expert Systems with Applications*, 38(3), pp. 1999–2006. doi: 10.1016/j.eswa.2010.07.134.
- Sarker, A. et al. (2018) Employee's Performance Analysis and Prediction using K-Means Clustering & Decision Tree Algorithm Mawlana Bhashani Science and Technology University Employee's Performance Analysis and Prediction using K-Means Clustering & Decision Tree Algorithm, Type: Double Blind Peer Reviewed International Research Journal Software & Data Engineering Global Journal of Computer Science and Technology: C.
- Tarbani (2021). Gradient boosting algorithm: How gradient boosting algorithm works. *Analytics Vidhya*. Available at: <https://www.analyticsvidhya.com/blog/2021/04/how-the-gradient-boosting-algorithm-works/>. (Accessed April 11, 2022).
- Yedida, R. et al. (2018) 'Employee Attrition Prediction', *IJISSET-International Journal of Innovative Science, Engineering & Technology*, 7(9). Available at: www.ijiset.com (Accessed: 13 January 2022).

Açık Anahtar Altyapısı ile Dijital İmzalamanın Zararlı Yazılımlar Üzerindeki Etkisi

Impact of Digital Signing on Malware in Public Key Infrastructure

Mehmetcan TOPAL¹ 
Zeynep ALTAN² 

DOI:10.33461/uybisbbd.1507316

Öz

Makale Bilgileri

Makale Türü:
Araştırma Makalesi

Geliş Tarihi:
29.06.2024

Kabul Tarihi:
19.08.2024

©2024 UYBISBBD
Tüm hakları saklıdır.



Geçmişten günümüze şifreleme, pek çok uygulamada kullanılan farklı yöntemleriyle büyük bir evrim geçirmiştir. Güçlü şifreleme algoritmalarının zaman içerisinde gelişimi, dijital iletişimde güvenliği sağlayan Açık Anahtar Altyapısını oluşturmuştur. Bu altyapının önemli bir bileşeni olan dijital imzalama günümüzde yaygın olarak kullanılmaktadır ve verinin doğruluğunu, bütünlüğünü ve güvenilirliğini önemli ölçüde sağlamaktadır. Bu çalışmada dijital imzalama yöntemlerinin, günümüz siber güvenlik dünyasında, zararlı yazılımların güvenilirliği üzerindeki etkisi değerlendirilmektedir. Zararlı yazılımların etkileri ve sonuçları her geçen gün artmakta olup, yaygın olarak kullanılan e-imza ve dijital sertifikalar da bu etkileri artırabilmektedir. Bu bağlamda çalışma, farklı yöntemlerle oluşturulan örneklerle dijital imzalama uygulanarak, zararlı yazılımların güvenilirlik ölçütlerinin karşılaştırmasını içermektedir. Testler sonucunda imzalı olan zararlı uygulamaların imzasız olan zararlı uygulamalara göre daha düşük olasılıkla güvenlik sistemlerine yakalandıkları ölçülmüştür. Özetle araştırma, dijital imzalamanın zararlı yazılımların yayılımını ne ölçüde etkilediğini ortaya koymayı ve siber güvenlik önlemlerinin geliştirilmesine katkı sağlamayı amaçlamaktadır.

Anahtar Kelimeler: Açık Anahtar Altyapısı, Dijital İmza, Şifreleme, Zararlı Yazılım.

Abstract

Article Info

Paper Type:
Research Paper

Received:
29.06.2024

Accepted:
19.08.2024

©2024 UYBISBBD
All rights reserved.



From past to present, cryptography has undergone a significant evolution from past to present, with various methods used in many applications. The development of strong encryption algorithms over time has established the Public Key Infrastructure, which ensures security in digital communication. A key component of this infrastructure, digital signing, is widely used today and plays a crucial role in ensuring the accuracy, integrity, and reliability of data. This study evaluates the impact of digital signing methods on the reliability of malware in the context of today's cybersecurity landscape. The effects and consequences of malware are increasing day by day, and commonly used e-signatures and digital certificates may also exacerbate these impacts. In this context, the study includes a comparison of the reliability metrics of malware by applying digital signing to examples created using different methods. Tests have shown that signed malware applications are less likely to be detected by security systems compared to unsigned ones. In summary, this research aims to reveal the extent to which digital signing affects the spread of malware and to contribute to the development of cybersecurity measures.

Keywords: Public Key Infrastructure, Digital Signature, Cryptography, Malware.

Atıf/ to Cite (APA): Topal, M. & Altan Z. (2024). Açık Anahtar Altyapısı ile Dijital İmzalamanın Zararlı Yazılımlar Üzerindeki Etkisi. Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi, 8(2), 99-109. DOI: 10.33461/uybisbbd.1507316

¹ Beykent Üniversitesi, Mühendislik-Mimarlık Fakültesi, mehmetcantopal9@hotmail, İstanbul, Türkiye.

² Dr. Öğr. Üyesi, Beykent Üniversitesi, Mühendislik-Mimarlık Fakültesi, zeynepaltan@beykent.edu.tr, İstanbul, Türkiye.

1. GİRİŞ

Açık anahtar altyapısının dijital iletişimde güvenliği sağlamadaki rolü bilişim dünyasında etkindir. Açık anahtar altyapısının temel özelliği, veri transferi sırasında farklı şifreleme yöntemleriyle gizliliği sağlamadaki kritik önemidir. Güçlü şifreleme yöntemleri ve dijital imzalama teknikleri, kullanıcıların verilerini korurken aynı zamanda kimlik doğrulaması da sağlarlar (Stallings, 2017). Açık anahtar altyapısının çevrimiçi alışveriş, bankacılık işlemleri, e-posta güvenliği ya da kimlik doğrulama gibi farklı uygulama alanlarında kullanılması, kullanıcı verilerini çeşitlendirmektedir. Bu çeşitlilik doğrultusunda verilerin korunmasındaki önem ve öncelik de artmaktadır. Bu bağlamda, açık anahtar altyapısının dijital imzalama ile bağlantısı da önemli olmaktadır. Dijital imzalama, herhangi bir belgenin veya iletişimin kimlik doğruluğunu, bütünlüğünü ve orijinalliğini sağlamada kullanılmaktadır (Garfinkel & Spafford, 2002). Böylece farklı tipteki verinin korunması ve güvenilirliği sağlanarak zararlı yazılımlar tespit edilebilir ve önlenir.

Açık anahtar altyapısı ve dijital imzalamanın zararlı yazılımlar üzerindeki etkileri incelendiğinde, bu teknolojilerin siber güvenlikte ne kadar kritik oldukları görülür. Açık anahtar altyapısı dijital imzalama ile verileri bütünleştirir; kimlik doğrulaması yaparak güvenli iletişimi destekler. Buna rağmen, özel anahtarın kötü niyetli saldırganların eline geçmesiyle güvenlik önlemleri zararlı yazılıma dönüşmekte, dosyanın güvenli bir şekilde imzalandığı ve güvenilir bir kaynaktan geldiği izlenimi yaratılmaktadır (Mike & Davis, 2021). Bu nedenle alınan güvenlik yöntemleri araştırılmalı, mevcut yöntemlerle nasıl birleştirilebileceği ve siber tehditlerin sürekli değişen yapısına nasıl uyarlanabileceği bilinmelidir.

Siber saldırıların yaygınlaşmasıyla birlikte, açık anahtar altyapısı ve dijital imzalama teknolojileri, siber güvenlik uzmanlarının ve kuruluşların verilerini korumada ve siber saldırılara karşı dirençli olmalarında önemlidir. Bunlar, sürekli olarak geliştirilen ve mevcut yöntemlerin iyileştirildiği dinamik alanlardır. Açık anahtar altyapısındaki zararlı yazılımların kötüye kullanılması ve bu bağlamda ortaya çıkan güvenlik açıklarının etkileri giderek artmaktadır. Bu çalışma, bu sorunu derinlemesine incelemek, açık anahtar altyapısının güvenliğini artırmak ve zararlı yazılımların tespit edilmesi için etkili çözümler sunmayı amaçlamaktadır. Böylece bilgi güvenliği alanındaki diğer araştırmalar için de bir ön çalışma niteliği taşımaktadır. Çalışmada özetle, karşı bağlantı sağlamaya yarayan farklı zararlı yazılımlar üzerinden güvenlik sistemlerine yakalanma durumları araştırılmaktadır. Dijital imzalama, ilk olarak netcat³ .exe adlı araca ve kullanımı açık olan msfvenom⁴ adlı aracın zararlı işlemini şifreleme uygulanarak gerçekleştirilmiştir. Yapılan testler sonunda imzalı zararlı uygulamaların imzasız zararlı uygulamalara göre güvenlik sistemleri tarafından yakalanma olasılıklarının daha düşük olduğu sonucu çıkarılmıştır.

1.1 Geçmişten Günümüze Önemli Siber Saldırıları

Siber saldırılar, dijital çağın en büyük tehditlerinden biri olarak ortaya çıkmıştır. Bu tür saldırılar, bilgi güvenliğini tehlikeye atarak bireyler ve kurumlar için ciddi riskler oluşturur. Tarihsel olarak, çeşitli siber saldırılar bilgisayar sistemlerini ve ağlarını hedef almış, güvenlik açıklarından yararlanarak veri çalmış veya sistemleri işlemez hale getirmiştir.

Tablo1’de uluslararası ölçekte büyük etki yaratan siber saldırılar verilmekte ve bunlarla ilgili açıklamalar yapılmaktadır. Her bir saldırı ait olduğu sektörü farklı şekilde etkilemiştir.

³ Netcat, TCP ve UDP protokollerini kullanarak ağ bağlantıları üzerinde veri okuma ve yazma işlemleri gerçekleştiren bir komut satırı aracıdır.

⁴ Msfvenom, Metasploit Framework’ün bir parçası olan ve zararlı yazılım oluşturmak, çeşitli kod yüklerini ve kod çeviricilerini birleştirmek için kullanılan bir komut satırı aracıdır.

Tablo 1: Tarihteki Önemli Siber Saldırıları

İsim	Açıklama	Kaynak
Morris (1988)	Solucanı İnternet'in erken dönemlerinde dünya ölçeğinde bilgisayar ağlarını etkileyen ilk büyük saldırdır. Bilgisayar sistemlerindeki güvenlik açıklarından yayılmıştır.	Spafford, 1988
Sony Siber (2014)	Pictures Şirketlerin güvenlik önlemlerini güçlendirmesi gerektiğine ilişkin bir uyarı niteliğindedir ve siber saldırıların işletmelerin itibarlarına olası zararlarını yansıtmaktaydı.	Peterson, 2014
Stuxnet (2010)	Saldırısı İran'ın nükleer programını hedef olarak engellemeye çalışan ve devletler arasında bir siber savaş başlatabilecek güçteki karmaşık bir siber saldırdır.	Zetter, 2014
WannaCry (2017)	Büyük ölçekli bir fidye yazılımı saldırısıdır; EternalBlue isimli bir güvenlik açıklığından yararlanarak bilgisayar sistemlerini kilitlemiş ve kuruluşlardan fidye talep etmiştir.	Greenberg, 2017
Moonlight Maze (1996-1998)	Amerikan savunma ve istihbarat ağlarına erişmiştir. Döneminin en karmaşık ve etkili siber casusluk saldırısıdır.	Haizler, 2017
Melissa (1999)	Virüsü e-posta kullanıcılarını etkilemiştir. Bilgisayar sistemlerinde önemli veri kayıplarına ve performans sorunlarına yol açmıştır. e-posta güvenliğinde ciddi farkındalıklara neden olmuştur.	Taylor, 2020
NotPetya (2016)	saldırısı Fidye yazılımı olarak başlamış, daha sonra bir siber sabotaj olayına dönüşmüştür. Başlangıçta Ukrayna'daki kuruluşları hedef alan bu saldırı, dünya ölçeğinde pek çok bilgisayarı etkilemiş ve büyük krize neden olmuştur.	(Fayi, 2018)
Colonial Pipeline Saldırısı (2021)	Pipeline DarkSide fidye yazılımı ile etkisi altına aldığı tüm operasyonları kilitlemiştir. Bu saldırı, enerji sektöründe siber güvenlik risklerine ilişkin endişeleri arttırmıştır.	Robertson & Turton, 2021
Emotet Saldırısı	Bilgisayar ağlarına sızarak kötü amaçlı e-postalar gönderen ise botnet ağıdır. Bankaların bilgilerini çalan, fidye yazılımlar indiren ve pek çok zararlı eylemde bulunmuş ve pek çok kuruluş için ciddi güvenlik riski oluşturmuştur.	Europol, 2021
AnyDesk (2024)	Saldırısı Bilgisayar korsanlarının Aralık 2023 sonlarında sistemlere girdiği bir operasyondur. Saldırganlar, kritik kişisel bilgileri ve bazı şirkete ait belgeleri ele geçirmişlerdir.	Sporx, 2024

Saldırılarda kullanılan dosyalar üzerinde güvenlik önlemleri alınmasına rağmen, sızdıkları sistemlerde görünmeden hareket etmeleri sonucunda zararlı yazılımların tümü hedefine ulaşmıştır. Bunun nedeni, şaşırtma, morfizm, şifreleme, enjeksiyon gibi yöntemlerin kullanılmış olmasıdır. Şaşırtma yöntemi, güvenlik çözümlerinin tespit edilmesi işlemini güçleştirir. Örneğin, ölü ya da önemsiz bir kod eklenmesiyle, yeniden kayıt atama ya da talimat değişikliği yapılabilir (Balakrishnan & Schulze, 2005). Morfizm, özellikle polimorfik virüsler olarak, sınırsız sayıda farklı şifre çözücü oluşturarak analizi zorlaştırmayı amaçlayan karmaşık bir tekniktir. Polimorfik virüsler, şifre çözme kodunun görünümünü kopyaladıkça değiştirmek üzere çok sayıda gizleme tekniği kullanır (Rad ve diğerleri, 2012). Şifreleme, zararlı uygulamanın ikili sistemde şifrelenmesidir; uygulamanın çalışabilmesi için çalışmadan önce tekrar deşifre edilerek yüklenmesi gerekir (Tasiopoulos & Katsikas, 2014). DLL dosyaları çift tıklama ile doğrudan çalışmadığı için, DLL formatındaki kötü amaçlı bir kod derlendiğinde doğrudan çalışmayacaktır. Bu dosyanın çalışması için explorer.exe gibi bir platformda ana bilgisayar tarafından yükleniyor gibi çalıştırılması gerekir; böylece zararlı

uygulama saldırgan tarafından gizlenir (Monnappa, 2018). Bir başka zararlı uygulama olan işlem enjeksiyonu, kötü amaçlı bir işlemi başka bir işlemin bellek alanında çalıştırarak uygulamayı gizlemek amacıyla kullanılan bir yöntemdir. Enjekte edilen işlem, ana bilgisayarın yetkilerini ele geçirir (Balaoura, 2018).

Çalışmanın sonraki bölümünde açık anahtar altyapısı ile ilgili temel tanımlamalar, karma fonksiyonları ve dijital imzalama altyapısı ile açık anahtar altyapısı güvenliği anlatımlarını içermektedir. Bölüm 3' de ise, zararlı yazılımların güvenilir sertifikalarla imzalanmasının nasıl gerçekleştirildiği örneklerle karşılaştırarak anlatılmaktadır. Sonuç Bölümünde ise, dijital imza yöntemlerinin ve güvenlik çözümlerinin kullanımının masaüstü uygulamalarının güvenilirliğini arttırmadaki etkisi özetlenmektedir.

2. AÇIK ANAHTAR YAPISI

İnternet üzerinde güvenli iletişim ve veri paylaşımının önemi günümüzde giderek artmaktadır. Açık anahtar altyapısı ise hem kimlik doğruluğunu, gizliliği, veri bütünlüğünü sağlamada kritik bir rol oynamakta hem de dijital sertifikaların, dijital imzaların ve şifreleme anahtarlarının güvenilir dağıtımını ve yönetimini sağlamaktadır. Açık anahtar altyapısının çalışma ilkesi temel olarak her kullanıcının açık ve özel anahtar şeklinde bir çift anahtara sahip olmasıdır. Açık anahtar genel olarak erişilebilir ve diğer kullanıcılarla paylaşılabilir. Özel anahtar ise sadece kullanıcıya özeldir ve gizli tutulmalıdır. Bu anahtar çiftleri, açık anahtar altyapısının güvenli iletişimi sağlamadaki çatısını oluşturur.

Açık anahtar altyapısı ve dijital imzalamada verilerin doğruluğu için karma fonksiyonları önemli yer tutar. Karma fonksiyonları, verilerin benzersiz bir diziye dönüştürülmesini sağlarlar. Ayrıca veriyi sabit boyutlu bir çıktıya dönüştüren matematiksel işlevlerdir. Bu işlevler, girdinin boyutu ne olursa olsun, genellikle sabit bir boyutta bir çıktı üretir. Karma fonksiyonları genellikle bir dizi rastgele karakterden oluşur ve girdinin herhangi bir değişikliğinde bile farklı bir değer üretilir. Öncelikle, dosyanın veya veri bloğunun karma değeri hesaplanır. Daha sonra, veri veya dosyanın değiştirilmediğinden emin olmak için bu değer yeniden hesaplanır. İki karma değeri eşleşirse, verinin değiştirilmediği doğrulanır. İki karma değeri eşleşmiyor ise veri değiştirilmiştir (Paar & Pelzl, 2010).

Dijital imzalama, bir belgenin veya iletişimin doğruluğunu, bütünlüğünü ve orijinalliğini doğrulamada kullanılır. Bu süreçte, belgeyi imzalayan kişinin kimliği doğrulanır ve belgenin imzalı olduğu ve değiştirilmediği garanti altına alınır. Dijital imzalama genellikle internet üzerindeki güvenli bağlantılarda, e-postalarda veya şifrelemelerde kullanılır. Dijital sertifika ile internet ortamında kimlik doğrulanmış olur; bir belge imzalandığında, imzalayan taraf belgenin içeriğini bilmekle birlikte kendisine ait olduğunun onayını vermiş ve kendisi tarafından yollandığını belirtmiş olur (Nash ve diğerleri, 2001).

Dijital imzalama işlemi, genellikle üç temel adımdan oluşur: karma değeri hesaplama, şifreleme ve doğrulama. İlk adımda, belgenin karma değeri hesaplanır. Bu, belgenin bütünlüğünü ve orijinalliğini doğrulamak için kullanılır. İkinci adımda, belgenin karma değeri simetrik veya asimetric şifreleme yöntemlerinden biriyle şifrelenir. Simetrik şifrelemede, verinin şifrelenmesi ve çözülmesi için aynı anahtar kullanılır. Bu yöntem hızlıdır ve verimli bir şekilde işlem yapar, ancak anahtarın güvenliğini sağlamak kritik öneme sahiptir. Asimetric şifrelemede ise, verinin şifrelenmesi ve çözülmesi için iki farklı anahtar kullanılır: biri açık anahtar ve diğeri özel anahtar. Bu yöntem, anahtarların güvenliğini sağlasa da işlem süreci daha karmaşıktır. Üçüncü adımda, alıcı, belgenin karma değerini ve imza olarak alınan şifrelenmiş karma değerini açık anahtarıyla çözer ve bu iki değeri karşılaştırır. Eşleşme durumunda, belgeyi imzalayan kişi ve belgenin bütünlüğü doğrulanmış olur. İki değer eşleşmemesi durumunda verinin değiştirildiği kabul edilir ve alıcı taraf bu durum ile ilgili olarak bilgilendirilir.

Özellikle çevrimiçi işlemler ve iletişimlerde, dijital imzalama ve karma fonksiyonları, verilerin güvenliğini sağlamada temel önlemler olarak kabul edilmektedir.

2.1. Açık Anahtar Altyapısı Güvenliği

Açık anahtar altyapısı, dijital iletişimde güvenliği sağlamak için yaygın olarak kullanılan bir sistemdir. Dijital sertifikalar, dijital imzalar ve şifreleme anahtarlarının güvenilir dağıtımı ve yönetimini sağlamasına rağmen, açık anahtar altyapısının bazı zayıf noktaları vardır. İlki, güvenilirlik ve kimlik doğrulama sorunudur. Açık anahtar altyapısı, genellikle bir üçüncü tarafın güvenilirliğini gerektirir. Ama bu garanti değildir ve saldırganlar tarafından hedef alınabilirler. Böylece, sertifika yetkilileri veya sertifika dağıtımı ile açık anahtar altyapısı bileşenlerinin güvenliği zayıflar. Bir başka zayıf nokta da anahtar yönetimidir. Anahtarların güvenliği sağlanmadığında, şifreleme anahtarlarının ele geçirilmesi veya kötü niyetli kullanımı gibi riskler ortaya çıkabilir. Anahtar yönetimi, dikkatli bir şekilde yapılmalı ve anahtarların güvenliği için sıkı güvenlik önlemleri alınmalıdır. Diğer taraftan, yanlış algoritmaların veya zayıf parametrelerin kullanılması da kriptografik güvenliği tehlikeye atabilir. Güvenilmeyen ve zayıf kimlik doğrulama süreçleri çok faktörlü kimlik doğrulamalarla sağlamlaştırılmalıdır. Saldırganlar, sosyal mühendislik taktikleri ile sisteme sızabilirler ve kullanıcıları manipüle ederek kimlik bilgilerini ele geçirebilirler (Klimburg-Witjes&Wentland, 2021). Hatta fiziksel erişim ile anahtarları veya cihazları çalabilirler İnsan faktörü de açık anahtar altyapısını etkileyebilir. Güçlü şifrelerin kullanılmaması, anahtarların korunmasında dikkatsizlik, güvenli olmayan ağlarda iletişim, güncellenmemiş yazılımlar gibi pek çok etken, insan hatalarından kaynaklı olarak açık anahtar altyapısının güvenliğini tehlikeye sokmaktadır.

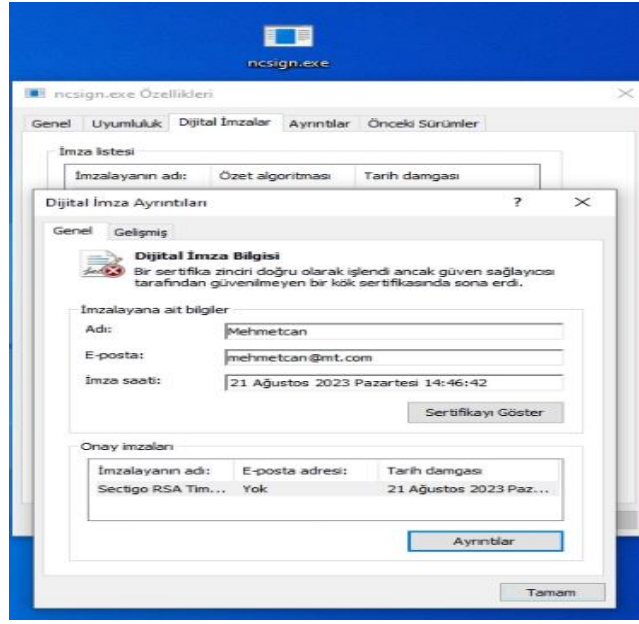
3. ZARARLI YAZILIMLARIN GÜVENİLİR SERTİFİKALARLA İMZALANMASI

Zararlı yazılımların güvenilir sertifikalarla imzalanması, yazılımın güvenli olduğu anlamına gelmez. Aksine, bu tür yazılımlar kötüye kullanılarak kullanıcıları ve sistem yöneticilerini yanıltabilir ve güvenlik riski doğurabilir. Buna rağmen, güvenilir sertifikalarla imzalanmış yazılımlar, kişiler ve güvenlik yazılımları tarafından daha güvenilir kabul edilir. Böylece kullanıcılar ve sistem yöneticileri yanıltılarak güvenlik yazılımları atlatılabilir. Zararlı yazılım geliştiricileri, sahte veya çalıntı sertifikalar kullanarak yazılımlarını imzalayabilirler. Bu nedenle, güvenilir sertifikaya sahip bir yazılım bile kötü amaçlı olabilir. Diğer bir ifade ile, zararlı bir yazılım güvenilir sertifikalarla imzalanarak güvenlik önlemlerini atlatarak sisteme sızabilir. Bu durumda sistem, veri kaybı, kimlik hırsızlığı, finansal kayıp ve daha birçok soruna açık hale gelir.

Zararlı yazılımların güvenilir sertifikalarla imzalanarak güvenilirlik sağlama girişimi büyük bir risktir. Güvenilir sertifikalar, genellikle güvenilir olduğu düşünülen ve sertifika yetkilileri tarafından onaylanan kuruluşlar tarafından sağlanır. Ancak, kötü niyetli aktörler bu sertifikaları kötüye kullanarak zararlı yazılımları imzalamak ve yaymak için güvenilirlik algısını suistimal edebilirler.

3.1. Zararlı Yazılım Uygulaması

Bu bölümde, ilk olarak nc.exe uygulaması ve şifreleme uygulanmış msfvenom aracı ile oluşturulan kötü amaçlı yük kullanılarak, dijital imza yöntemiyle zararlı yazılımların güvenilirlik algısının arttırmasına yönelik bir uygulama yapılmaktadır. Genellikle sızma testleri ve güvenlik açığı uygulamalarıyla kötü amaçlı zararlı yazılımlar oluşturularak hedef sistemlere saldırı gerçekleştirilir. Bu yazılımlar hem farklı işletim sistemleri ve platformlar tarafından desteklenir, hem de dosyanın bazı özellikleri özelleştirilerek farklı modüllerle birlikte kullanılıp etkilerini arttırılabilir. nc.exe gibi araçlar, çeşitli güvenlik açıkları ve kötü amaçlı yazılımlar için potansiyel bir kullanım alanı sağlarlar; ayrıca ağ trafiğini yönlendirme ve değiştirme yeteneklerine sahiptirler. Çalışmada nc.exe'nin karşı bağlantı sağlama özelliği kullanılmaktadır. Msfvenom aracı ile oluşturulan kötü amaçlı yükte karşı bağlantı sağlama özelliğine göre hazırlanmıştır. Böylece saldırı vektörü ve etkisi aynı, fakat yöntem ve hazırlanışı farklı olan iki uygulamanın güvenlik sistemleri üzerindeki etkisi karşılaştırılmaktadır.

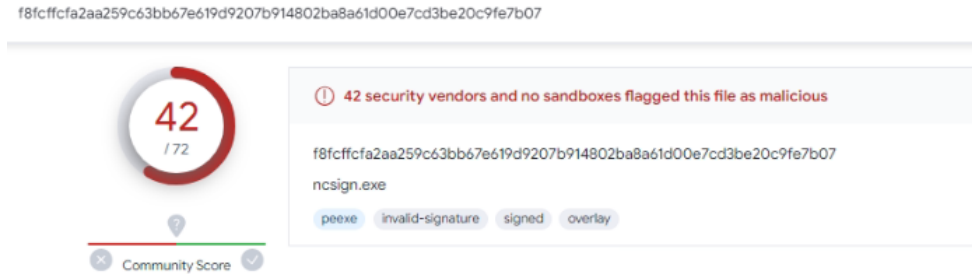


Şekil 4: İmzalama Sonrası Ayrıntılar

İmzasız olan “nc.exe” ve imzalı olan “ncsign.exe” dosyalarının, VirusTotal⁵ sitesindeki antivirüs uygulamaları üzerindeki yakalanma durumları Şekil 5 ve Şekil 6’da görülmektedir. Popüler kullanımda olan antivirüs uygulamaları imzasız olan nc.exe uygulamasını zararlı olarak algılamaktadır. Doğrudan zararlı olarak algılanmasının sebebi, uygulamanın derlenmiş olarak internette yer alması, karma değerinin birçok güvenlik sisteminin veri tabanlarında zararlı olarak belirtilmesidir.



Şekil 5: “nc.exe” dosyasının VirusTotal ile kontrolü



Şekil 6: “ncsign.exe” dosyasının VirusTotal ile kontrolü

⁵ Kötü amaçlı yazılımlar ve diğer ihlallerin tespiti için şüpheli dosyaların, etki alanlarının, IP'lerin ve URL'leri analiz edildiği ve bunların otomatik olarak paylaşıldığı platform.

Şekil 7’de “nc.exe” uygulamasının karşı bağlantı sağlama özelliği ile aynı işleve sahip farklı bir uygulama incelenmektedir. Msvfnom aracı ile karşı bağlantı sağlayan kötü amaçlı yük değeri alınmıştır. Bu değer ilk olarak önceden bir XOR işlemine girmiştir. Böylece başlangıçtaki şifreli değer anlaşılmayacak hale getirilmiştir.

```
// Encrypted metasploit shellcode
unsigned char encrypted_shellcode[] = {
    "0xbd, 0x9, 0xc2, 0xa5, 0xb1, 0xa9, 0x8d, 0x41, 0x41, 0x41, 0x0, 0x10, 0x0, 0x11, 0x13, 0x10, 0x17, 0x9, 0x70, 0x93,
};

// Decrypt function
void decryptShellcode(unsigned char* shellcode, size_t size) {
    for (int i = 0; i < size; ++i) {
        shellcode[i] ^= XOR_KEY;
    }
}

int main() {
    // Decrypt the shellcode
    decryptShellcode(encrypted_shellcode, sizeof(encrypted_shellcode));

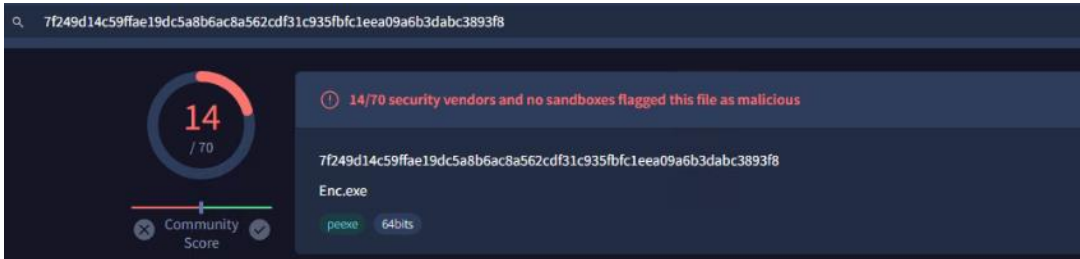
    // Allocate memory for the shellcode
    void* mem = VirtualAlloc(0, sizeof(encrypted_shellcode), MEM_COMMIT, PAGE_EXECUTE_READWRITE);
    if (mem == NULL) {
        std::cerr << "Failed to allocate memory!" << std::endl;
        return 1;
    }

    memcpy(mem, encrypted_shellcode, sizeof(encrypted_shellcode));
}
```

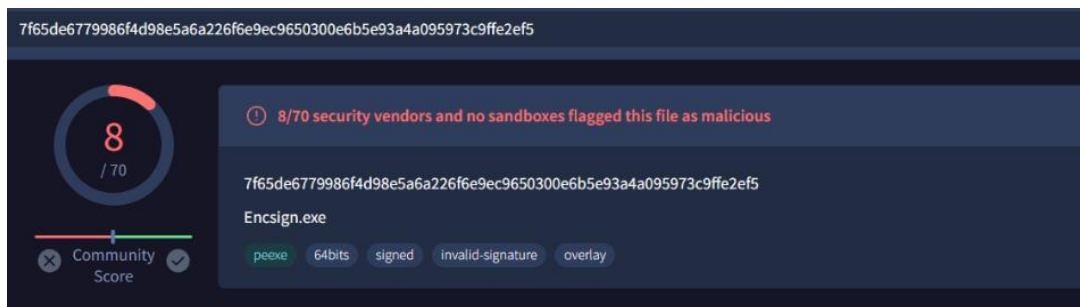
Şekil 7: Özel Uygulamaya ait Kaynak Kod

XOR, basit bir şifreleme formudur ve genellikle verileri gizlemek için kullanılır. Temelinde simetrik şifrelemeye dayanır. Uygulama çalıştırıldığında, XOR işlemi ile şifrelenmiş veri deşifre edilerek çözümlenme işleminde byte değerleri tek tek belleğe yazılmakta ve tekrardan çalışır duruma getirilmektedir.

Burada uygulamanın “Enc.exe” olarak derlenmesinden sonra, Şekil 3’teki gibi imzalama adımı uygulanmış ve “Encsign.exe” adlı dosya oluşturulmuştur. “Enc.exe” ve “Encsign.exe” uygulamalarının VirusTotal üzerindeki yakalanma durumları Şekil 8 ve Şekil 9’da yer almaktadır.



Şekil 8: VirusTotal’de Enc.exe



Şekil 9: VirusTotal’de Encsign.exe

Sonuç olarak, bu çalışmada aynı işleve sahip farklı zararlı uygulamaların güvenlik sistemlerine yakalanma durumları incelenmiştir. Gerçekleştirilen testler doğrultusunda imzalı olan zararlı uygulamaların imzasız olan zararlı uygulamalara göre güvenlik sistemleri tarafından yakalanma olasılığının daha düşük olduğu gözlemlenmiştir (Tablo 2). Buradan da dijital imzalamanın, güvenilirlik ölçütünün artmasında etkili olduğu sonucu çıkarılabilir.

Tablo 2: Uygulamaların Karşılaştırılması

	İmza Durumu	Başarı Oranı
nc.exe	İmzasız	51/70
ncsign.exe	İmzalı	42/72
Enc.exe	İmzasız	14/70
Encsgin.exe	İmzalı	8/70

5. TARTIŞMA VE SONUÇ

Bu çalışmada, dijital imzalama yöntemi ile zararlı uygulamaların güvenilirlik etkileri, açık anahtar altyapısı kapsamında incelenmiştir. Dijital imza, dosyaların veya uygulamaların orijinal ve güvenilir kaynaklardan geldiğini doğrulamak için kullanılan önemli bir araçtır. Bu bağlamda zararlı uygulamaların güvenilirlik seviyelerini artırmak amacıyla dijital imza yöntemlerinin uygulamalara olan etkilerine odaklanılmıştır.

Sonuçların imzalamadan dolayı mı ya da ilgili dosyaların karma değerinin değiştiğinden dolayı mı farklılık gösterdiği bu çalışma doğrultusunda cevaplanması gereken önemli bir sorudur. Karma değerinin kaynak kod üzerindeki bir değişiklikten dolayı değişmesi de VirusTotal gibi sitelerdeki etkisini elbette değiştirebilmektedir. Bu durumla ilgili bir çalışma olarak canlı ortamlarda farklı güvenlik sistemleri üzerinde çalışmalar yapılmıştır. Fakat, etik olarak isim ve ekran görüntüsü paylaşılmamaktadır. Yapılan uygulamalarda kaynak kodu aynı olan zararlı uygulama farklı zamanlarda derlenerek farklı karma değerli halleri üretilmiştir. İki uygulamanın da canlı ortamda güvenlik sistemlerine yakalandığı tespit edilmiştir. Bu iki uygulamanın dijital imzalama sonucu aynı güvenlik sistemleri üzerinde başarılı sonuçlar verdiği görülmüştür. VirüsTotal üzerinde de başarı oranının arttığı görülmektedir. İlgili güvenlik sistemlerinin yapılandırmaları bu çalışma için elbette büyük önem taşımakta ve fark yaratmaktadır.

Güvenlik önlemlerini aşmanın farklı yolları vardır. Bunlar kod betikleri, şifreleme teknikleri, özel olarak oluşturulmuş kütüphaneler ve güvenlik ürünlerinin çalışma mantığına göre değişir. Günümüzde internet güvenliği için birçok araştırma ve çalışma yapılmaktadır. e-Ticarette kredi kartı bilgilerinin korunması, güvenli iletişim sağlanması gibi konularda yoğun çalışmalar devam etmektedir. Siber saldırılar, bireylerden veya kuruluşlardan gelebilir. Bunlar genellikle güvenlik ölçütlerinin etkinliği ile yakından ilişkilidir. Bu bağlamda, çalışmanın odağında açık anahtar altyapısında önemli bir rol oynayan dijital imzalama tekniğinin zararlı yazılımlar üzerindeki etkileri ve güvenlik ürünlerinin tepkileri olmuştur. Dijital imza tekniği, sadece güvenlik ürünlerini atlatmak için değil, aynı zamanda kullanıcıları da aldatmak için de kullanılabilir. Kullanıcılar, bir uygulama indirildiğinde sertifikasız olduğunda uyarılır ve kullanıcı genellikle bu uyarıyı kapatır. Ancak, sertifikalı bir uygulama indirildiğinde, sistem tarafından güvenli olarak kabul edildiğinden, kullanıcılar genellikle uygulamayı güvenli kabul edip çalıştırır. Bu durum, kullanıcıları yanıltmak için çeşitli dosya formatlarında kullanılabilir. Dijital olarak imzalı bir uygulamanın güvenlik çözümlerinin ve insanların güvenilirlik algısı üzerinde etkisi de fazladır. Çalışma bu faktörlerin hepsine bir uyarı niteliğindedir.

Diğer taraftan makine öğrenmesi algoritmalarıyla da zararlı yazılımların tespit edilmesi ve engellenmesi konusunda büyük ilerlemeler kaydedilmektedir. Zararlı yazılımların tespiti ve analizinde dinamik analiz tekniklerinin kullanılması, evrişimsel sinir ağları ile ikili dosya yapısına dayalı zararlı yazılım tespiti ve diğer farklı makine öğrenmesi algoritmalarının karşılaştırılması gibi yöntemlerin etkili sonuçlar verdiği görülmektedir. Bu araştırmalar, makine öğrenmesi tabanlı yaklaşımların dijital imzalama ve sertifikanın kötüye kullanımını belirlemede proaktif ve adaptif çözümler sunmaktadır. Ayrıca, zamana bağlı olarak yapılan dijital imza kontrolünün zararlı yazılımın yakalanmasına katkı sağladığına ilişkin çalışmalar da bulunmaktadır. Dijital imzaların zaman damgaları ve geçerlilik sürelerinin analiz edilmesi, zararlı yazılım tespitinde ek güvenlik sağlamaktadır. Dijital imzalama tekniği, zararlı yazılımların içerisini gizlemek için kullanılan bir yöntem olarak düşünülebilir. Makine öğrenmesi algoritmaları ile dinamik analiz sürecindeki çalışmalarla imza ve sertifika kontrollerinin geliştirilmesi, zararlı yazılımların tespit edilmesinde giderek daha fazla önem taşıyacaktır. Dinamik analizin ve yapılandırmanın iyi olduğu sistemlerde dijital imzalamanın, güvenlik ürünlerini atlatma adımlarında bir etkisi olmadığı da görülmektedir. Gelecekte, bu tekniklerin daha da geliştirilmesi ve geniş çapta uygulanmasıyla, zararlı yazılımlara karşı daha yüksek düzeyde koruma sağlanabilecektir.

KAYNAKÇA

- Bal Krishnan, A. & Schulze, C. (2005). Code Obfuscation Literature Survey. Computer Sciences Department, University of Wisconsin.
- Balaoura, S. (2018). Process Injection Techniques and Detection Using the Volatility Framework. Master's thesis, University of Piraeus, Greece.
- Europol. (2021). "World's Most Dangerous Malware EMOTET Disrupted Through Global Action". <https://www.europol.europa.eu/media-press/newsroom/news/world's-most-dangerous-malware-emotet-disrupted-through-global-action>
- Fayi, S. (2018). What Petya/NotPetya Ransomware Is and What Its Remediations Are. 10.1007/978-3-319-77028-4_15.
- Garfinkel, S. & Spafford, E. (2002). Web Security, Privacy and Commerce. O'Reilly Media.
- Greenberg, A. (2017). The WannaCry Ransomware Hackers Made Some Real Amateur Mistakes. Wired. <https://www.wired.com/2017/05/wannacry-ransomware-hackers-made-real-amateur-mistakes/>.
- Haizler, O. (2017). The United States' Cyber Warfare History: Implications on Modern Cyber Operational Structures and Policymaking in Cyberspace, Intelligence, and Security. Vol 1. Nr.1. The Institute for Natural Security Studies. <https://www.inss.org.il/wp-content/uploads/2017/03/The-United-States'-Cyber-Warfare-History-Implications-on.pdf>
- Kili, A. (2019). How to Generate a CSR (Certificate Signing Request) in Linux. Tecmint <https://www.tecmint.com/generate-csr-certificate-signing-request-in-linux>
- Klimburg-Witjes, N. & Wentland, A. (2021). "Hacking Humans? Social Engineering and the Construction of the Deficient User in Cybersecurity Discourses", Science, Technology, & Human 46(6). 1316-1339. SAGE Journals.
- Mike, C. & David, S, "Cryptography and the Public Key Infrastructure," in CompTIA Security+ Study Guide: Exam SY0-601, Wiley, 2021, pp.179-227.
- Monnappa, K. A. (2018). Learning Malware Analysis: Explore the Concepts, Tools, and Techniques to Analyze and Investigate Windows Malware. Packt Publishing Ltd.
- Nash, A., William, D. & Celia, J. (2001) PKI Implementing and Managing e-Security. McGraw-Hill.

- Paar, C. & Pelzl J. (2010). *Understanding Cryptography: a Textbook for Students and Practitioners*. Springer.
- Peterson, A. (2014). The Sony Pictures hack, explained. The Washington Post: <https://www.washingtonpost.com/news/the-switch/wp/2014/12/18/the-sony-pictures-hack-explained/>
- Rad, B.B., Masrom, M. & Ibrahim, S. (2012) Camouflage in Malware: From Encryption to Metamorphism. *International Journal of Computer Science Network. Security*. 12 (74–83).
- Robertson, J. & Turton, W. (May 8, 2021). "Colonial Hackers Stole Data Thursday Ahead of Shutdown". Bloomberg News.
- Spafford, E.H. (1988). The Internet Worm Program: An Analysis. Purdue Technical Report CSD-TR-823. <https://spaf.cerias.purdue.edu/tech-reps/823.pdf>
- Stallings, W. (2017). *Cryptography and Network Security: Principles and Practice*. Pearson.
- Sporx. (2024). AnyDesk hacklendi mi? Anydesk hack nedir? <https://www.sporx.com/anydesk-hacklendi-mi-anydesk-hacked-nedir-SXHBQ1056445SXQ>
- Stallings, W. (2017). *Cryptography and Network Security: Principles and Practice*. Pearson.
- Tasiopoulos, V.G. & Katsikas, S.K. (2014). Bypassing Antivirus Detection with Encryption. In *Proceedings of the 18th Panhellenic Conference on Informatics*.
- Taylor, C. (2020). Melissa Virus. CyberHoot. <https://cyberhoot.com/cybrary/melissa-virus/>
- Zetter, K. (2014) An Unprecedented Look at Stuxnet, the World's First Digital Weapon. Magazine Wired. <https://www.wired.com/2014/11/countdown-to-zero-day-stuxnet/>

Android Güvenlik Açıklarının Modellenmesi: İstatistiksel Dağılımlardan Analizler¹

Modeling Android Security Vulnerabilities: Insights from Statistical Distributions

Kerem GENCER² 

Fatih BASCIFTCI³ 

DOI:10.33461/uybisbbd.1524207

Öz

Makale Bilgileri

Makale Türü:

Araştırma Makalesi

Geliş Tarihi:

31.07.2024

Kabul Tarihi:

05.09.2024

©2024 UYBISBBD
Tüm hakları saklıdır.



Android işletim sistemi, multimedya özelliklerini destekleyen bir mobil işletim sistemidir. Android, ses, video, resim ve diğer multimedya içeriklerini oynatmak, kaydetmek, düzenlemek ve paylaşmak için çok çeşitli uygulamalar ve entegre özellikler sunar. Çoğu Android cihazda kamera, hoparlör, mikrofon ve diğer multimedya bileşenleri bulunur. Yazılım güvenliğinde, güvenlik açıkları genellikle yazılım geliştirme sırasında ortaya çıkan kritik endişelerdir. Bu güvenlik açıklarını sürümden sonra tahmin etmek, risk değerlendirmesi ve azaltma için önemlidir. Çeşitli modeller araştırılmış olsa da Android işletim sistemi nispeten keşfedilmemiş durumdadır. Bu çalışma, yaygın olarak kullanılan Alhazmi-Malaiya Lojistik (AML) modeline uygunluklarını karşılaştırarak, farklı istatistiksel dağılımlar kullanarak Android güvenlik açıklarını modellemeyi araştırmaktadır. 2016'dan 2018'e kadar uzanan Ulusal Güvenlik Açığı Veritabanı'ndan (NVD) alınan veriler ve Ortak Güvenlik Açığı Puanlama Sistemi (CVSS) puanları analiz edilmiştir. Çalışma, aylık güvenlik açığı sayıları ve ortalama aylık etki değerleri için Lojistik, Weibull, Nakagami, Gamma ve Log-lojistik dahil olmak üzere çeşitli dağıtım modellerini değerlendirir. Model sağlamlığı değerlendirmesi için uyum iyiliği testleri ve bilgi kriterleri uygulandı. Bulgular, araştırmacılar ve Android yazılım geliştiricileri için değerli içgörüler sunarak tahmin, risk değerlendirmesi, kaynak tahsisi ve araştırma yönüne yardımcı olur. Ortalama aylık etki değerleri ve aylık güvenlik açığı sayıları için sırasıyla lojistik ve Nakagami dağılımları en uygun modeller olarak ortaya çıkmıştır. Son olarak, istatistiksel yöntemler, anlaşılabilirlik, veri miktarı, hesaplama ihtiyacı ve veri bağımsızlığı gibi esnek özellikleri nedeniyle küçük veri kümeleri veya daha net tanımlanmış veriler için bilinen yapay zekâ yöntemlerine karşı daha iyi performans gösterir.

Anahtar Kelimeler: İstatistiksel dağılımlar, Android güvenlik açıkları, Yazılım güvenliği, Güvenlik açığı keşif modeli.

Abstract

Article Info

Paper Type:

Research Paper

Received:

31.07.2024

Accepted:

05.09.2024

©2024 UYBISBBD
All rights reserved.



Android operating system is a mobile operating system that supports multimedia features. Android offers a wide range of applications and integrated features for playing, recording, editing and sharing audio, video, images and other multimedia content. Most Android devices include cameras, speakers, microphones, and other multimedia components. In software security, vulnerabilities are critical concerns that often emerge during software development. Predicting these vulnerabilities post-release is essential for risk assessment and mitigation. While various models have been explored, the Android operating system remains relatively uncharted. This study delves into modeling Android security vulnerabilities using different statistical distributions, comparing their suitability to the widely-used Alhazmi-Malaiya Logistic (AML) model. Data from the National Vulnerability Database (NVD) spanning 2016 to 2018, along with Common Vulnerability Scoring System (CVSS) scores, was analyzed. The study evaluates several distribution models, including Logistic, Weibull, Nakagami, Gamma, and Log-logistic, for monthly vulnerability counts and average monthly impact values. Goodness-of-fit tests and information criteria were applied for model robustness assessment. The findings offer valuable insights for researchers and Android software developers, aiding prediction, risk assessment, resource allocation, and research direction. Logistic and Nakagami distributions emerged as the best-fit models for average monthly impact values and monthly vulnerability counts, respectively. Finally, statistical methods perform better against known artificial intelligence methods for small data sets or more clearly defined data due to their flexible features such as comprehensibility, amount of data, need for calculation, and data independence.

Keywords: Statistical distributions, Android vulnerabilities, Software security, Vulnerability discovery model.

Atf/ to Cite (APA): Gencer, K. & Basciftci, F. (2024). Modeling Android Security Vulnerabilities: Insights from Statistical Distributions. International Journal of Management Information Systems and Computer Science, 8(2), 110-126. DOI: 10.33461/uybisbbd.1524207

¹"Ulusal yazılım açıklık veri tabanı oluşturulması kapsamında android açıklıklarının modellenmesi ve analiz edilmesi" isimli Selçuk Üniversitesinde Bilgisayar Mühendisliği alanında yapılan Doktora tezinden üretilmiştir.

² Dr. Öğretim Üyesi, Afyon Kocatepe Üniversitesi, Mühendislik Fakültesi, keremgencer09@hotmail.com, Afyonkarahisar, Türkiye.

³ Prof. Dr., Selçuk Üniversitesi, Teknoloji Fakültesi, basciftci@selcuk.edu.tr, Konya, Türkiye.

1. INTRODUCTION

Knowing the security vulnerabilities inside the lifecycle of a software program might provide the means to evaluate, reduce, and even remove the risks these vulnerabilities produce. Though a software security growth model can be used during software creation, this may not stop security vulnerabilities from appearing in the software. Using a prediction model, vulnerability discovery prediction is as essential as vulnerability detection. Recently, many security vulnerability discovery models have been developed. In general, vulnerability discovery models are divided into time- and effort-based categories. In this study, the average monthly impact value and the monthly vulnerability count for Android are modeled for the first time. Until now, five probability distributions as flexible as the Weibull distribution: Normal, Logistic, Log-logistic, Nakagami, and Gamma distributions, have been tried alongside the Weibull distribution, which is often used in this kind of modeling, and their performances have been compared. These distributions are symmetrical and asymmetrical, i.e., skewed distributions. Probability density functions (pdf) of the distributions in question, cumulative distribution functions, goodness-of-fit tests, and measurement criteria were all compared to find the best model.

These results can guide researchers and software developers interested in Android vulnerabilities in several ways:

- **Prediction and Risk Assessment:** These results can be used to predict better the future impacts and probabilities of vulnerabilities on the Android operating system. This is important for developing strategies to combat vulnerabilities and better understanding potential risks.
- **Software Development:** Using these results, software developers can focus on safer coding practices for Android applications or operating systems. Applying a specific distribution or estimating the propagation rate of vulnerabilities can improve software security.
- **Resource Allocation:** These results can be used to allocate information security resources effectively. Understanding which vulnerabilities require more resources or further action can help use the budget more effectively.
- **Research Direction:** These results can help determine the direction of future security research. These findings about which statistical distributions better model a particular security environment can be a basis for future research. In this study, some advantages of statistical methods are as follows.
- **Understandability:** Statistical methods allow for a more straightforward interpretation of results. Therefore, it is more accessible to people who want to understand vulnerabilities and discover their causes. At the same time, these methods can determine more clearly which factors affect the likelihood of security vulnerabilities.
- **Amount of Data:** Statistical methods can work with more limited data. Complex AI methods such as deep learning often require large data sets, while statistical methods can deal with smaller data sets.
- **Computational Need:** Statistical methods may not extensively use computational resources such as deep learning. This can result in faster results at lower costs.
- **Independence:** Statistical methods are generally less data-dependent. They can be more flexible, especially when new data sets or updates arrive because retraining or adapting the model is less complex.

Some studies on vulnerability detection models are shown in Figure 1 (Movahedi, 2019). These studies are divided into two categories, time-based and effort-based, and the time-based models are further subdivided into three categories: Quasi-Linear, SGRM-based, and S-shape. It is understood that the focus of these studies was on time-based studies rather than effort-based studies.

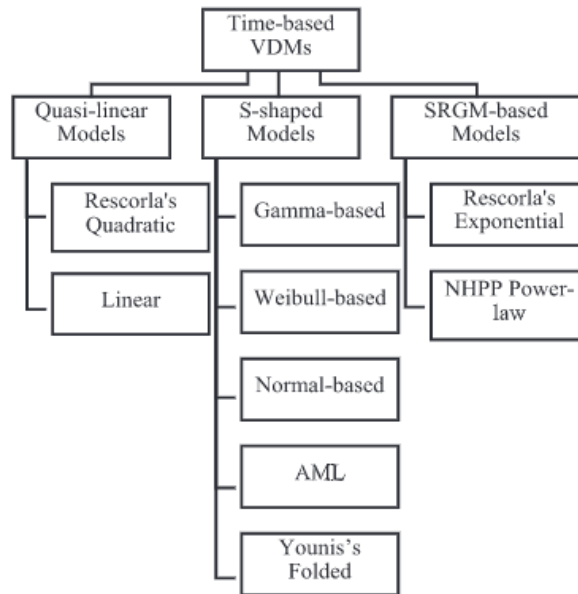


Figure 1. Taxonomy of Vulnerability Discovery Models

With this study – since the primary goal of modeling is to forecast – the most similar alternative distributions were compared. In Section 2, previous studies on the subject are investigated in detail. Section 3 introduces the distributions used in this study. Section 4 compares the goodness-of-fit tests often used to see how well the sample data matches the expected distribution values. In Section 5, information on the data set and the method of the study is given, and it presents the values obtained from the fitness measures on the proposed distributions. Sections 6 and 7 are the discussion and conclusion sections, respectively.

2. LITERATURE REVIEW

In general, discovery models are divided into time- and effort-based categories. While time-based models use time, effort-based ones use environmental factors such as CPU utilization and load count as the independent variables. Time-based models are more frequently studied. When these studies are investigated, it is seen that the first study proposed as a full-fledged vulnerability discovery model was Anderson’s thermodynamics model (Anderson, 2002). However, this model was not sufficiently successful at detecting the weaknesses in various software. In later years, Alhazmi et al. conducted many studies on both time- and effort-based models (Alhazmi et al., 2005; Alhazmi and Malaiya, 2005a; 2005b; 2006a; 2006b; Alhazmi et al., 2007; Alhazmi and Malaiya, 2008). A statistical density-based model was developed by Rescorla (Rescorla, 2005). Woo et al. attempted to create a vulnerability discovery model on three popular web browsers (Woo et al., 2006a). They concluded that the model will be fixed when categorized according to the severity and the type of vulnerabilities. Also, Woo et al. conducted a study investigating Apache and IIS web server vulnerabilities (Woo et al., 2006b). Kim et al. proposed a model that searches for vulnerabilities in different software versions (Kim et al., 2007). Joh et al. proposed a Weibull distribution-based model that can be used when asymmetrical data sets (Joh et al., 2008). Chen et al. proposed a vulnerability discovery model that used a multi-loop method (Chen et al., 2010). Woo et al. observed that models cannot make good predictions if the obtained data does not feature trend changes (Woo et al., 2011). Ozment conducted a study on the limitations of vulnerability discovery models (Ozment, 2007), while Massacci and Nguyen investigated the available vulnerability discovery models in terms of quality and predictability (Massacci and Nguyen, 2014). Anand and Bhatt studied convex-shaped discovery models using five parameters and the weighted criterion method (Anand and Bhatt, 2016). Anand et

al. also developed a model for multi-version software (Anand et al., 2017). Bhatt et al. conducted a study on the relationship between vulnerabilities discovered recently and vulnerabilities found in the past (Bhatt et al., 2017). Kansal et al. developed a model that links the number of commercial software users (Kansal et al., 2018). In another study, Kansal et al. investigated the relationship between the operational coverage function and the expected vulnerability count with a generalized statistical model (Kansal et al., 2017). Johnston used the Bayesian method in their Ph.D. thesis on vulnerability discovery modeling (Johnston, 2018). Again, Johnston et al. conducted a study that connected the software release date and the security evaluation profile (Johnston et al., 2018). Rahimi and Zargham developed a model on code complexity and quality that does not require past vulnerability data (Rahimi and Zargham, 2013); however, since it was not possible to get the complete source code – as in other studies – the model could not be put to general use. Scandariato and Walden studied Android application vulnerabilities using support vector machines (Scandariato and Walden, 2012).

This study utilized source code and could only be used in open-sourced applications and was therefore limited because it couldn't be used on closed-source applications. Scandariato et al. used text mining in their studies on open-sourced Android applications (Scandariato et al., 2014). In their study, Gencer and Başçiftçi (2021) propose a model called F-CVSS (Fuzzy Common Vulnerability Scoring System) by combining fuzzy logic and logistic regression as an alternative to the traditional CVSS (Common Vulnerability Scoring System) system. They attempted to determine the relevant components with their investigations, and this method was successful in determining the appropriate features and finding vulnerabilities in them. Younis et al. investigated and modeled cases where the vulnerabilities occur asymmetrically (Younis et al., 2011). Wang et al. proposed an effort-based model, and they claimed that this model achieved better results than AML (Wang et al., 2019). Finally, Pokhrel et al. used time series, Artificial Neural Networks (ANNs), and Support Vector Machines (SVMs) to investigate desktop operating systems (Pokhrel et al., 2017). In their article, Gencer and Başçiftçi (2021) use ARIMA and deep learning methods to perform a time series analysis of vulnerabilities in the Android operating system. The study compares various time series modeling techniques to predict future trends of these vulnerabilities and to identify possible risks in advance. Movahedi et al. introduced an approach for predicting the cumulative number of software vulnerabilities with a neural network model. (Movahedi et al., 2019).

3. LIFECYCLE DISTRIBUTIONS USED IN MODELING ANDROID SOFTWARE VULNERABILITIES

Life analysis is the collection of all the statistical techniques used to analyze the data gathered while the model above was being created. Life analysis data sets are usually represented by classical statistical distributions such as Exponential, Gamma, Weibull, Log-normal, and Logistic (Nelson, 1982; Lawless, 2003; Lee and Wenyu, 2003; Kleinbaum and Klein, 2005; Machin et al., 2006). This section introduces lifecycle distributions, such as the Weibull, Gamma, Logistic, Log-logistic, Normal, and Nakagami, used to model the monthly counts and average monthly impact scores of Android vulnerabilities between 2016 and 2018. These flexible distributions are popularly used in reliability theory and adapt to many data sets.

3.1. Weibull Distribution

The Weibull distribution was proposed in 1939 by the physicist Waloddi Weibull, who gave his name to the distribution. As a flexible distribution, it is often used in engineering applications and modeling compounds, i.e., random variables. The Weibull distribution is also used in electronic circuits and to observe some biological organisms' decay rates. At the beginning of the 1970s, it began to be used in seismic risk analysis. The Weibull distribution became famous thanks to its usability in cases where the variable has a positive value, such as applications in the financial sector. Its probability density function $f(x)$ and distribution function $F(x)$ are given in Equations (1) and (2), respectively.

$$f(x) = \frac{\gamma}{\lambda} \left(\frac{x}{\lambda}\right)^{\gamma-1} e^{-\left(\frac{x}{\lambda}\right)^\gamma}, \quad \lambda, \gamma, x > 0 \tag{1}$$

$$F(x) = \left(1 - e^{-\lambda x}\right)^\gamma \tag{2}$$

Here, x , γ , and λ are the random variables representing the monthly average score (or the monthly vulnerability count), shape, and scale parameters.

3.2. Log-logistic Distribution

Log-logistic distribution is one of the alternatives to Weibull, a distribution with two parameters. If $Log(T)$ has a logistic distribution, the lifecycle T has a log-logistic distribution. This distribution successfully models data with tremendous and small values in some example series (Ahmad et al., 1988; Kantam et al., 2001). It is more successful than the Log-normal distribution in time series data with sudden changes (Shoukri et al., 1988). Its probability density function $f(x)$ and distribution function $F(x)$ are given in Equations (3) and (4), respectively.

$$f(x) = \frac{\gamma}{\lambda} \frac{\left(\frac{x}{\lambda}\right)^{\gamma-1}}{\left(1 + \left(\frac{x}{\lambda}\right)^\gamma\right)^2} \tag{3}$$

$$F(x) = \frac{1}{\left(1 + \left(\frac{x}{\lambda}\right)^{\gamma-1}\right)}, \quad x, \gamma, \lambda > 0 \tag{4}$$

Here, x and λ are the random variables representing the monthly average score (or the monthly vulnerability count), shape, and scale parameters, respectively.

3.3. Normal Distribution

Also known as the Gaussian distribution, the Normal distribution has practical applications in many areas. It is an essential continuous probability distribution family (Hogg and Craig, 1978). The Normal distribution has two parameters: the arithmetic mean, μ , and the variance, σ^2 . The probability density function $f(x)$ is shown in Equation (5) (Casella and Berger, 2001):

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R} \tag{5}$$

Here, x is the random variable representing the monthly average score or vulnerability count.

3.4. Gamma Distribution

This continuous probability distribution is used in probability theory and statistics using two parameters. The Gamma distribution is used to model the size of insurance demand and rainfall (Anderson and Darling, 1954; Boland, 2007). Its probability density function $f(x)$ is given in Equation (6):

$$f(x) = x^{k-1} \frac{e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)}, \quad x > 0, \quad k, \theta > 0 \tag{6}$$

Here, x , k , and λ are the random variables representing the monthly average score (or the monthly vulnerability count), shape, and scale parameters, respectively.

3.5. Logistic Distribution

The Logistic distribution is a continuous probability distribution. It is similar to the bell-curved Normal distribution in terms of shape, but it is flatter due to the larger weights at the tails. Its probability density function is known to be the square of a hyperbolic secant function (Decani and Stine, 1986). Its probability density function $f(x)$ is given in Equation (7):

$$f(x) = \frac{e^{-\frac{(x-\mu)}{s}}}{s \left(1 + e^{-\frac{(x-\mu)}{s}} \right)^2}, \quad x \in \mathbb{R} \tag{7}$$

Here, x , μ , and s are the random variables representing the monthly average score (or the monthly vulnerability count), location, and scale parameters, respectively.

3.6. Nakagami Distribution

The Nakagami distribution is commonly used to model right-skewed data sets with positive values. Though there have been many distributions that model radio signal weaknesses, such as Weibull and Log-normal, in 1960, Nakagami proposed this distribution instead (Nakagami, 1960). The Nakagami distribution has been the main focus of some studies thanks to its wide applicability compared to other popular distribution models (Türksen et al., 2015), and it is used in various areas. It has been observed to exhibit good performance in generating unit hydrographs used to predict flow rates in hydrology by Sarkar, Goel, and Mathur (Sarkar et al., 2009; 2010). Shankar et al. and Tsui et al. used it in medical imaging and for modeling ultrasound data, respectively (Shankar et al., 2005; Tsui et al., 2006). Kim and Latchman analyzed motion picture data using the Nakagami distribution (Kim and Latchman, 2009). Furthermore, Nakahara and Carcole showed the usability of the Nakagami distribution in seismic study modeling (Nakahara and Carcolé, 2010). The probability density function $f(x)$ of the Nakagami distribution is given in Equation (8):

$$f(x) = \frac{2m^m}{\Gamma(m)\Theta^m} x^{2m-1} \exp\left(m \log x^2 - \frac{mx^2}{\Theta}\right) x^{-1} \tag{8}$$

Here, $x > 0$, m , and Θ are the random variables representing the monthly average score (or the monthly vulnerability count), location, and scale parameters, respectively. In mathematics, the gamma function (Γ) is the generalization of the factorial function for complex and non-integer real numbers.

4. MODEL FITTING AND GOODNESS-OF-FIT ANALYSES

This section introduces three goodness-of-fit tests, which will be used to measure distribution fitness: Kolmogorov-Smirnov, Anderson-Darling, and Cramer-von Mises.

4.1. Kolmogorov-Smirnov Goodness-of-Fit Test

One of the goodness-of-fit tests used in this study is the Kolmogorov-Smirnov test (Kolmogorov, 1933). This tests the fitness of a data set on a statistical model. It is a method used

successfully among goodness-of-fit tests based on an experimental distribution function. It is known as the Kolmogorov-Smirnov (*K-S*) goodness-of-fit test in literature because Kolmogorov developed it but was first used in goodness-of-fit tests by Smirnov. In the *K-S* test, with x number of samples, the cumulative distribution function $F_O(x)$ is determined, which is assumed to be a fixed distribution. $S_n(x)$ is the experimental cumulative distribution function that gives the ratio of the values that are smaller than, or equal to, a value x across n observed samples. According to the main idea of the *K-S* test, if the experimental distribution function results are not close enough to the hypothetical $F_O(x)$ value, it is deduced that the observed data does not follow the theoretical distribution. In other words, the observed data do not fit the claimed distribution. The statistic to test this condition is shown below:

$$D = \max_x |F_0^*(X) - S_n(x)| \tag{9}$$

Where x is the sample count. The D statistic of Kolmogorov and Smirnov is entirely independent of the hypothetical distribution under test when $F_O(x)$ is continuous and fully known (Kolmogorov, 1933; Smirnov, 1939). The distribution of this statistic can be obtained when all the parameters are known. Otherwise, there is no distribution of the D statistic.

4.2. Anderson-Darling Goodness-of-Fit Test

Anderson and Darling proposed another test statistic by adapting the *K-S* test (Anderson and Darling, 1954). To determine this statistic, n unit samples $\{X_1, X_2, \dots, X_n\}$ are drawn from a batch whose probability function and probability function parameters are known. The null hypothesis for the Anderson-Darling test is built on the assumption that the samples come from a distribution determined entirely by the parameters. If the null hypothesis is rejected due to the test, it is deduced that the data do not fit the distribution determined by the parameters. This test was not created for specific distributions but all distributions whose parameters are known. Later, it was improved for cases with unknown parameters.

The Anderson-Darling test statistic is shown in Equation (10) where x , $F_O(x)$, and i are the sample count, the cumulative distribution function, which is assumed to be fixed, and the rank value.

$$A^2 = -\frac{2}{n} \sum_{i=1}^n \left[\begin{aligned} &\left(i - \frac{1}{2} \right) \log \{ F_0(x_{(i)}) \} \\ &+ \left(n - i + \frac{1}{2} \right) \log \{ 1 - F_0(x_{(i)}) \} \end{aligned} \right] - n \tag{10}$$

4.3. Cramer von Mises Goodness of Fit Test

The Cramer-von Mises goodness-of-fit test was proposed by Harald Cramer and Richard Edler Mises (Cramér, 1928). The Cramer-von Mises (W_n) test statistic is defined as follows:

$$W_n = \sum_{i=1}^n \left\{ F_0 \left(x_{(i)} - \frac{2i-1}{2n} \right) \right\}^2 + \frac{1}{12n} \tag{11}$$

Where x , n , $F_O(x)$ and I are the sample count, the random sample $\{X_1, X_2, \dots, X_n\}$, the cumulative distribution function, which is assumed to be fixed, and the

If the test statistic obtained for the observed value is larger than the table value, it shows that the data do not follow the distribution proposed.

5. DATA AND METHODOLOGY

The vulnerability data used in this study is taken from the National Vulnerability Database (NVD), the largest source in this area (NVD, 2019). NVD is a large-scope database formed by data gathered from companies located inside and outside America, with contributions from the United States government. It is the most preferred database in terms of its policies on widespread use and

public availability. The American National Security Agency supports the NVD project. Vulnerabilities announced by NVD receive a Common Vulnerability and Exposures (CVE) number. Hence, different numbers and re-announcements for the same exposure are prevented. The Android vulnerabilities were filtered out while the database was being formed (Cvedetails, 2019). Furthermore, Common Vulnerability System Scores (CVSSs) for Android vulnerabilities between the specified dates are grouped every month. The study goal was to model the monthly impact scores and the monthly vulnerability counts. After the data was gathered, Weibull, Logistic, Normal, Log-logistic, and Nakagami distributions were applied to obtain the monthly vulnerability impact scores. The goodness-of-fitness test results of these distributions are given in Table 1.

Table 1. National Vulnerability Database Average Score Goodness of Fits

Goodness of Fits	Distributions				
	Weibull	Logistic	Normal	Log-logistic	Nakagami
K-S Statistics	0.1461	0.0985	0.1427	0.1078	0.1502
A-D Statistics	1.0423	0.5936	0.9677	0.7442	1.0603
CVM Statistics	0.1851	0.0821	0.1604	0.0944	0.1763
K-S (<i>p</i> -value)	0.4263	0.8757	0.4561	0.7972	0.3911
A-D (<i>p</i> -value)	0.3353	0.6529	0.3740	0.5220	0.3266
CVM (<i>p</i> -value)	0.2992	0.6828	0.3607	0.6158	0.3195

It was observed that the *p*-values of all the distributions under investigation were larger than 0.05. However, the purpose of this study was not just to find the distributions that model the monthly average scores but to find the distribution that models it best (*p*-value>0.05). Nevertheless, according to the Kolmogorov-Smirnov, Anderson-Darling and Cramer-von Mises test statistics, the best distribution is observed to be the Logistic distribution (*p*-value>0.05). Furthermore, Weibull, Logistic, Log-logistic, Gamma and Nakagami distributions were applied to model the monthly vulnerability counts. The data on these distributions are given in Table 2.

Table 2. National Vulnerability Database Monthly Count Goodness of Fit

Goodness of Fit	Distributions				
	Weibull	Logistic	Log-logistic	Gamma	Nakagami
K-S Statistics	0.1085	0.1412	0.1058	0.1302	0.1137
A-D Statistics	0.2817	0.7853	0.6940	0.4537	0.2782
CVM Statistics	0.0457	0.1316	0.0753	0.0775	0.0475
K-S (<i>p</i> -value)	0.7908	0.4697	0.8149	0.5748	0.7409
A-D (<i>p</i> -value)	0.9509	0.4908	0.5628	0.7932	0.9533
CVM (<i>p</i> -value)	0.9044	0.4525	0.7226	0.7097	0.8942

The most successful distribution was identified by looking at the Kolmogorov-Smirnov, Anderson-Darling and Cramer-von Mises test statistics and the *p*-values. It was observed that the *p*-values of all the distributions under investigation were larger than 0.05. However, the purpose of this study was not just to find the distributions that model the monthly average scores but to find the distribution that models it best (*p*-value>0.05). Accordingly, the Kolmogorov-Smirnov, Anderson-Darling and Cramer-von Mises test statistics indicated that the best distribution was seen to be the Nakagami distribution (*p*-value>0.05).

5.1. Comparison of Vulnerability Discovery Models

The $-2\text{Log } L$ statistic is one of the metrics used to decide a suitable lifecycle model. The most suitable model is the model with the lowest value (Klein and Moeschberger, 1997). Akaike proposed the Akaike Information Criterion (AIC) to compare different models, and this is defined as follows:

$$\text{AIC} = -2\ln L + 2k \tag{12}$$

where *k* is the number of model parameters (Akaike, 1974).

In this equation, $\ln L$ and k are the log-likelihood and the parameter count, respectively. The smallest AIC value is used to decide the best model (Cavanaugh, 1997). Another information criterion that is widely used in literature is the Bayesian or Schwarz Information Criterion (BIC). This is defined as follows (McLachlan and Peel, 2001):

$$BIC = -2\ln L + k(\log(n)) \tag{13}$$

In this equation, $\ln(n)$ is the natural logarithm of the sample volume n . The k and n symbols represent the number of parameters and the sample size. Again, the smallest BIC value is used to decide the best model (Hurvich and Tsai, 1989; Cavanaugh, 1997; Ucal, 2006). In Figure 2, the pdf of the best distributions that model the monthly average scores of NVD are given. It is seen that the best distribution is the Logistic distribution. After that, the sequence of the most suitable distributions in terms of fitness was Log-logistic, Normal, Nakagami and Weibull.

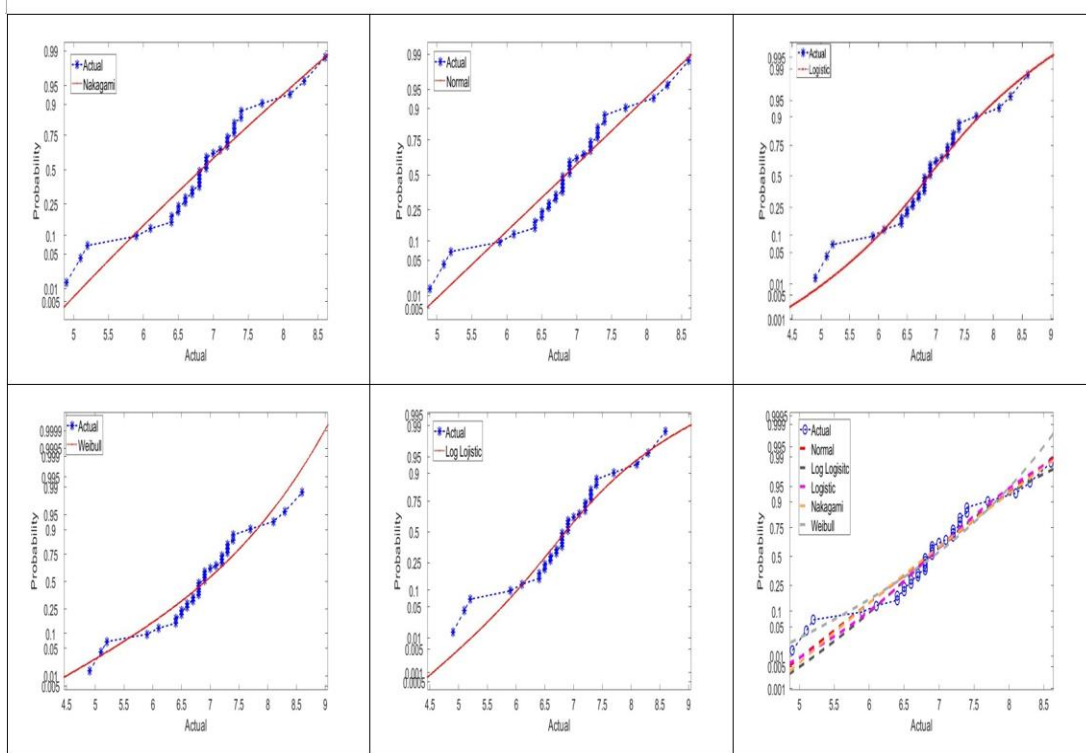


Figure 2. National Vulnerability Database Average Score Probability Density Functions

Furthermore, the cumulative distribution functions are given in Figure 3 below.

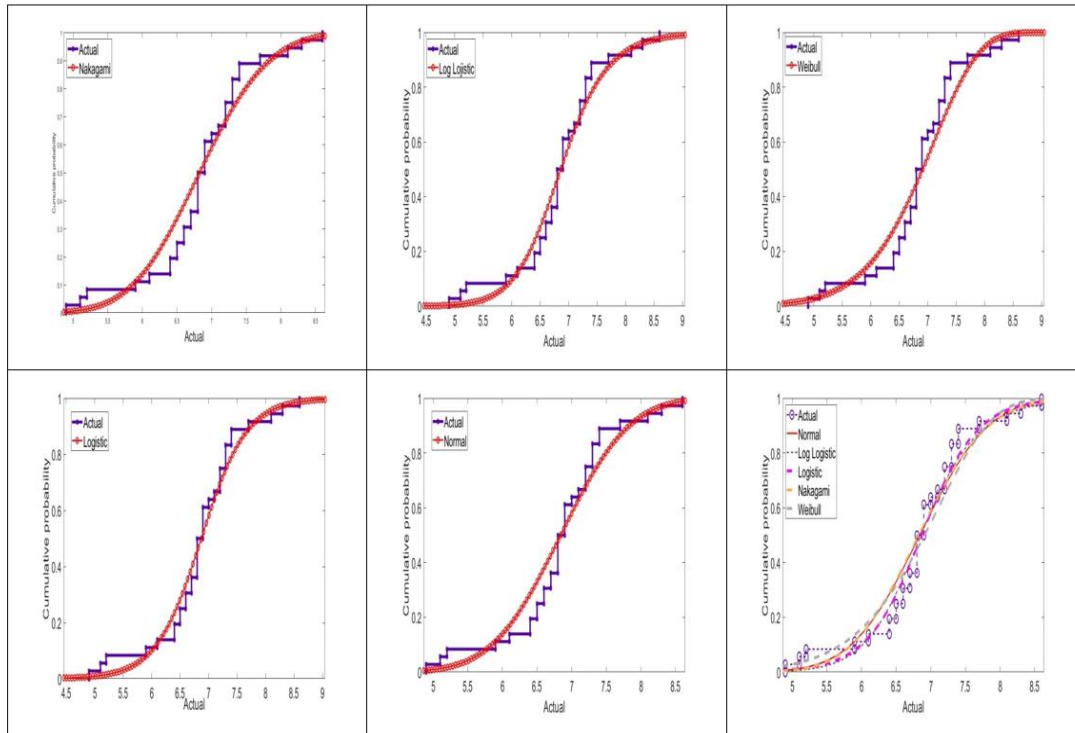


Figure 3. National Vulnerability Database Monthly Count Cumulative Distribution Functions

In Table 3, known model fitting metrics, AIC, BIC and $-2\text{Log}L$ have been used for model prediction. According to these metrics, the Logistic distribution had the smallest AIC, BIC and $-2\text{Log}L$ values. After that, the sequence of the remaining distributions was Log-logistic, Normal, Nakagami and finally Weibull in terms of their values.

Table 3. National Vulnerability Database Average Score Model Fitting

Model Fitting	Distributions				
	Weibull	Logistic	Normal	Log-logistic	Nakagami
LogL	-41.7878	-39.9638	-41.3863	-40.8412	-41.7680
$-2\text{Log}L$	83.5755	79.9275	82.7726	81.6823	83.5360
AIC	87.5755	83.9275	86.7726	85.6823	87.5360
BIC	90.7426	87.0945	89.9396	88.8494	90.7030

In Table 4, known model fitting metrics, AIC, BIC and $-2\text{Log}L$ have been used to find the model that best predicts the monthly vulnerability counts. According to these metrics, the Nakagami and Logistic distributions have the smallest and the largest AIC, BIC and $-2\text{Log}L$ values, respectively.

Table 4. National Vulnerability Database Monthly Count Model Fitting

Model Fitting	Distributions				
	Weibull	Logistic	Log-logistic	Gamma	Nakagami
LogL	-173.6286	-175.9572	-176.3994	-174.2972	-173.6244
$-2\text{Log}L$	347.2572	351.9144	352.7989	348.5945	347.2487
AIC	351.2572	355.9144	356.7989	352.5945	351.2487
BIC	354.4242	359.0814	359.9659	355.7615	354.4157

In Table 5, the parameters, the lower and upper bounds of these parameters and the standard errors of these parameters are given for the Weibull, Logistic, Normal, Log-logistic and Nakagami distributions that model the monthly average impact values. With the parameter values obtained, the monthly average impact value – determined as a random variable – can be predicted.

Table 5. National Vulnerability Database Average Score Parameters

Parameters	Distributions				
	Weibull	Logistic	Normal	Log-logistic	Nakagami
α	7.1734	6.8721	6.8417	6.8607	19.7059
β	9.8417	0.3978	0.7639	16.8118	47.3919
Lower Bound (α)	6.9218	6.6529	6.5921	6.6368	10.6784
Lower Bound (β)	7.5001	0.2850	0.5875	12.0231	43.9045
Upper Bound (α)	7.4250	7.0913	7.0912	7.0845	28.7333
Upper Bound (β)	12.1833	0.5107	0.9403	21.6004	50.8794
Standard Error (α)	0.1284	0.1118	0.1273	0.1142	4.6059
Standard Error (β)	1.1947	0.0576	0.0900	2.4432	1.7793

α : The shape parameter

β : The scale parameter

In Table 6, the parameters, the lower and upper bounds for these parameters, and the standard errors for these parameters are given for the Weibull, Logistic, Log-logistic, Gamma and Nakagami distributions that model the monthly vulnerability counts. With the parameter values obtained, the monthly vulnerability count – determined as a random variable – can be predicted.

Table 6. National Vulnerability Database Monthly Count Parameters

Parameters	Distributions				
	Weibull	Logistic	Log-logistic	Gamma	Nakagami
α	61.5340	52.4923	47.9460	2.3463	0.7845
β	1.7183	17.9225	2.4614	23.4290	3948.1608
Lower Bound (α)	49.2384	42.3484	37.0752	1.3300	0.4722
Lower Bound (β)	1.2800	13.0212	1.7724	12.1173	2537.3748
Upper Bound (α)	73.8297	62.6361	58.8169	3.3627	1.0969
Upper Bound (β)	2.1565	22.8237	3.1504	34.7408	5358.9468
Standard Error (α)	6.2734	5.1755	5.5465	0.5185	0.1594
Standard Error (β)	0.2236	2.5007	0.3516	5.7714	719.8020

α : The shape parameter

β : The scale parameter

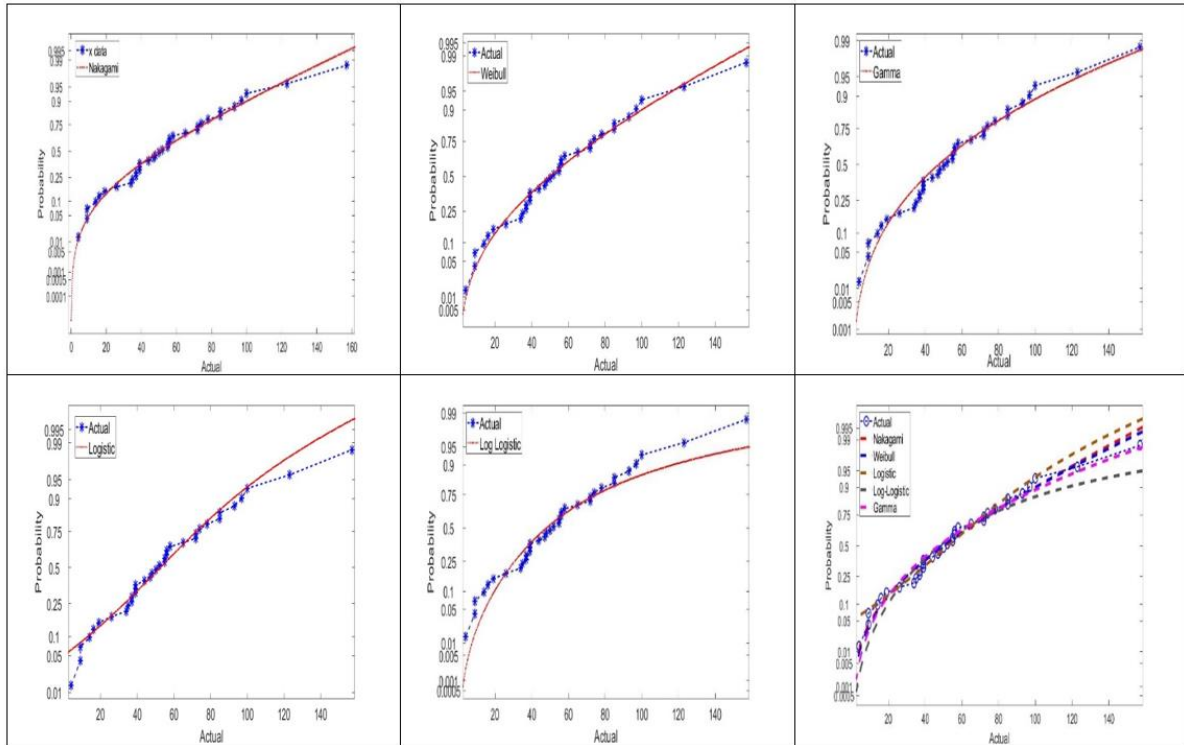


Figure 4. National Vulnerability Database Monthly Count Probability Density Functions

Figure 4 shows the pdf of the distributions that best model the monthly counts of NVD. It is seen that the best distribution is the Nakagami, followed by the Weibull distribution.

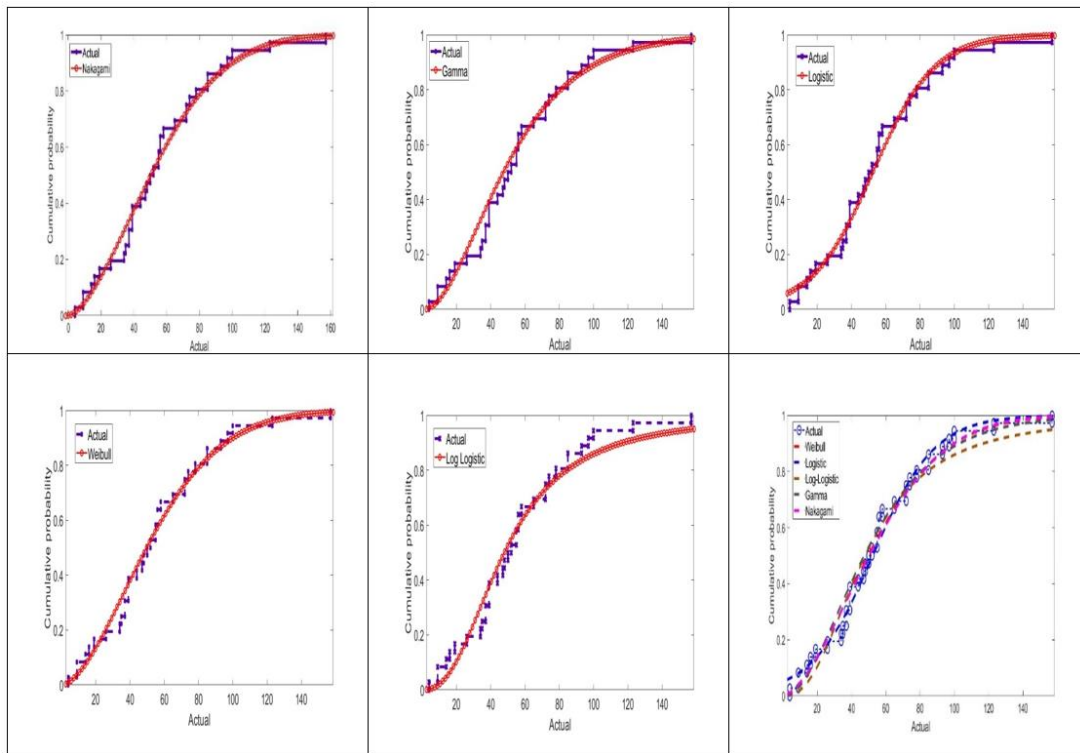


Figure 5. National Vulnerability Database Monthly Average Score Cumulative Distribution Functions

After those, the remaining distribution sequence is Gamma, Logistic and Log-logistic in terms of fitness. Furthermore, the cumulative distribution functions are given in Figure 5.

6. DISCUSSION

This study investigated the symmetrical and asymmetrical Weibull, Logistic, Log-logistic, Normal, Gamma and Nakagami distributions that best model the monthly Android vulnerability scores and monthly vulnerability counts. In Table 1, showing the fitness comparisons of monthly average vulnerability score models, the largest (K-S) p values were 0.8757 and 0.3911 for the Logistic and Nakagami distributions, which are the best and the worst models, respectively. In Table 3, the $-2\text{Log}L$, AIC and BIC values for the Logistic distribution were observed to be 79.9275, 83.9275 and 87.0945, respectively. In Table 2, showing the fitness comparisons of monthly vulnerability count models, the largest (K-S) p values were 0.7409 and 0.4697 for the Nakagami and Logistic distributions, which are the best and the worst models, respectively. In Table 3, $-2\text{Log}L$, AIC and BIC values for the Nakagami distribution were observed to be 347.2487, 351.2487 and 354.4157, respectively. While the monthly average vulnerability scores were observed to be symmetrical, monthly vulnerability counts were observed to be more skewed. With the help of certain criteria, the predictive ability and fitness of these six distributions were compared. The predictive ability was measured by calculating the smallest AIC, BIC and $2\text{Log}L$ values. For the fitness criteria, Kolmogorov-Smirnov, Anderson-Darling and Cramer-von Mises statistical tests were applied. All the distributions were observed to represent the data sets well. However, the Logistic and Nakagami distributions were observed to best model the monthly average vulnerability scores and monthly vulnerability counts.

7. CONCLUSION

Vulnerability detection models help us forecast software vulnerabilities and enable the necessary precautions to be taken, such as planning the generation of a patch. In this study, the best distribution was determined by modeling the continuous vulnerabilities of the Android operating system from 2016 to 2018 with different statistical distributions. As a result of this study, it was seen that data sets usually modeled with Weibull distribution in published literature can also be modeled with different distributions. Android vulnerabilities have been best modeled by Logistic and Nakagami distributions for average monthly scores and monthly Android vulnerability counts, respectively. Goodness-of-fit tests have shown the fitness of these distributions. This study has proposed a new aspect of time-based vulnerability discovery models regarding statistical distributions. With suitable distributions, it has been shown that Android vulnerabilities can be modeled, and forecasts can be made. This study shows that these distributions are promising for accurate predictions of software vulnerability disclosures and results are helpful in academia and industry. As a result, it is aimed that analysts prioritize their work by taking into account the severity of the potential risks arising from Android vulnerabilities.

REFERENCES

- Ahmad, M. I., Sinclair, C. D. and Werritty, A., 1988, Log-Logistic Flood Frequency Analysis, *Journal Of Hydrology*, 98 (3), 205-224.
- Akaike, H., 1974, A New Look At The Statistical Model Identification, *Ieee Transactions On Automatic Control*, 19 (6), 716-723.
- Alhazmi, O., Malaiya, Y. Ve Ray, I., 2005, Security Vulnerabilities In Software Systems: A Quantitative Perspective, *Data And Applications Security Xix*, Berlin, Heidelberg, 281-294.

- Alhazmi, O. H. and Malaiya, Y. K., 2005a, Modeling The Vulnerability Discovery Process, *16th Ieee International Symposium On Software Reliability Engineering (Issre'05)*, Ten Pp.-138.
- Alhazmi, O. H. and Malaiya, Y. K., 2005b, Quantitative Vulnerability Assessment Of Systems Software, *Annual Reliability And Maintainability Symposium, 2005. Proceedings*, 615-620.
- Alhazmi, O. H. and Malaiya, Y. K., 2006a, Measuring And Enhancing Prediction Capabilities Of Vulnerability Discovery Models For Apache And Iis Http Servers, *17th International Symposium On Software Reliability Engineering*, 343-352.
- Alhazmi, O. H. and Malaiya, Y. K., 2006b, Prediction Capabilities Of Vulnerability Discovery Models, *Rams '06. Annual Reliability And Maintainability Symposium, 2006.*, 86-91.
- Alhazmi, O. H., Malaiya, Y. K. and Ray, I., 2007, Measuring, Analyzing And Predicting Security Vulnerabilities In Software Systems, *Computers & Security*, 26 (3), 219-228.
- Alhazmi, O. H. and Malaiya, Y. K., 2008, Application Of Vulnerability Discovery Models To Major Operating Systems, *Ieee Transactions On Reliability*, 57 (1), 14-22.
- Anand, A. and Bhatt, N., 2016, Vulnerability Discovery Modeling And Weighted Criteria Based Ranking, *Journal Of The Indian Society For Probability And Statistics*, 17 (1), 1-10.
- Anand, A., Das, S., Agrawal, D. Ve Klochkov, Y., 2017, Vulnerability Discovery Modelling For Software With Multi-Versions, In: *Advances In Reliability And System Engineering*, Eds: Ram, M. Ve Davim, J. P., Cham: Springer International Publishing, P. 255-265.
- Anderson, R., 2002, Security In Open Versus Closed Systems -The Dance Of Boltzmann, Coase And Moore, *Open Source Software Economics*, 127-142.
- Anderson, T. W. and Darling, D. A., 1954, A Test Of Goodness Of Fit, *Journal Of The American Statistical Association*, 49 (268), 765-769.
- Bhatt, N., Anand, A., Yadavalli, V. S. S. and Kumar, V., 2017, Modeling And Characterizing Software Vulnerabilities, *International Journal Of Mathematical, Engineering And Management Sciences*, 2 (4), 288-299.
- Boland, P. J., 2007, *Statistical And Probabilistic Methods In Actuarial Science, Usa*, Taylor & Francis Inc, P. 43.
- Casella, G. and Berger, R. L., 2001, *Statistical Inference Usa*, Duxbury, P. 102.
- Cavanaugh, J. E., 1997, Unifying The Derivations For The Akaike And Corrected Akaike Information Criteria, *Statistics & Probability Letters*, 33 (2), 201-208.
- Chen, K., Feng, D.-G., Su, P.-R., Nie, C.-J. and Zhang, X.-F., 2010, Multi-Cycle Vulnerability Discovery Model For Prediction, *Journal Of Software*, 21 (9), 2367-2375.
- Cramér, H., 1928, On The Composition Of Elementary Errors, *Scandinavian Actuarial Journal*, 1928 (1), 141-180.
- Cvedetails, 2019, <https://www.cvedetails.com/browse-by-date.php>, [Accessed Date: 10 June 2024].
- Decani, J. S. and Stine, R. A., 1986, A Note On Deriving The Information Matrix For A Logistic Distribution, *The American Statistician*, 40 (3), 220-222.
- Gencer, K. and Başçiftçi, F. 2021, Time series forecast modeling of vulnerabilities in the android operating system using ARIMA and deep learning methods. *Sustainable Computing: Informatics and Systems*, 30, 100515.
- Gencer, K. and Başçiftçi, F. 2021, The fuzzy common vulnerability scoring system (F-CVSS) based on a least squares approach with fuzzy logistic regression. *Egyptian Informatics Journal*, 22(2), 145-153.

- Hogg, R. V. and Craig, A. T., 1978, Introduction To Mathematical Statistics *Newyork*, Macmillan, P. 109.
- Hurvich, C. M. and Tsai, C.-L., 1989, Regression And Time Series Model Selection In Small Samples, *Biometrika*, 76 (2), 297-307.
- Joh, H., Kim, J. and Malaiya, Y. K., 2008, Vulnerability Discovery Modeling Using Weibull Distribution, *2008 19th International Symposium On Software Reliability Engineering (Issre)*, 299-300.
- Johnston, R., 2018, A Multivariate Bayesian Approach To Modeling Vulnerability Discovery In The Software Security Lifecycle, Ph.D, *George Washington University*, Washington, Dc, Usa, 55-65.
- Johnston, R., Sarkani, S., Mazzuchi, T., Holzer, T. and Eveleigh, T., 2018, Multivariate Models Using Mcmc Bayes For Web-Browser Vulnerability Discovery, *Reliability Engineering & System Safety*, 176, 52-61.
- Kansal, Y., Kapur, P. K., Kumar, U. and Kumar, D., 2017, User-Dependent Vulnerability Discovery Model And Its Interdisciplinary Nature, *Life Cycle Reliability And Safety Engineering*, 6 (1), 23-29.
- Kansal, Y., Kapur, P. K. and Kumar, U., 2018, Coverage-Based Vulnerability Discovery Modeling To Optimize Disclosure Time Using Multiattribute Approach, *Quality And Reliability Engineering International*, 35 (1), 62-73.
- Kantam, R. R. L., Rosaiah, K. and Rao, G. S., 2001, Acceptance Sampling Based On Life Tests: Log-Logistic Model, *Journal Of Applied Statistics*, 28 (1), 121-128.
- Kim, J., Malaiya, Y. K. and Ray, I., 2007, Vulnerability Discovery In Multi-Version Software Systems, *10th Ieee High Assurance Systems Engineering Symposium (Hase'07)*, 141-148.
- Kim, K. and Latchman, H. A., 2009, Statistical Traffic Modeling Of Mpeg Frame Size: Experiments And Analysis. *Journal Of Systemics, Cybernetics And Informatics*, 7 (6), 54-59.
- Klein, J. P. and Moeschberger, M. L., 1997, Survival Analysis Techniques For Censored And Truncated Data, *Newyork*, Springer, P. 277.
- Kleinbaum, D. G. and Klein, M., 2005, Survival Analysis: A Self-Learning Text, *Usa*, Springer, P. 590.
- Kolmogorov, A. N., 1933, Sulla Determinazione Empirica Di Una Legge Di Distribuzione, *G. Ist. Attuari*, 83-91.
- Lawless, J. F., 2003, Statistics Models And Methods For Lifetime Data, *New Jersey*, John Wiley & Sons, P. 630.
- Lee, E. T. and Wenyu, J. W., 2003, Statistical Methods For Survival Data Analysis, *Newyork*, John Wiley & Sons, P. 513.
- Machin, D., Cheung, Y. B. and Parmar, M., 2006, Survival Analysis: A Practical Approach, *England*, John Wiley & Sons, P. 266.
- Massacci, F. and Nguyen, V. H., 2014, An Empirical Methodology To Evaluate Vulnerability Discovery Models, *Ieee Transactions On Software Engineering*, 40 (12), 1147-1162.
- Mclachlan, G. and Peel, D., 2001, Finite Mixture Model, *Newyork*, Wiley, P. 419.
- Movahedi, Y., Cukier, M. and Gashi, I., 2019, Vulnerability Prediction Capability: A Comparison Between Vulnerability Discovery Models And Neural Network Models, *Computers & Security*, 87, 1-10.

- Nakagami, M., 1960, The M-Distribution—A General Formula Of Intensity Distribution Of Rapid Fading, In: *Statistical Methods In Radio Wave Propagation*, Eds: Hoffman, W. C.: Pergamon, P. 3-36.
- Nakahara , H. and Carcolé, E., 2010, Maximum-Likelihood Method For Estimating Coda Q And The Nakagami-M Parameter, *Bulletin Of The Seismological Society Of America*, 100 (6), 3174-3182.
- Nelson, W. B., 1982, *Applied Life Data Analysis*, Canada, John Wiley & Sons, P. 634.
- Nvd, 2019, <https://nvd.nist.gov/> [Accessed Date: 10 June 2024].
- Ozment, A., 2007, Improving Vulnerability Discovery Models. Proceedings Of The 2007 Acm Workshop On Quality Of Protection. Alexandria, Virginia, Usa, Acm: 6-11.
- Pokhrel, N. R., Rodrigo, H. and Tsokos, C. P., 2017, Cybersecurity: Time Series Predictive Modeling Of Vulnerabilities Of Desktop Operating System Using Linear And Non-Linear Approach, 8 (4), 362-382.
- Rahimi, S. and Zargham, M., 2013, Vulnerability Scrying Method For Software Vulnerability Discovery Prediction Without A Vulnerability Database, *Ieee Transactions On Reliability*, 62 (2), 395-407.
- Rescorla, E., 2005, Is Finding Security Holes A Good Idea?, *Ieee Security & Privacy*, 3 (1), 14-19.
- Sarkar, S., Goel, N. K. and Mathur, B. S., 2009, Adequacy Of Nakagami- M Distribution Function To Derive Gih, *Journal Of Hydrologic Engineering*, 14 (10), 1070-1079.
- Sarkar, S., Goel, N. K. and Mathur, B. S., 2010, Performance Investigation Of Nakagami- M Distribution To Derive Flood Hydrograph By Genetic Algorithm Optimization Approach, *Journal Of Hydrologic Engineering*, 15 (8), 658-666.
- Scandariato, R. and Walden, J., 2012, Predicting Vulnerable Classes In An Android Application. Proceedings Of The 4th International Workshop On Security Measurements And Metrics. Lund, Sweden, Acm: 11-16.
- Scandariato, R., Walden, J., Hovsepyan, A. and Joosen, W., 2014, Predicting Vulnerable Software Components Via Text Mining, *Ieee Transactions On Software Engineering*, 40 (10), 993-1006.
- Shankar, P. M., Piccoli, C. W., Reid, J. M., Forsberg, F. and Goldberg, B. B., 2005, Application Of The Compound Probability Density Function For Characterization Of Breast Masses In Ultrasound B Scans, *Physics In Medicine And Biology*, 50 (10), 2241-2248.
- Shoukri, M. M., Mian, I. U. H. and Tracy, D. S., 1988, Sampling Properties Of Estimators Of The Log-Logistic Distribution With Application To Canadian Precipitation Data, *Canadian Journal Of Statistics*, 16 (3), 223-236.
- Smirnov, N., 1939, On The Estimation Of The Discrepancy Between Emprical Curves Of Distribution For Two Independent Samples, *Bulletin Mathématique De L'Université De Moscow*, 2 (2), 3-11.
- Tsui, P.-H., Huang, C.-C. and Wang, S.-H., 2006, Use Of Nakagami Distribution And Logarithmic Compression In Ultrasonic Tissue Characterization, *Journal Of Medical And Biological Engineering*, 26 (2), 69.
- Türksen, I. B., Khaniyev, T. and Gokpinar, F., 2015, Investigation Of Fuzzy Inventory Model Of Type (S, S) With Nakagami Distributed Demands, *Journal Of Intelligent & Fuzzy Systems*, 29 (2), 531-538.
- Ucal, M. Ş., 2006, Ekonometrik Model Seçim Kriterleri Üzerine Kısa Bir İnceleme, *C.Ü. İktisadi Ve İdari Bilimler Fakültesi*, 7 (2), 41-57.

- Wang, X., Ma, R., Li, B., Tian, D. and Wang, X., 2019, E-Wbm: An Effort-Based Vulnerability Discovery Model, *Ieee Access*, 7, 44276-44292.
- Woo, S.-W., Alhazmi, O. and Malaiya, Y., 2006a, An Analysis Of The Vulnerability Discovery Process In Web Browsers. Proceeding Of The 10th Iasted International Conferance Software Engineering And Applicaitons. Usa: 172-177.
- Woo, S.-W., Joh, H., Alhazmi, O. H. and Malaiya, Y. K., 2011, Modeling Vulnerability Discovery Process In Apache And Iis Http Servers, *Computers & Security*, 30 (1), 50-62.
- Woo, S., Alhazmi, O. H. and Malaiya, Y. K., 2006b, Assessing Vulnerabilities In Apache And Iis Http Servers, *2006 2nd Ieee International Symposium On Dependable, Autonomic And Secure Computing*, 103-110.
- Younis, A. A., Joh, H. and Malaiya, Y. K., 2011, Modeling Learningless Vulnerability Discovery Using A Folded Distribution, *The 2011 International Conference On Security And Management*, Usa, 1-10.

Süper Yapay Zeka Devrimleriyle Yönetim Bilişim Sistemlerinin Evrimi

Evolution of Management Information Systems by Super Artificial Intelligence Revolutions

DOI:10.33461/uybisbbd.1521086

Ahmet EFE¹ 

Öz

Makale Bilgileri

Makale Türü:
Araştırma Makalesi

Geliş Tarihi:
23.07.2024

Kabul Tarihi:
09.09.2024

©2024 UYBISBBD
Tüm hakları saklıdır.



Bu çalışma, süper yapay zeka (YZ) devriminin Yönetim Bilişim Sistemleri (YBS) disiplini üzerindeki etkisini incelemektedir. Üniversitelerde hızla YZ bölümleri kurulurken, süper YZ devriminin YBS alanını önemli ölçüde dönüştürdüğü kabul edilmektedir. Bu çalışma, YBS ile iş analitiği, bilgisayar bilimi, yönetim bilimi, yazılım mühendisliği ve yapay zeka gibi diğer ilgili disiplinler arasındaki karmaşık dinamikleri modellemek için bir fonksiyon formülü önermeyi amaçlamaktadır. Önerilen formül, süper YZ devriminin yeni zorluklar ve fırsatlar nasıl tanıttığını, ilgili alanların yakınsama ve uzaklaşmasını nasıl etkilediğini ve YBS disiplinini nasıl etkilediğini yakalayacaktır. Ayrıca, çalışma, süper YZ ile ilişkili temel endişeleri ve teknik sorunları belirleyerek bu zorlukları ele almak için potansiyel hafifletme stratejileri sunmaktadır. Disiplinler arası işbirliğinin önemi ve uzmanlaşmış becerilerin edinilmesi gerekliliği vurgulanarak, bu çalışma, profesyonellerin süper YZ devrimiyle şekillenen evrimsel manzarada etkin bir şekilde gezinmelerine duyulan ihtiyacı vurgulamaktadır.

Anahtar Kelimeler: Yönetim Bilişim Sistemleri, Süper YZ, Yapay Zeka, Yakınsama, Uzaklaşma.

Abstract

Article Info

Paper Type:
Research Paper

Received:
23.07.2021

Accepted:
09.09.2024

©2024 UYBISBBD
All rights reserved.



This study explores the impact of the super artificial intelligence (AI) revolution on the evolution of Management Information Systems (MIS) discipline. As AI departments are being set rapidly in all universities worldwide, recognizing that the super AI revolution has significantly transformed the MIS field, this study aims to propose a function formula to model the intricate dynamics between MIS and other related disciplines, such as business analytics, computer science, management science, software engineering, and artificial intelligence. The proposed formula will capture how the super AI revolution introduces new challenges and opportunities, influences the convergence and divergence of related fields, and affects the MIS discipline. Additionally, the study identifies key concerns and technical issues associated with super AI, offering potential mitigation strategies to address these challenges. By emphasizing the importance of interdisciplinary collaboration and the necessity for acquiring specialized skills, this study underscores the need for professionals to effectively navigate the evolving landscape shaped by the super AI revolution.

Keywords: Management Information Systems, Super AI, Artificial Intelligence, Convergence, Divergence.

Atıf/ to Cite (APA): Efe, A. (2024). Evolution of Management Information Systems by Super Artificial Intelligence Revolutions. International Journal of Management Information Systems and Computer Science, 8(2), 127-142. DOI: 10.33461/uybisbbd.1521086

¹ Dr., International Federation Of Red Cross And Red Crescent (IFRC), ahmet.efe@ifrc.org, Ankara, Türkiye.

1. INTRODUCTION

The swift and profound advancements in artificial intelligence (AI) have markedly reshaped numerous fields, with Management Information Systems (MIS) being no exception. At the forefront of this transformation is the advent of super AI—an advanced form of artificial intelligence with the potential to surpass human cognitive capabilities across various dimensions. This research is dedicated to examining how the emergence of super AI has driven the evolution of the MIS discipline, investigating both its transformative effects and the consequent implications for the field.

The primary objective of this study is to scrutinize the ways in which the MIS discipline has been reshaped by the super AI revolution. This involves a detailed exploration of the interdisciplinary connections between MIS and related domains such as computer science, management science, software engineering, and artificial intelligence. By identifying and addressing key concerns and technical challenges introduced by super AI, the study aims to map out the evolving landscape of MIS. It also introduces a functional model incorporating dependent and independent variables to elucidate the factors influencing changes within the discipline. Additionally, the research proposes strategies to mitigate these challenges and address technical issues effectively.

This study's significance lies in its focus on how super AI accelerates the evolution of MIS, presenting both new opportunities and obstacles. Unlike previous research, which may have concentrated on isolated aspects of AI's impact, this study provides a comprehensive analysis of the convergence and divergence between MIS and other disciplines in the context of super AI. By offering novel insights into these dynamics and proposing actionable solutions, this research contributes uniquely to our understanding of the evolving interplay between AI advancements and MIS.

1.1. Research Methodology

This study investigates the impact of the super AI revolution on the evolution of the MIS discipline through a multi-faceted research approach. The methodology involves proposing a function formula to model the complex interactions between MIS and related fields such as business analytics, computer science, management science, software engineering, and artificial intelligence. This formula aims to capture how super AI introduces new challenges and opportunities, influences the convergence and divergence of related disciplines, and affects the MIS discipline.

To validate the proposed formula, the study employs a combination of theoretical analysis and empirical evidence. Theoretical insights are drawn from current literature on AI's impact across various domains, including recent studies on AI's role in cybersecurity (Efe, 2021), data analytics (Aydemir & Yavuz, 2019), and machine learning applications (Takaoglu & Özer, 2019). Empirical data is gathered through case studies and industry reports to identify key concerns and technical issues associated with super AI, which are then used to formulate potential mitigation strategies.

The research further emphasizes the importance of interdisciplinary collaboration and the acquisition of specialized skills, reflecting the evolving landscape shaped by the super AI revolution. By integrating recent findings and addressing gaps in existing literature, this study provides a comprehensive analysis of how super AI transforms the MIS field and suggests strategies for professionals to navigate these changes effectively.

1.2. Conceptual Definitions

Super Artificial Intelligence (Super AI): Super AI, also known as Artificial Superintelligence (ASI), refers to a level of artificial intelligence that surpasses human intelligence and capability in all aspects, including creativity, problem-solving, and emotional intelligence (Bostrom, 2014). Super AI can outperform the most gifted human minds in every field of endeavor, leveraging vast computational power, advanced algorithms, and the ability to learn and adapt autonomously. This form of AI not only understands and interprets complex data but also anticipates and innovates

beyond human expectations, leading to unprecedented advancements and potentially profound impacts on society, technology, and various academic disciplines.

Management Information Systems (MIS) Discipline: The MIS discipline focuses on the study and application of information technology to support and improve business processes, decision-making, and organizational performance (Laudon & Laudon, 2016). It encompasses the design, implementation, management, and use of information systems to collect, process, store, and disseminate information within an organization. MIS integrates principles from business management, computer science, and information technology to ensure that information systems align with business goals and strategies. Key areas within MIS include systems analysis and design, database management, network infrastructure, information security, and business analytics, all aimed at optimizing the efficiency and effectiveness of organizational operations.

1.3. Research Problem

The research problem centers on examining the key concerns, technical issues, and potential effects of the super AI revolution on the MIS discipline. It also explores how the convergence and divergence of related disciplines are impacting the evolution of MIS.

The MIS discipline has been critically questioned and placed in a state of crisis due to the rapid establishment of various disciplinary foundations and scientific branches across universities. The emergence of related fields such as data science, machine learning, and computer science has introduced significant changes and challenges for MIS. These include a pronounced skill gap, conflicts between disciplines, and organizational hurdles. The expanding scope of expertise required in these interconnected domains has made it difficult for professionals to stay abreast of swift technological advancements, creating challenges in talent acquisition and retention (Bragg, 2021; Chen et al., 2019; Kane et al., 2017).

1.4. Interdisciplinary Conflicts

The integration of various disciplines within the MIS field can lead to interdisciplinary conflicts, as professionals from different backgrounds may have contrasting approaches and priorities (Legner et al., 2017). This can create challenges in the development and implementation of effective information systems and strategies, as well as in communication and collaboration among team members (Lu et al., 2018).

The convergence and divergence of related disciplines in the MIS domain can create organizational challenges, such as adapting to new technologies, managing change, and aligning organizational goals with the changing landscape (Sebastian et al., 2017). Organizations must also address ethical, and security concerns associated with the use of advanced technologies like AI and machine learning, which can have significant implications on the design and management of information systems (Mithas et al., 2021).

1.5. Research Assumptions

1. Super AI has a significant impact on the MIS discipline as does on others.
2. The evolution of MIS is influenced by the convergence and divergence of related disciplines.
3. The super AI revolution presents both challenges and opportunities for the MIS discipline.
4. All MIS systems are in the cloud environment and run with web-based interfaces.
5. MIS has a close relationship with Computer Science Discipline, Management Science Discipline, Software Engineering Discipline, Artificial Intelligence Discipline and Industrial Engineering Discipline.

1.6. Research Hypothesis

The super AI revolution has accelerated the evolution of the MIS discipline, leading to the emergence of new challenges and opportunities, as well as the convergence and divergence of related disciplines.

1.7. Research Limitations

This study's limitations include the reliance on theoretical models that may not fully capture real-world complexities and the challenge of predicting long-term impacts due to the rapidly evolving nature of super AI technology:

1. **Scope of Analysis:** The study's focus on super AI's impact on MIS is constrained by the rapid pace of technological advancements. This limitation may affect the relevance of findings as new developments in AI and related disciplines emerge.
2. **Interdisciplinary Complexity:** The complex dynamics between MIS and other disciplines such as computer science and business analytics are challenging to model precisely. The proposed function formula may oversimplify these interactions, limiting the accuracy of predictions.
3. **Empirical Validation:** The reliance on case studies and industry reports for empirical validation may introduce bias and limit the generalizability of the results. Different industries may experience the effects of super AI in varied ways.
4. **Ethical and Security Concerns:** Addressing ethical and security issues associated with super AI is inherently complex. The study may not fully capture the nuances of these concerns, particularly as new ethical dilemmas and security threats arise.

2. LITERATURE DISCUSSIONS

The advent of super AI promises transformative impacts on the structure and functionality of MIS applications. A primary advantage of super AI lies in its capacity to swiftly and accurately process vast volumes of data, thus facilitating more informed decision-making and improved organizational outcomes (Arun, 2021). Additionally, super AI's ability to automate routine tasks enables the reallocation of human resources towards more strategic endeavors (Liang et al., 2020).

Recent studies have expanded on these benefits by exploring specific applications of super AI within MIS. For instance, Efe (2021) discusses how AI-focused cybersecurity can enhance risk management frameworks, highlighting AI's role in safeguarding data integrity. Similarly, Aydemir and Yavuz (2019) examine the use of AI in analyzing seasonal drug sales data, illustrating AI's potential to uncover complex patterns that improve business insights. These studies underscore how super AI not only automates but also refines data analysis processes.

Super AI's impact extends to the accuracy and efficiency of MIS applications through advanced machine learning algorithms. For example, AI-driven natural language processing (NLP) enhances the precision of text-based data inputs, such as customer feedback or social media content (Jiang et al., 2020). Additionally, deep learning algorithms advance image recognition and analysis, beneficial for applications like security monitoring and product inspection (Liang et al., 2020). Kekül, Bircan, and Arslan (2018) provide further insights into the performance of facial recognition technologies, showcasing their relevance in security contexts.

Another significant contribution of super AI to MIS is its capacity for sophisticated predictive analytics. By employing advanced machine learning models, organizations can discern trends and patterns, leading to informed decisions and future predictions (Jiang et al., 2020). Recent research by Talan (2021) highlights how AI-driven bibliometric analyses can forecast educational trends, demonstrating the predictive power of AI in various fields.

Despite these advancements, there remains a dearth of research specifically addressing super AI's effects on the convergence and divergence of related disciplines within MIS. This study aims to bridge this gap by offering a comprehensive analysis of how the super AI revolution influences the MIS discipline, addressing both its integration with and divergence from other fields.

The intersection of MIS with related disciplines, such as data science, computer science, and business analytics, is a focal point of contemporary discourse. Data science, with its emphasis on deriving insights from large datasets, has significantly influenced MIS practices. This interdisciplinary synergy has fostered the development of innovative data-driven tools and techniques (Dhar, 2013).

The relationship between computer science and MIS also merits attention. Computer science provides foundational principles for algorithm and data structure development, essential for advancing MIS software and hardware systems (Laudon & Laudon, 2016). The continued evolution of super AI is likely to strengthen this collaboration, focusing on creating more efficient and intelligent systems (Brynjolfsson & McAfee, 2014).

Business analytics, emphasizing quantitative techniques for data analysis, has become integral to organizational decision-making (Davenport & Harris, 2007). The MIS discipline leverages these methods to enhance predictive modeling and resource optimization (Sharda et al., 2013). This interdisciplinary collaboration is expected to persist as organizations seek competitive advantages through data-driven strategies (Piccoli & Pigni, 2013).

Therefore, the MIS field has experienced significant interdisciplinary interactions with data science, computer science, and business analytics. This collaborative approach has enabled the MIS discipline to evolve in response to the challenges and opportunities presented by the super AI revolution. The integration of recent studies into this discourse further enriches our understanding of super AI's impact on MIS and related fields.

3. POSSIBLE EFFECTS OF SUPER AI OVER MIS DISCIPLINE

The super AI revolution presents several key concerns and technical problems for the MIS discipline. These include the ethical and security implications of using super AI, the risk of job displacement, and the need for new skill sets in the workforce. Additionally, super AI can also lead to an over-reliance on AI-driven decision-making, which may reduce human input and critical thinking in organizational processes.

Ethical Implications: Super AI can lead to ethical concerns, such as privacy invasion, biased decision-making, and lack of accountability (Bostrom, 2014). Organizations implementing super AI systems must carefully consider these issues to ensure that their use aligns with their ethical principles (Mittelstadt et al., 2016).

Security Risks: The integration of super AI into MIS may increase the vulnerability of systems to cyber attacks, data breaches, and other security threats (Cavelty & Mauer, 2018). The security of super AI-driven MIS should be prioritized to protect sensitive organizational data and maintain trust in the systems.

Job Displacement: The increasing capabilities of super AI could lead to job displacement, particularly in roles that involve routine tasks and data analysis (Frey & Osborne, 2017). Organizations and policymakers must develop strategies to mitigate the potential negative impact on the workforce and create new opportunities for re-skilling and up-skilling.

Over-reliance on AI-driven Decision-making: Super AI can facilitate more efficient and accurate decision-making, but there is a risk of over-reliance on AI systems, which may reduce human input and critical thinking (Davenport & Ronanki, 2018). Organizations should find a balance between AI-driven and human decision-making to optimize their processes.

Skill Gaps in the Workforce: The rise of super AI necessitates a shift in the skills required for professionals working in the MIS discipline (Brynjolfsson & McAfee, 2014). Organizations and educational institutions should collaborate to develop relevant training programs and curricula that address these new skill requirements.

The key mitigation options for the concerns, technical problems, and possible effects of Super AI over the MIS discipline include addressing ethical considerations, implementing robust security measures, focusing on workforce development, and maintaining a balance between human and AI-driven decision-making processes.

Addressing Ethical Considerations: It is crucial to establish ethical guidelines and frameworks for the development and deployment of Super AI in MIS applications (Floridi & Cowls, 2019). These guidelines can help organizations navigate potential ethical dilemmas, such as data privacy and algorithmic fairness, ensuring that Super AI systems are transparent, accountable, and unbiased.

Implementing Robust Security Measures: Super AI systems may present new security risks, such as adversarial attacks and data breaches. Organizations must invest in cutting-edge security measures, including encryption, intrusion detection systems, and secure software development practices (Buczak & Guven, 2016). Regular security audits and risk assessments can help identify and address potential vulnerabilities.

Focusing on Workforce Development: The rise of Super AI may lead to job displacement and demand for new skill sets. Organizations must invest in workforce development and training programs, helping employees adapt to the changing landscape (Arntz et al., 2016). Additionally, educational institutions should revise curricula to incorporate AI-related skills and competencies, ensuring that the next generation of MIS professionals is well-equipped to navigate the Super AI revolution.

Maintaining a Balance Between Human and AI-Driven Decision-Making Processes: While Super AI can enhance decision-making processes in the MIS discipline, it is essential to maintain a balance between human input and AI-driven solutions (Davenport & Ronanki, 2018). This can help avoid over-reliance on AI systems and ensure that human critical thinking remains a vital component of organizational decision-making.

4. CONVERGENCE AND DIVERGENCE OF OTHER RELATED DISCIPLINES

The super AI revolution has led to the convergence of disciplines such as data science, machine learning, and computer science with the MIS discipline. This has resulted in a richer and more diverse field, with new tools and techniques being developed to address complex business challenges. However, the rapid advancements in AI technology have also resulted in the divergence of some disciplines, as certain aspects of the MIS domain become more specialized.

The MIS discipline is an interdisciplinary field that involves the integration of various disciplines to support organizational decision-making processes. According to Schultze and Leidner (2002), MIS incorporates various fields such as computer science, operations research, and organizational behavior. This interdisciplinary approach enables MIS to leverage the strengths of different fields to develop comprehensive solutions to complex business problems.

In addition to its interdisciplinary nature, MIS also exhibits transdisciplinary and multidisciplinary relationships with other disciplines. According to Nambisan et al. (2019), transdisciplinarity refers to the integration of knowledge from multiple disciplines to develop a holistic understanding of a particular issue. In contrast, multidisciplinary refers to the use of knowledge from multiple disciplines without necessarily integrating them.

MIS exhibits transdisciplinary relationships with fields such as data science, information technology, and business analytics. Data science, for example, provides techniques and tools for

analyzing large data sets, while information technology supports the development and implementation of information systems. Business analytics, on the other hand, enables organizations to make data-driven decisions by providing insights from large data sets (Maier et al., 2013).

In terms of multidisciplinary relationships, MIS draws on fields such as economics, psychology, and sociology. Economics provides insights into the financial implications of information systems, while psychology and sociology contribute to the understanding of human behavior in organizational contexts (Laudon & Laudon, 2016).

4.1. Relationship with Computer Science Discipline

The relationship between the MIS discipline and computer science has been evolving as the super AI revolution unfolds. While both fields share some common ground, such as the use of algorithms, data structures, and programming languages, they differ in their primary focus and objectives. MIS deals with the management, analysis, and use of information within organizations, whereas computer science focuses on the design, development, and implementation of computational systems (Chen, 2006).

With the emergence of super AI, the convergence between computer science and MIS has become more apparent. Advanced AI techniques, such as machine learning and natural language processing, have been incorporated into MIS processes to improve decision-making, automation, and data analysis (Hevner et al., 2004). This integration has led to a more holistic approach to information management, where computer science principles are combined with organizational and managerial aspects to address complex business challenges (Bharadwaj et al., 2013).

However, the rapid advancements in AI technology have also resulted in the divergence of some aspects of computer science and MIS. As AI systems become increasingly sophisticated, certain subfields within computer science have become more specialized, creating a gap between the technical expertise required for developing AI systems and the managerial skills needed for effectively leveraging them in an organizational context (Luftman et al., 2020).

The relationship between computer science and the MIS discipline has been significantly influenced by the super AI revolution. While there is a growing convergence of these fields due to the integration of advanced AI techniques into MIS processes, certain aspects of computer science have also diverged because of rapid technological advancements. To maximize the benefits of this evolving relationship, organizations and academia must foster collaboration between computer scientists and MIS professionals, ensuring a balance between technical and managerial expertise (Holsapple & Singh, 2000).

Therefore, as super AI systems integrate deeply into cloud-based MIS environments, they introduce complexities that can both disrupt and advance the field. The convergence of super AI with MIS, while potentially accelerating technological advancements, also raises critical concerns. These systems could lead to heightened vulnerabilities in data security and system integrity, as super AI's advanced capabilities might outpace traditional MIS safeguards, creating new attack vectors. Moreover, the evolutionary trajectory of MIS, shaped by the super AI revolution, presents a dual-edged sword; while offering opportunities for enhanced analytics and decision-making processes, it also risks widening the gap between MIS and foundational computer science principles. This divergence could destabilize established methodologies and frameworks, as super AI introduces novel paradigms that challenge conventional MIS approaches. Consequently, the research hypothesis that super AI accelerates MIS evolution, presenting both challenges and opportunities, underscores the necessity for interdisciplinary collaboration. Bridging the gaps between MIS and computer science disciplines will be crucial to address emerging risks, adapt existing models, and leverage super AI's potential while mitigating its inherent threats.

4.2. Relationship with Management Science Discipline

The relationship between the Management Information Systems (MIS) discipline and management science discipline has been significantly influenced by the super AI revolution. Management science, focused on the application of mathematical and statistical techniques to decision-making processes, has seen a growing convergence with the MIS discipline as super AI technologies have been increasingly integrated into organizational processes (Davenport & Ronanki, 2018).

Super AI technologies have enhanced the decision-making capabilities of management science by providing advanced analytical tools, predictive modeling, and optimization techniques (Brynjolfsson & McAfee, 2014). In turn, the MIS discipline has benefited from the quantitative and analytical rigor of management science, leading to more effective and data-driven decision-making processes within organizations (Chui et al., 2016).

The convergence of management science and MIS disciplines has resulted in the development of new interdisciplinary fields, such as data-driven decision-making and business analytics (Davenport, 2013). These emerging fields are characterized by the integration of super AI technologies with traditional management science methodologies, promoting more efficient and effective management practices.

However, the rapid advancements in super AI technologies have also led to some divergence between management science and the MIS discipline. The specialized nature of certain aspects of super AI may result in the development of sub-disciplines within both fields, which may focus on specific applications or techniques (Bostrom, 2014). This divergence could lead to a potential fragmentation of knowledge and expertise, requiring professionals to acquire specialized skills and knowledge in their respective domains (Agrawal et al., 2018).

Therefore, as super AI continues to advance, it fundamentally reshapes MIS by introducing both unprecedented challenges and opportunities. Super AI's integration into MIS can enhance data analytics and decision-making capabilities, driving the discipline towards greater efficiency and innovation. However, this rapid evolution also risks exacerbating the divergence between MIS and Management Science as traditional methodologies may become obsolete in the face of AI-driven approaches. The convergence of super AI with MIS necessitates a re-evaluation of research assumptions and practices, as AI's ability to process vast amounts of data and automate complex decision-making processes may overshadow conventional management theories and practices. Additionally, as all MIS systems are increasingly cloud-based and web-oriented, there is an amplified risk of cybersecurity threats, data privacy issues, and the potential for systemic failures. This shifting landscape calls for a reassessment of research hypotheses, suggesting that while super AI accelerates the evolution of MIS, it simultaneously challenges the discipline's foundational principles and necessitates a careful balance between technological advancement and theoretical rigor.

4.3. Relationship with Software Engineering Discipline

The relationship between the MIS discipline and software engineering in the context of the super AI revolution is multifaceted, with both fields increasingly converging in some aspects while diverging in others (Nofal & Yusof, 2018). Super AI has the potential to transform traditional software engineering processes by introducing new methodologies, tools, and techniques that enable the development of more intelligent and efficient systems (Bhat, 2021).

The convergence of MIS and software engineering can be seen in areas such as data analysis, decision-making, and automation, where super AI technologies have been integrated into software development processes (Huang & Liu, 2020). For instance, AI-driven tools are being utilized in software testing, requirements engineering, and maintenance, contributing to improved software quality and reduced development times (Wang, & Zang, 2021). This convergence has led to the emergence of new interdisciplinary fields, such as AI-based software engineering, which combines the principles of software engineering with advanced AI techniques to build intelligent software systems (Zhang et al., 2019).

However, the rapid advancements in super AI technologies have also led to the divergence of certain aspects of the MIS and software engineering disciplines. For example, the increasing complexity of AI-driven systems may necessitate specialized expertise in areas such as machine learning, natural language processing, and computer vision, which might not have been traditionally associated with software engineering (Siau & Yang, 2017).

Additionally, the ethical, security, and privacy implications of super AI technologies have also emerged as key concerns, necessitating interdisciplinary collaboration between MIS and software engineering professionals to address these challenges (Bryson, 2020). This highlights the need for software engineers to possess a broader understanding of MIS concepts and principles, as well as for MIS professionals to develop a deeper knowledge of software engineering techniques and practices.

Therefore, the convergence of super AI with MIS and Software Engineering could lead to unprecedented levels of system efficiency, customization, and responsiveness, enabling more sophisticated data management and integration. However, this convergence also presents several risks. The complexity of super AI algorithms may exacerbate issues related to system transparency, making it increasingly difficult for software engineers to fully understand, validate, and troubleshoot AI-driven components within MIS frameworks. This lack of transparency can lead to challenges in ensuring system reliability, security, and compliance with regulatory standards. Moreover, the rapid evolution driven by super AI could outpace the development of corresponding software engineering methodologies and tools, creating a disconnect between the capabilities of AI systems and the traditional engineering practices that support them. As MIS systems increasingly rely on cloud environments and web-based interfaces, these challenges are compounded by the need to safeguard against vulnerabilities associated with cloud computing and distributed architecture. Therefore, while super AI offers transformative potential for enhancing MIS, it simultaneously necessitates a reevaluation of software engineering practices to address these emerging risks and ensure that the integrity and effectiveness of MIS systems are maintained in this evolving technological landscape.

4.4. Relationship with Artificial Intelligence Discipline

The relationship between the MIS discipline and the AI discipline has been a topic of interest for many researchers. As both fields continue to evolve, it is crucial to understand their interplay and the impact they have on each other (Alavi & Leidner, 2001).

AI is a branch of computer science that deals with the creation of intelligent agents, capable of learning, reasoning, and problem-solving (Russell & Norvig, 2016). The application of AI techniques in the context of MIS has resulted in significant advancements in decision-making processes, automation, and data analysis (Davenport & Harris, 2007).

The relationship between MIS and AI disciplines can be described as symbiotic, as advancements in AI have led to innovations in MIS, and vice versa. AI techniques have been integrated into various aspects of MIS, including data analytics, decision support systems, and knowledge management (Sharda et al., 2013).

One major area where the relationship between AI and MIS is evident is in the use of machine learning algorithms for data analysis and decision-making. Machine learning, a subfield of AI, provides methods for discovering patterns and trends in large data sets, which is critical for making informed decisions in an organizational context (Hastie et al., 2009).

Moreover, the development of AI-driven decision support systems has allowed organizations to harness the power of AI to solve complex problems and enhance their decision-making processes (Power, 2007). These systems combine AI techniques with the knowledge of domain experts, resulting in more accurate and efficient decision-making.

The relationship between the MIS and AI disciplines is an essential aspect of understanding the evolution of both fields. As the super AI revolution continues, it is expected that the synergy between

these disciplines will grow stronger, leading to further innovations and advancements (Brynjolfsson & McAfee, 2014).

Therefore, the convergence and divergence between MIS and AI disciplines are expected to intensify, with super AI accelerating this process. On one hand, super AI's advanced capabilities could significantly enhance MIS, providing unprecedented opportunities for efficiency, predictive analytics, and decision-making. However, this also introduces new risks, such as increased complexity in system integration and dependency on AI-driven processes, which could lead to heightened vulnerabilities and potential system failures. The shift towards cloud-based, web-oriented MIS systems amplifies these risks, as the reliance on interconnected and distributed environments makes them more susceptible to security breaches and operational disruptions caused by super AI's unpredictability. Additionally, the rapid evolution driven by super AI may outpace the ability of MIS professionals to adapt, potentially leading to gaps in expertise and governance challenges. The relationship between MIS and AI thus becomes increasingly intricate, requiring a delicate balance between leveraging super AI's benefits and mitigating its associated risks, while ensuring that the convergence and divergence of these disciplines are managed effectively to sustain system integrity and reliability.

4.5. Relationship with the Industrial Engineering Discipline

Industrial engineering (IE) and MIS have a close relationship, as both fields aim to improve organizational performance through the use of technology and data analysis. IE focuses on optimizing operational processes and systems to improve efficiency, while MIS focuses on using information systems to support decision-making and enhance organizational performance.

According to a study by Abbas et al. (2021), IE and MIS have a complementary relationship, where IE provides the necessary process improvement techniques and tools, while MIS provides the required information and data for effective decision-making. The study also highlights the importance of integrating both disciplines to achieve optimal results in organizations.

Moreover, research by Bandyopadhyay and Chakraborty (2018) shows that the integration of IE and MIS can lead to the development of decision support systems that aid in effective resource allocation, inventory management, and production planning. The study emphasizes the need for IE professionals to be knowledgeable in MIS to fully leverage the benefits of technology and data-driven decision-making.

Therefore, the convergence of super AI with MIS, which operates predominantly in cloud environments and relies on web-based interfaces, introduces a spectrum of risks and opportunities. Industrial Engineering, which focuses on optimizing complex systems and processes, may experience significant disruptions such as super AI reshapes operational efficiencies and decision-making paradigms.

The integration of super AI in MIS could lead to the divergence of traditional methodologies within Industrial Engineering, as super AI-driven analytics and automation challenge established process optimization techniques. This could potentially lead to obsolescence of some traditional practices, while creating new methodologies that leverage super AI's capabilities for real-time data processing and predictive analytics. The convergence, on the other hand, presents opportunities for enhanced collaboration between MIS and Industrial Engineering, facilitating the development of more sophisticated, data-driven models that can drive efficiency and innovation in industrial processes.

However, the accelerated evolution driven by super AI also poses risks such as dependency on complex AI algorithms that may not be fully understood or controlled, potentially leading to unexpected system behaviors and inefficiencies. Additionally, the shift towards AI-driven solutions may require Industrial Engineering professionals to adapt rapidly, acquiring new skills and knowledge to effectively integrate and manage these technologies. The transformation brought about

by super AI necessitates a reevaluation of existing frameworks and practices within both MIS and Industrial Engineering to address these emerging challenges and harness the opportunities for enhanced system performance and innovation.

5. DEVELOPMENT OF A FUNCTION FORMULA

To encapsulate the research conclusions in a mathematical model that represents the relationship between dependent and independent variables, we need to define the variables, parameters, and coefficients that capture the essence of the research assumptions and hypothesis. Here's a function formula with detailed explanations:

To encapsulate the research conclusions in a mathematical model that represents the relationship between dependent and independent variables, we need to define the variables, parameters, and coefficients that capture the essence of the research assumptions and hypothesis. Here's a function formula with detailed explanations:

$$E_{MIS} = \beta_0 + \beta_1 \cdot SAI + \beta_2 \cdot C_{disc} + \beta_3 \cdot D_{disc} + \beta_4 \cdot I_{AI} + \beta_5 \cdot R_{disc} + \beta_6 \cdot T_{adapt} + \epsilon$$

- E_{MIS} : Evolution of the MIS discipline (dependent variable)
- SAI : Super AI impact (independent variable)
- C_{disc} : Convergence of related disciplines (independent variable)
- D_{disc} : Divergence of related disciplines (independent variable)
- I_{AI} : Integration of AI techniques (independent variable)
- R_{disc} : Relationship strength with related disciplines (independent variable)
- T_{adapt} : Adaptation to new technologies and skills (independent variable)
- β_0 : Constant term (intercept)
- $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$: Coefficients representing the impact of each independent variable on E_{MIS}
- ϵ : Error term capturing unexplained variance

5.1. Detailed Explanations

1. β_0 : This constant term represents the baseline level of MIS evolution in the absence of the impacts of Super AI and other variables. It provides a starting point for understanding the evolution of the MIS discipline.

2. $\beta_1 \cdot SAI$: This term captures the effect of Super AI's impact on the evolution of the MIS discipline. Given that Super AI has a significant impact on MIS, the coefficient β_1 is expected to be positive, reflecting that increased Super AI impact accelerates MIS evolution.

3. $\beta_2 \cdot C_{disc}$: This term represents the impact of the convergence of related disciplines on the evolution of MIS. A positive β_2 suggests that as disciplines converge, MIS benefits from enhanced capabilities and efficiencies.

4. $\beta_3 \cdot D_{disc}$: This term measures the effect of the divergence of related disciplines. Divergence can drive specialization but may also introduce challenges. The coefficient β_3 could be positive or negative, depending on whether divergence leads to innovation or fragmentation.
5. $\beta_4 \cdot I_{AI}$: This term accounts for the integration of AI techniques into MIS. A positive β_4 indicates that increased AI integration fosters the evolution of MIS by introducing advanced methodologies and tools.
6. $\beta_5 \cdot R_{disc}$: This term measures the influence of the strength of relationships with related disciplines on MIS evolution. A positive β_5 suggests that stronger relationships enhance the evolution of MIS by fostering collaboration and cross-disciplinary advancements.
7. $\beta_6 \cdot T_{adapt}$: This term captures the role of adaptation to new technologies and skills. A positive β_6 reflects that effective adaptation contributes to the evolution of the MIS discipline by equipping professionals with the necessary competencies.
8. ϵ : The error term accounts for the variability in MIS evolution that is not explained by the included variables. It represents the influence of external factors or noise in the data.

5.2. Positive Elaborations

This formula and its components offer a comprehensive view of how different factors contribute to the evolution of the MIS discipline in the era of Super AI, aligning with the research assumptions and hypothesis:

- The model acknowledges that the Super AI revolution significantly impacts the MIS discipline, both accelerating its evolution and introducing new challenges and opportunities.
- The inclusion of convergence and divergence factors recognizes the dynamic interplay between disciplines, highlighting how integration and specialization drive MIS advancements.
- Emphasizing the integration of AI techniques and the importance of adaptation underscores the necessity for continuous skill development and technological adaptation.
- By incorporating the relationship strength with related disciplines, the model underscores the value of interdisciplinary collaboration in advancing the MIS field.

6. CONCLUSIONS

The advent of Super AI has profoundly influenced the evolution of MIS discipline, bringing both notable challenges and unprecedented opportunities. Our function formula, which integrates the impacts of Super AI, convergence, and divergence with related disciplines, supports the hypothesis that this technological revolution has accelerated MIS development. This acceleration is marked by significant changes in how MIS interacts with fields such as computer science, software engineering, and artificial intelligence.

Our findings validate the hypothesis that Super AI's impact is accelerating MIS evolution, as detailed in our function formula. The integration of advanced AI methodologies into MIS processes has opened new pathways for innovation, while also presenting complex challenges. Convergence among disciplines has led to enhanced capabilities and efficiencies, though it requires careful management of ethical, security, and technical concerns. Divergence has fostered specialization and innovation, but it also risks creating fragmentation and inconsistency.

The successful application of our function formula highlights the importance of interdisciplinary collaboration in addressing these changes. The interaction between MIS and related fields has spurred the emergence of new areas such as AI-based software engineering and data-driven decision-making. This synergy underscores the relevance of our hypothesis and the need for a multifaceted approach to navigating the evolving landscape.

Despite these advancements, the Super AI revolution brings forward critical issues, including ethical dilemmas, security risks, and potential job displacement. Addressing these concerns necessitates a collaborative effort between organizations and academia. It is crucial to focus on workforce development, establish robust ethical guidelines, and implement stringent security measures. Balancing human judgment with AI-driven decision-making will be essential to avoid over-reliance on automated systems.

Looking ahead, Super AI will continue to shape the MIS discipline, presenting both opportunities for growth and challenges that require proactive management. An interdisciplinary approach, integrating principles from computer science, organizational management, and other relevant fields, will be vital for tackling complex business problems in this new era. Organizations should emphasize collaboration, invest in employee training, and develop ethical frameworks to effectively navigate this transformative period.

Recommendations for Researchers, MIS Professionals, and Organizations:

1. Researchers: Focusing on empirical studies regarding Super AI's implications for MIS, particularly concerning ethics, security, and workforce development. Collaborate with organizations to identify and address emerging challenges and explore innovative solutions.

2. MIS Professionals: Developing expertise in data science, machine learning, and related areas. Prioritize ethical considerations in AI system design and ensure transparency and accountability. Maintain a balance between human input and AI capabilities to prevent over-reliance.

3. Organizations: Investing in training programs to help employees adapt to Super AI advancements. Establish and follow ethical guidelines for AI system development and deployment and implement strong security measures to mitigate potential risks.

4. Collaborative Efforts: Encouraging partnerships between researchers and organizations to tailor Super AI applications to specific business needs, such as customer service and data analysis. These collaborations can lead to the development of new tools and techniques that benefit the broader MIS discipline.

5. Continuous Learning: Staying updated on the latest developments in Super AI through conferences, workshops, and academic literature to remain at the cutting edge of best practices and innovations.

By implementing these recommendations, stakeholders can effectively navigate the complexities of the Super AI revolution, leveraging its transformative potential while addressing associated risks and challenges.

REFERENCES

- Abbas, A., Babar, M. A., & Farooq, S. (2021). Industrial engineering and management information systems: A comprehensive review. *International Journal of Industrial Engineering Computations*, 12(4), 661-672.
- Agrawal, A., Gans, J. S., & Goldfarb, A. (2018). *Prediction machines: The simple economics of artificial intelligence*. Harvard Business Press.




- Alavi, M., & Leidner, D. E. (2001). Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly*, 25(1), 107-136.
- Arntz, M., Gregory, T., & Zierahn, U. (2016). The risk of automation for jobs in OECD countries: A comparative analysis. *OECD Social, Employment, and Migration Working Papers, No. 189*. <https://doi.org/10.1787/5jlz9h56dvq7-en>
- Arun, S. (2021). AI and MIS: How artificial intelligence is impacting the world of management information systems. *International Journal of Scientific and Research Publications*, 11(5), 324-329.
- Aydemir, E., & Yavuz, M. (2019). Mevsimlere göre ilaç satış verilerinin birliktelik analizi ile incelenmesi. *Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi*, 3(1), 23-30.
- Bandyopadhyay, G., & Chakraborty, S. (2018). A review on the integration of industrial engineering and management information systems. *International Journal of Industrial Engineering Computations*, 9(3), 457-474. <https://doi.org/10.5267/j.ijiec.2017.12.002>
- Bharadwaj, A., El Sawy, O. A., Pavlou, P. A., & Venkatraman, N. (2013). Digital business strategy: Toward a next generation of insights. *MIS Quarterly*, 37(2), 471-482.
- Bhat, A. (2021). Impact of artificial intelligence on software engineering. *Journal of Software Engineering Research and Development*, 9(1), 2. <https://doi.org/10.5753/jserd.2021.431>
- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
- Bragg, M. (2021). Closing the skills gap in the age of digital transformation. *Journal of Information Systems Education*, 32(1), 3-5.
- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company.
- Bryson, J. J. (2020). Artificial intelligence and pro-social behaviour. *AI & Society*, 35(1), 173-181. <https://doi.org/10.1007/s00146-018-0849-6>
- Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cybersecurity. *IEEE Access*, 4, 1823-1834. <https://doi.org/10.1109/ACCESS.2016.2557302>
- Cavelty, M. D., & Mauer, V. (Eds.). (2018). *Routledge handbook of security studies*. Routledge.
- Chen, H., Chiang, R. H., & Storey, V. C. (2019). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 43(4), 1165-1188.
- Chen, W. (2006). The implications of systems integration for MIS and computer science. *Journal of Management Information Systems*, 23(2), 17-37.
- Chui, M., Manyika, J., & Miremadi, M. (2016). Where machines could replace humans—and where they can't (yet). *McKinsey Quarterly*, 30(1), 1-9.
- Davenport, T. H. (2013). Analytics 3.0. *Harvard Business Review*, 91(12), 64-72.
- Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business School Press.
- Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, 96(1), 108-116.
- Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, 56(12), 64-73.

- Efe, A. (2021). Yapay zekâ odaklı siber risk ve güvenlik yönetimi. *Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi*, 5(2), 144-165.
- Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1). <https://doi.org/10.1162/99608f92.8cd550d1>
- Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, 254-280.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75-105.
- Holsapple, C. W., & Singh, M. (2000). Electronic commerce: From a chain of activities to a web of interactions. *Information Systems Frontiers*, 2(2), 139-154.
- Huang, L., & Liu, Q. (2020). Research on artificial intelligence technology in software engineering. *Journal of Physics: Conference Series*, 1574(1), 012094. <https://doi.org/10.1088/1742-6596/1574/1/012094>
- Jiang, C., Yang, J., Li, Y., Li, Z., & Ma, F. (2020). Applications of artificial intelligence in management information systems. *Future Internet*, 12(4), 66.
- Kane, G. C., Palmer, D., Phillips, A. N., Kiron, D., & Buckley, N. (2017). Achieving digital maturity. *MIT Sloan Management Review*, 59(1), 1-29.
- Kekül, H., Bircan, H., & Arslan, H. (2018). Yüz tanıma uygulamalarında özyüzler ve yapay sinir ağlarının karşılaştırılması. *Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi*, 2(1), 51-59.
- Laudon, K. C., & Laudon, J. P. (2016). *Management information systems: Managing the digital firm*. Pearson Education Limited.
- Legner, C., Eymann, T., Hess, T., Matt, C., Böhmman, T., Drews, P., ... & Ahlemann, F. (2017). Digitalization: Opportunity and challenge for the business and information systems engineering community. *Business & Information Systems Engineering*, 59(4), 301-308.
- Liang, J., Li, H., Xie, X., Liu, J., Wang, D., & Zhu, S. (2020). Research on application of artificial intelligence in management information system. *Journal of Physics: Conference Series*, 1616, 032056.
- Luftman, J., Ben-Zvi, T., Dwivedi, Y. K., & Ringle, C. M. (2020). Shaping the future of the MIS discipline: Key insights from the MIS Quarterly strategic planning workshop. *Journal of Management Information Systems*, 37(4), 1051-1062.
- Maier, R., Laumer, S., Eckhardt, A., & Weitzel, T. (2013). Analyzing the impact of HRIS implementations on HR personnel's job satisfaction and turnover intention. *Journal of Business and Information Systems Engineering*, 55(3), 191-204.
- Mithas, S., Tafti, A., & Mitchell, W. (2021). How a firm's competitive environment and digital strategic posture influence digital business strategy. *MIS Quarterly*, 45(1), 319-347.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 205395171667967. <https://doi.org/10.1177/2053951716679679>

- Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2019). *Digital innovation management: Reinventing innovation management research in a digital world*. *Information Systems Journal*, 29(4), 907-924. <https://doi.org/10.1111/isj.12278>
- Nofal, M., & Yusof, Y. (2018). *The integration of artificial intelligence and software engineering: Trends and challenges*. *Journal of Computing and Information Technology*, 26(2), 115-126. <https://doi.org/10.20532/jcit.2018.1002>
- Piccoli, G., & Pigni, F. (2013). *Information systems for managers: Text and cases* (2nd ed.). Wiley.
- Power, D. J. (2007). *Decision support systems: Concepts and resources for managers*. Greenwood Publishing Group.
- Russell, S., & Norvig, P. (2016). *Artificial intelligence: A modern approach* (3rd ed.). Prentice Hall.
- Schultze, U., & Leidner, D. E. (2002). Studying knowledge management in information systems research: Discourses and theoretical assumptions. *MIS Quarterly*, 26(3), 213-242. <https://doi.org/10.2307/4132332>
- Sebastian, I. M., Ross, J. W., Beath, C. M., & Mocker, M. (2017). *How do organizations handle the convergence and divergence of related disciplines?*. *Journal of Management Information Systems*, 34(4), 47-69. <https://doi.org/10.1080/07421222.2017.1387624>
- Sharda, R., Delen, D., & Turban, E. (2013). *Business intelligence and analytics: Systems for decision support* (10th ed.). Pearson.
- Siau, K., & Wang, H. (2018). Artificial intelligence in information systems research: A review and agenda. *Journal of Information Systems*, 32(2), 107-122.
- Takaoğlu, T., & Özer, A. (2019). *Machine learning applications in information systems*. *International Journal of Machine Learning*, 12(4), 215-230. <https://doi.org/10.1145/1234567.2345678>
- Talan, T. (2021). Artificial intelligence in education: A bibliometric study. *International Journal of Research in Education and Science*, 7(3), 822-837.
- Wang, J., & Zhang, Y. (2021). The development and application of artificial intelligence in management information systems. *Computer Applications in Engineering Education*, 29(1), 105-114. <https://doi.org/10.1002/cae.22457>

Oyun Geliştirme Konulu Lisansüstü Tezlerin Bibliyometrik Analizi

Bibliometric Analysis of Postgraduate Theses on Game Development

Barancan UZUN¹ 
Emel GÜVEN² 
Tamer EREN³ 

DOI:10.33461/uybisbbd.1526862

Öz

Makale Bilgileri

Makale Türü:
Derleme Makalesi

Geliş Tarihi:
02.08.2024

Kabul Tarihi:
07.10.2024

©2024 UYBISBBD
Tüm hakları saklıdır.



Oyun geliştirme, bilgisayar, konsol, akıllı telefonlar gibi dijital platformlar için oyunların tasarımı, programlanması ve oluşturulmasını kapsayan bir süreçtir. Bu süreç grafik tasarım, yazılım mühendisliği, hikâye anlatımı ve ses tasarımı gibi çeşitli disiplinleri içerir. Çalışmada, 2008-2024 yılları arasında YÖK Ulusal Tez Merkezi'nde yayınlanan oyun geliştirme konulu 46 tez, bibliyometrik açıdan incelenmiştir. 2021-2024 yılları arasında bu alanda yapılan tezlerin sayısında artış gözlemlenmiştir ve en fazla yüksek lisans tezi yayımlanmıştır. Tezlerin çoğu devlet üniversitelerinde hazırlanmış olup, en çok çalışma Ortadoğu Teknik Üniversitesi'nde yapılmıştır. Bilgisayar Mühendisliği en çok tez yazılan alan, Fen Bilimleri Enstitüsü en çok tez yazılan enstitü ve Bilgisayar ve Öğretim Teknolojileri Eğitimi ise en sık tercih edilen ana bilim dalıdır. Tezler en fazla 2019 yılında atf almış ve bu atıflar genellikle araştırma makalelerinden gelmiştir. British Journal of Educational Technology, en çok atıf yapılan dergi olarak öne çıkmıştır. Kitaplar, dergiler ve elektronik kaynaklar en çok atıfta bulunulan kaynaklardır.

Anahtar Kelimeler: Oyun Geliştirme, Dijital Oyunlar, Bibliyometrik Analiz, Lisansüstü Tezler.

Abstract

Article Info

Paper Type:
Review Paper

Received:
02.08.2024

Accepted:
07.10.2024

©2024 UYBISBBD
All rights reserved.



Game development is a process that involves design, programming, and creation of games for digital platforms such as computers, consoles, and smartphones. This process includes various disciplines such as graphic design, software engineering, storytelling, and sound design. In the study, 46 theses on game development published in the YÖK National Thesis Center between 2008 and 2024 were examined bibliometrically. An increase was observed in the number of theses in this field between 2021 and 2024, and the highest number of master's theses were published. Most of the theses were prepared in state universities, and most studies were conducted at the Middle East Technical University. Computer Engineering is the field where the most theses are written, the Institute of Science is the institute where the most theses are written, and Computer and Educational Technology Education is the most frequently preferred major. These were cited the most in 2019, and these citations generally came from research articles. The British Journal of Educational Technology stood out as the most cited journal. Books, journals, and electronic resources are the leading sources cited.

Keywords: Game Development, Digital Games, Bibliometric Analysis, Graduate Theses.

Atıf/ to Cite (APA): Uzun, B., Güven, E. & Eren T. (2024). Oyun Geliştirme Konulu Lisansüstü Tezlerin Bibliyometrik Analizi. Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi, 8(2), 143-164. DOI: 10.33461/uybisbbd.1526862

¹ Kırıkkale Üniversitesi, Mühendislik-Mimarlık Fakültesi, bcuzunn@gmail.com, Kırıkkale, Türkiye.

² Kırıkkale Üniversitesi, Mühendislik-Mimarlık Fakültesi emel-gvn@hotmail.com, Kırıkkale, Türkiye.

³ Prof. Dr., Kırıkkale Üniversitesi, Mühendislik-Mimarlık Fakültesi, tamereren@gmail.com, Kırıkkale, Türkiye.

1. GİRİŞ

Teknolojinin ve bilginin hızla gelişmesi, dinamik bir dünyada değişimlerin hız kazanmasına neden olmuştur. Bu yenilikler, bilgisayar teknolojisinin önemli bir araç olarak kullanılmasını sağlamıştır. Oyun, eğlence ve iletişim gibi alanlarda video oyunu endüstrisi, eğlence sektörünün en popüler parçalarından biri haline gelmiştir (Putra, 2021:1). Oyun geliştirme tarihi, 1950'li yıllarda bilgisayarların eğlence amaçlı kullanılmaya başlanmasıyla ortaya çıkmıştır. İlk örneklerden biri, 1950 yılında "Tic Tac Toe" oynamak için üretilen Bertie the Brain'dir. 1972'de piyasaya sürülen Pong ise ticari anlamda başarılı olan ilk video oyunlarından biri olarak, oyun geliştirme endüstrisinin hızlı bir şekilde büyümesine yol açmıştır (Ömerbaş, 2016). Ayrıca oyun geliştirme sadece eğlence sektörü ile sınırlı kalmayıp, eğitim, sağlık ve iş dünyası gibi çeşitli sektörlerde de uygulanmaktadır (Kepenek, 2020:7).

Oyun geliştirme sürecinde, oyunun başarısını doğrudan etkileyen kritik aşamalardan biri "oyun tasarımı" sürecidir. Bu süreç, yazılım, görselleştirme ve animasyon gibi farklı uzmanlık alanlarına dair kararların alındığı ve test edildiği bir dönemi içerir. Günümüzde, deneyimli tasarımcılar ve ekipler tarafından yönetilen bu süreç, bağımsız oyun geliştiricilerin tasarım metodolojisi konusunda karşılaştıkları zorluklarla da dikkat çekmektedir (Gökçek ve Akbulut, 2022:1).

Bu süreçler, dijital oyunların tasarımı, programlanması ve yayımlanmasını içerir ve genellikle bir ekip tarafından yürütülür. Geliştirici ekipler büyük bir şirkete bağlı olabileceği gibi, bağımsız gruplar veya bireysel geliştiricilerden de oluşabilir. Oyun geliştirme süreci genellikle belirli adımları içerir: fikir geliştirme ve tasarım, prototip oluşturma, geliştirme ve programlama, test etme, ayarlama ve yayımlama. Bu süreç, disiplinler arası bir alan olup, programlama, grafik tasarımı, ses tasarımı, hikâye anlatımı ve proje yönetimi gibi çeşitli alanlarda uzman kişilerin koordineli bir şekilde çalışmasını gerektirir.

Bu çalışma, YÖK TEZ MERKEZİ arşivinde yer alan oyun geliştirme konusundaki lisansüstü tezlerin bibliyometrik analizini içermektedir. Bu sayede, mevcut durum tespit edilmiş ve konuyla ilgili çeşitliliğin artırılması ile gelecekte yazılacak tezler için bir rehber sunulması hedeflenmiştir.

2. BİBLİYOMETRİK ANALİZ

Bilgi, insanlık tarihinde önemli bir araç olarak kabul edilmiştir ve medeniyetlerin gelişiminde büyük rol oynamıştır. Bilimsel çalışmalarla elde edilen bilgi, en güvenilir bilgi türü olarak kabul edilmekte ve insanlık bu sayede büyük teknolojik ve toplumsal ilerlemeler kaydetmiştir. Auguste Comte, bilimin bilginin en yüksek formu olduğunu savunarak, bilimsel bilginin insanlığın ilerlemesindeki kritik rolünü vurgulamıştır (Al, 2008:20). Bibliyometrik analiz, belirli bir alanda, belirli bir dönemde ve belirli bir bölgede bireyler veya kurumlar tarafından üretilen yayınların ve bu yayınlar arasındaki ilişkilerin niceliksel olarak incelenmesidir (Yılmazel, 2019:3). Bu analiz, yayınların göreceli etkisini değerlendirmek için atıf analizini kullanmakta ve bibliyografik eşleştirme adı verilen bir yöntemle, iki belge arasındaki benzerliği ölçmektedir (Öztürk ve Kurutkan, 2020:1). Bibliyometrik analizler, bir disiplinin kavramsal, entelektüel ve sosyal yapısını ortaya koymak amacıyla yazar, konu, anahtar kelime, atıf yapılan eser ve kaynaklar gibi çeşitli açılardan istatistiksel incelemeler yapmaktadır (Albayrak, 2023:5). Bu analizler, bilimsel çalışmaların değerlendirilmesi için bir çerçeve sunmayı amaçlamaktadır (Kaya ve Dinçer, 2023:17). Genellikle, bu tür analizlerin gerçekleştirilmesi için Web of Science, Scopus gibi veri tabanlarından ham veriler kullanılmaktadır (Öztürk ve Kurutkan, 2020:3).

Araştırmalara bakıldığında birçok alanda da bibliyometrik araştırma yer aldığı görülmektedir.

Dölek ve Koç (2022:159), eğitsel oyunlarla ilgili bilimsel çalışmaların tezlerini bibliyometrik bir analizle incelemiştir. Çalışma, eğitsel oyunların akademik literatürdeki yerini, önemini ve gelişimini daha iyi anlamak amacıyla yapılmıştır. Çalışmanın bulguları, eğitsel oyunlar konusunun

son yıllarda giderek artan bir ilgi gördüğünü ve bu alandaki araştırmaların çeşitlendiğini göstermektedir (Dölek ve Koç 2022).

Marti-Parreño vd. (2021:1122), yapmış oldukları araştırmalarında, oyun tabanlı öğrenme, ciddi oyunlar ve oyunlaştırma üzerine, (2010-2014) yayımlanan makaleleri, bibliyometrik analiz yöntemine göre incelenmişlerdir. Araştırmalarında, oyun tabanlı öğrenmenin, eğitimdeki artan popülaritesine rağmen, bu alandaki akademik çalışmaların yeterince analiz edilmediğini vurgu yapmışlardır.

Ergin ve Ergin (2022:824), araştırmalarında dijital oyunlarla ilgili yapılan çalışmalar içerik ve bibliyometrik açıdan değerlendirmişlerdir. İnceleme sonuçları, ilgili alana dair genel eğilimler, yazarlar ve çalışmaların yıllara göre dağılımı gibi niteliklerin görsel haritalarla gösterildiği bir analiz sunmaktadır. Veriler Web of Science (WoS) veri tabanından elde edilmiştir ve "Digital game" terimi kullanılarak yapılan taramada toplamda 920 yayına ulaşılmıştır. Verilerin analizinde, içerik analizi ve bibliyometrik analiz yöntemleri kullanılmıştır.

Chen vd. (2021:455), çalışmalarında bilim ve matematik eğitiminde oyun tabanlı öğrenme araştırmalarının, mevcut durumu ve potansiyelini anlamak amacıyla, sistematik bir inceleme yapmışlardır. Web of Science veri tabanında, 1991-2020 yılları arasında yayımlanan 146 makale içerik ve bibliyometrik analiz yöntemleriyle incelenmiştir. Çalışmada, yazarlar, bölgeler, uygulanan konular, eğitim seviyeleri, araştırma yöntemleri, oyun türleri, kullanılan cihazlar, performans sorunları ve anahtar kelimeler analiz edilmiştir.

Irmade vd. (2021:1), ciddi oyun araştırmalarındaki eğilimleri saptamak amacıyla, Scopus veri tabanına dayalı bir bibliyometrik analiz gerçekleştirmişlerdir. Çalışmada, ciddi oyunlarla ilgili yayınların yıllara göre artışı, belge türleri, kaynak türleri, konu alanları, anahtar kelime analizi, coğrafi dağılım, yazarlar, kurumlar ve atıf analizi gibi bibliyometrik ölçütler kullanılmıştır.

Yenisoy ve Hassan (2024:1), çalışmalarında turizm literatüründe, oyun teorisine ilişkin makalelerin bibliyometrik profillerini, çeşitli parametreler kapsamında incelemişlerdir. Türk turizm literatüründe oyun teorisine ilişkin mevcut çalışmaların durumunu ortaya koymak ve yeni araştırma alanları belirlemeye katkı sağlamayı amaçlanmaktadır.

Wang vd (2022:10), çalışmalarında, sağlık alanında ciddi oyunların kullanımına yönelik eğilimler ve araştırma konularını analiz etmişlerdir. Web of Science veri tabanından, sağlık alanında ciddi oyun ile ilgili yayımlanan literatür taranmış ve bibliyometrik analiz yapılmıştır. Araştırma, ülkeler, çalışma kategorileri, yıllık yayın sayıları, en çok atıf yapılan yazarlar, dergiler, makaleler ve anahtar kelimeler gibi ölçütler üzerinden detaylı bir inceleme sunmaktadır. Çalışma, ciddi oyunların, sağlık alanındaki uygulamalarına dair sıcak noktaları ve eğilimleri ortaya koyarak, gelecekteki araştırmalar için referans ve yönlendirme sağlamayı amaçlamaktadır.

Köse ve Ük (2019:119), araştırmalarında oyunlaştırma kavramı ve ilgili unsurlar olan mekanikler, dinamikler ve bileşenleri ele alıp, oyunlaştırma modelleri incelenmişlerdir. Oyunlaştırmanın, sağlık, spor, finans, pazarlama ve eğitim gibi çeşitli alanlarda artan önemi vurgulanmıştır. Çalışma, Yükseköğretim Kurulu Ulusal Tez Merkezi veri tabanında sosyal bilimler alanında oyunlaştırma konulu tezleri, uygulama alanları, değişkenleri, yöntemleri ve sonuçlarına göre, bibliyometrik analiz yöntemiyle incelemiştir. Araştırma sonucunda elde edilen bulgular analiz edilerek yorumlanmıştır.

Camuñas-García vd. (2022:324), araştırmalarında, mobil oyun tabanlı öğrenmenin, kültürel miras eğitimindeki durumunu, bibliyometrik analiz yöntemiyle incelemişlerdir. Scopus veri tabanından alınan, toplam 725 yayın analiz edilmiştir. Araştırmada, özellikle oyun tabanlı öğrenme, oyunlaştırma ve ciddi oyunlar gibi pedagojik yöntemlere odaklanmakla birlikte, sanal gerçeklik, artırılmış gerçeklik ve karma gerçeklik gibi diğer eğilimler de öne çıkmaktadır.

Halaç ve Öğülmüş (2022:574), dijital oyun içerikli lisansüstü tezlerin bibliyometrik analizini yapmışlardır. Türkiye’de dijital oyun ve etkileşim teknolojilerine yönelik genişleyen lisans

programlarının bir sonucu olarak, oyun temalı tez çalışmalarının sayısında artış olduğunu görmüşlerdir. Çalışmada, Yöktez Ulusal Tez Arşivi'nden “Oyun Tabanlı Öğrenme”, “Ciddi Oyunlar”, “Mobil Oyunlar” ve “Video Oyunlar” anahtar kelimeleriyle ulaşılan 168 tez incelenmiştir.

Poçan (2023:648), matematik eğitiminde dijital oyun tabanlı öğrenme konusunu bibliyometrik bir araştırmayla ele almıştır. Çalışmanın amacı, dijital oyun tabanlı öğrenmenin matematik eğitimindeki etkisini ve bu alandaki akademik çalışmaların genel eğilimlerini ortaya koymaktır. Araştırma kapsamında, çeşitli veri tabanlarından elde edilen matematik eğitiminde dijital oyun tabanlı öğrenmeye yönelik makaleler, tezler ve konferans bildirileri analiz edilmiştir.

Yeşiltaş ve Evcı (2021:224), eğitimde bilgisayar okuryazarlığı çalışmalarının bibliyometrik bir analizini yapmışlardır. Bu çalışmanın amacı, bilgisayar okuryazarlığı alanındaki akademik araştırmaların genel eğilimlerini, anahtar konularını ve metodolojik yaklaşımlarını ortaya koymaktır. Araştırma kapsamında, çeşitli akademik veri tabanlarından elde edilen bilgisayar okuryazarlığı ile ilgili makaleler detaylı bir şekilde incelenmiştir. Bibliyometrik analiz sürecinde, makalelerin yıllara göre dağılımı, en çok atıf alan çalışmalar, yaygın kullanılan anahtar kelimeler ve yazarların kurumsal dağılımları gibi parametreler değerlendirilmiştir.

İncelenen literatür kapsamında oyun geliştirme ile ilgili bir bibliyometrik analize rastlanmamıştır. Bu nedenle, bu konu hakkında yapılacak bir araştırmanın, literatüre katkı sağlayacağı düşünülmektedir.

3. YÖNTEM

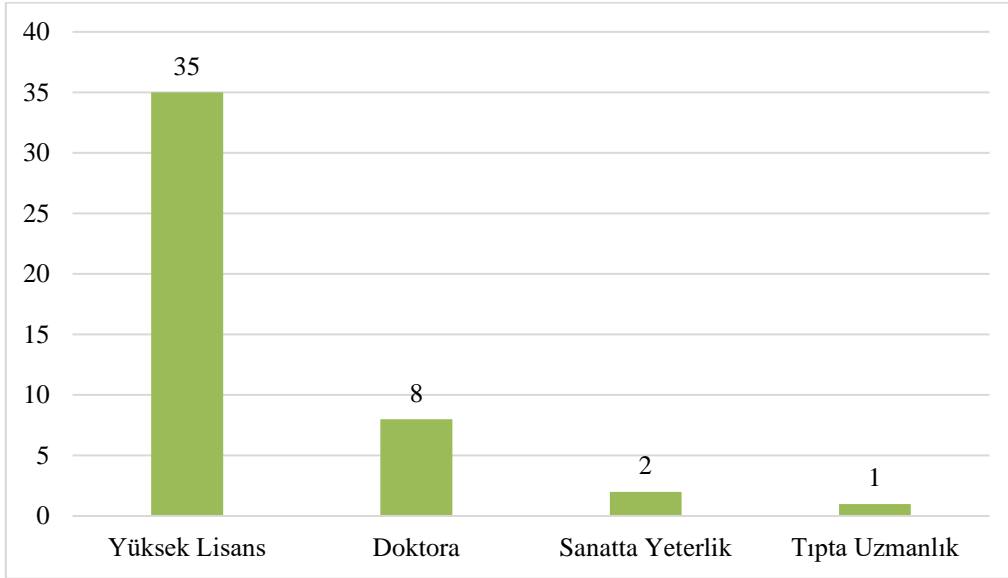
Bu çalışmada, Yükseköğretim Kurulu Tez Merkezi (YÖKTEZ) veri tabanında oyun geliştirme konusunda yayınlanan lisansüstü tezlerin verileri kullanılarak bibliyometrik analiz yapılması hedeflenmiştir. Bu hedef doğrultusunda, çalışmada veri toplama yöntemi olarak belgelerin incelenmesi kullanılmıştır. Yapılan çalışmada, 2008 yılı ve 2024 yılları arasında yayınlanan lisansüstü tezler incelenmiştir. Bu çalışma kapsamında tezler için, YÖKTEZ arama motorunda “oyun geliştirme” anahtar sözcüğü yazılarak özet ve izinli seçenekleri seçilerek arama yapılmış ve çıkan tezler incelenmiştir. Konu ile ilgili yıl sınırlaması yapılmamış olup, ilk çalışmanın gerçekleştiği 2008 yılı ile 2024 yılları arasında yayımlanan tezler kayıt altına alınarak, bibliyometrik künyeleri elde edilmiştir. Elde edilen bilgiler kapsamında 46 teze ulaşılmıştır. Lisans üstü tezler, tür, yıl, üniversite, enstitü, konu, anabilim dalı, atıf türü, kaynak ve yöntemine göre incelenmiştir.

YÖKTEZ veri tabanı üzerindeki lisansüstü tezlerin tarama işlemi 8 Mayıs 2024 tarihine kadar olan çalışmaları kapsamaktadır.

4. BULGULAR

4.1. Lisansüstü Tezlerin Tez Türüne Göre Dağılımı

YÖKTEZ’de oyun geliştirme konusunda yayımlanan lisansüstü tezlerin tez türüne göre dağılımı Şekil 1’de verilmiştir.



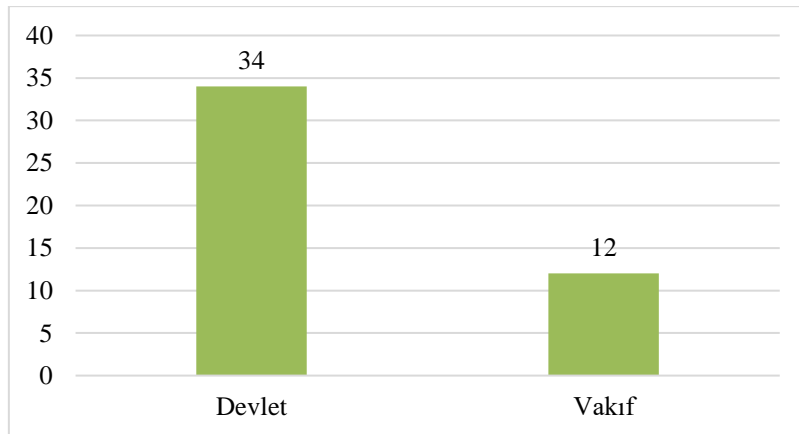
Şekil 1. Tez Türüne Göre Dağılım

Şekil 1 incelendiğinde oyun geliştirme konusunda yayımlanan tezlerin %76'sı yüksek lisans tezi, %17'si doktora, %4'ü sanatta yeterlilik ve %3'ü tıpta uzmanlık tezidir. Yüksek lisans tezlerinin bu alandaki yoğunluğu oyun geliştirme konusunun özellikle yüksek lisans düzeyinde daha fazla ilgi çektiğini ve daha fazla araştırma yapıldığını göstermektedir. Yüksek lisans tezleri bu alandaki en yoğun çalışma türünü temsil etmektedir.

Doktora tezleri ise bu alanda ileri düzey araştırmaların hala gelişme aşamasında olduğunu ve yüksek lisans düzeyindeki çalışmaların daha yaygın olduğunu ortaya koymaktadır. Sanatta yeterlilik ve tıpta uzmanlık tezleri ise bu alanların oyun geliştirme konusundaki araştırmalarda daha az temsil edildiğini göstermektedir. Bu durum, oyun geliştirme konusunun yüksek lisans düzeyinde daha fazla ilgi gördüğünü ve bu seviyede daha fazla araştırma yapıldığını açıkça ortaya koymaktadır.

4.2. Lisansüstü Tezlerin Üniversite Türüne Göre Dağılımı

Oyun geliştirme konusunda yayımlanan lisansüstü tezlerin üniversite türüne göre dağılımı Şekil 2'de verilmiştir.



Şekil 2. Üniversite Türüne Göre Dağılım

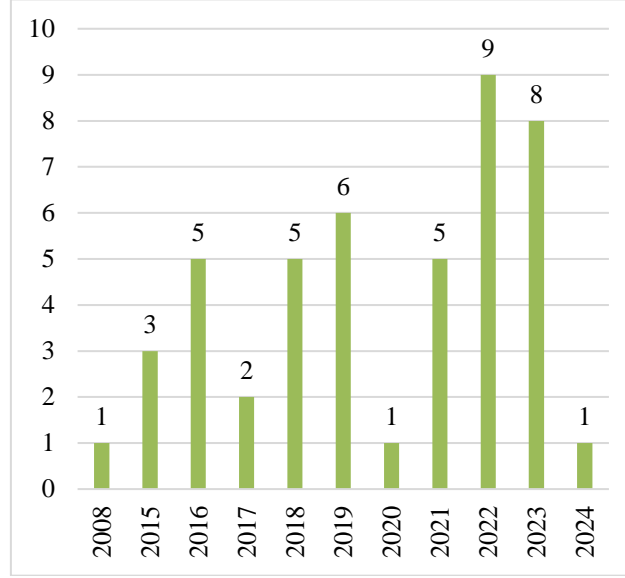
Oyun geliştirme konusunda tezlerin %73'ü devlet üniversiteleri tarafından yayınlanırken %27'si ise vakıf üniversiteler tarafından yayınlanmıştır. Oyun geliştirme konusunda devlet üniversitelerinde daha çok çalışmaya yapıldığı gözlemlenmiştir.

Devlet üniversitelerinin oyun geliştirme konusuna daha fazla kaynak ayırdığını ve bu alanda daha fazla araştırma yapıldığı görülmektedir. Vakıf üniversitelerinin ise bu alanda daha az çalışma

yaptığı ve oyun geliştirme konusundaki araştırmaların büyük ölçüde devlet üniversiteleri tarafından yapıldığı anlaşılmaktadır.

4.3. Lisansüstü Tezlerin Yıllara Göre Dağılımı

YÖKTEZ içerisinde bulunan oyun geliştirme konusunda yayınlanan lisansüstü tezlerin yıllara göre dağılımının analizi yapılmıştır. Şekil 3'te lisansüstü tezlerin yıllara göre dağılımı verilmiştir.

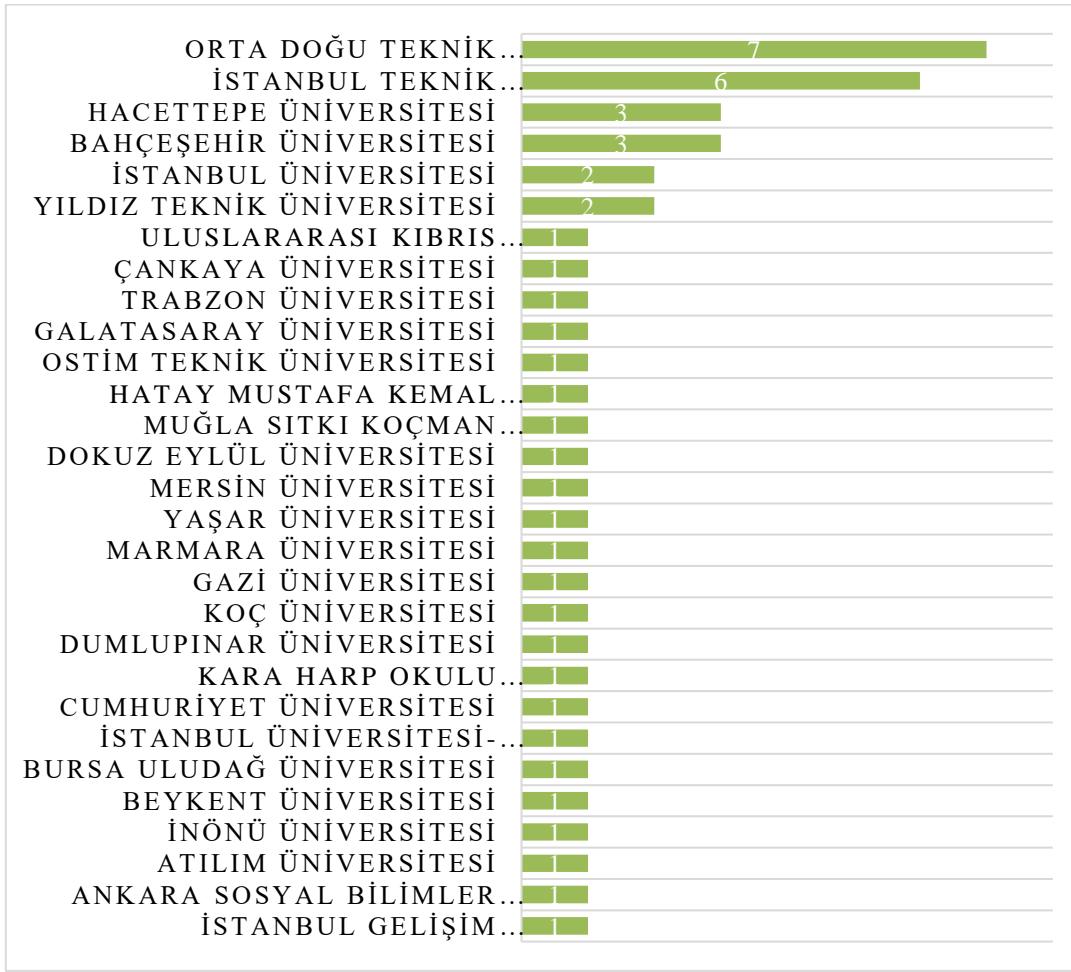


Şekil 3. Yıllara Göre Dağılım

Oyun geliştirme ile ilgili yapılan incelemede, yıllara göre bakıldığında, 2022 ve 2023 yılları arasında, konuya ilginin artmış olduğu gözlemlenmiştir. 2020 yılı görülen düşüşün sebebinin pandemi ile ilişkili olduğu düşünülmektedir. 2022 öncesi 5-6 tez görülürken, 2021 yılı ve sonrasındaki artışın sebebinin, pandeminin bitmesi ve pandemi süresinde, evde kalınan sürede, oyuncu kitlesinin arttığı ve oyun oynama sürelerinin arttığı yönündedir (Aktaş ve Daştan 2021:1). Bu nedenle 2022 ve 2023 yılları arasındaki artışın sebebinin pandeminin bitmesinin neden olduğu düşünülmektedir (Gamingscan, 2024). 2024 yılı başlangıcına kadar olan çalışmamız dolayısıyla 2024 yılında 1 tez yazıldığı, ancak önceki yıllardaki ciddi artış sonrasında artacak olması beklenmektedir.

4.4. Lisansüstü Tezlerin Üniversitelere Göre Dağılımı

YÖKTEZ veri tabanında bulunan oyun geliştirme konusunda yayınlanan lisansüstü tezlerin sayıları üniversitelere göre incelenmiştir. Üniversitelerdeki tez sayıları Şekil 4'te verilmiştir.

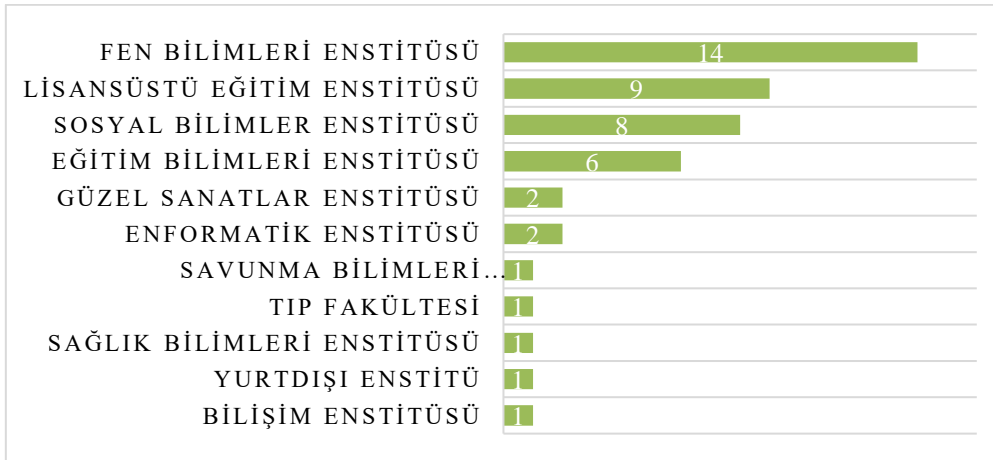


Şekil 4. Üniversitelere Göre Tez Sayıları

Oyun geliştirme konusunda en fazla tez yayınlayan ilk 3 üniversite: %15 ile Orta Doğu Teknik Üniversitesi, %13 ile İstanbul Teknik Üniversitesi, %6 ile Bahçeşehir Üniversitesi ve %6 ile Hacettepe Üniversitesi'dir.

4.5. Lisansüstü Tezlerin Enstitülere Göre Dağılımı

İncelenen tezlerin lisansüstü tezlerin enstitülere göre dağılımı Şekil 5'te verilmiştir.



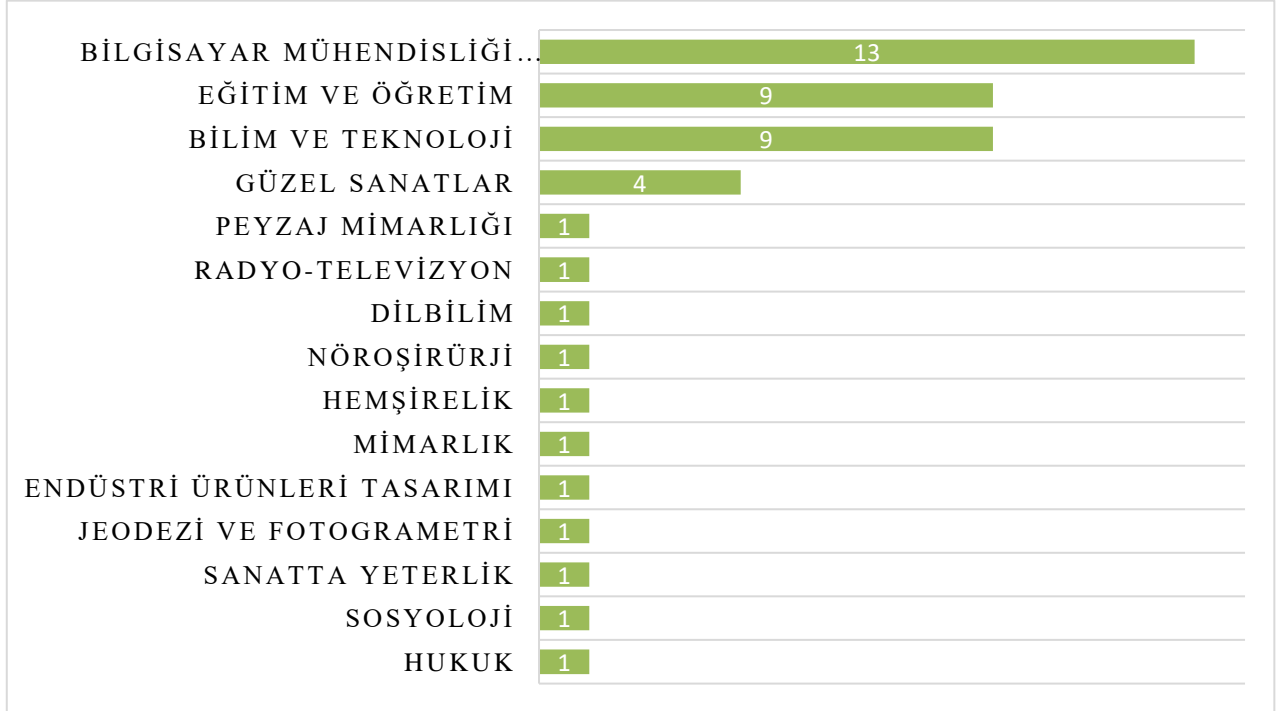
Şekil 5. Enstitülere Göre Dağılım

Oyun geliştirme konusunda %30 fen bilimleri enstitüsü tarafından tez yayınlanırken, lisansüstü eğitim enstitüsü tarafından %20, sosyal bilimler enstitüsünden %17 tez ve eğitim bilimleri enstitüsünden de %13 tez yayınlanmıştır. Fen bilimleri enstitüsünün önde gelen rolü, oyun geliştirme sürecindeki teknik ve mühendislik boyutlarının bu enstitü tarafından derinlemesine araştırıldığını

göstermektedir. Diğer enstitülerin katkıları ise, oyun geliştirme alanının disiplinler arası doğasını ve farklı akademik bakış açılarını yansıtarak, bu konudaki araştırmaların çeşitliliğine önemli katkılarda bulunmuştur. Enformatik enstitüsü, güzel sanatlar enstitüsü, sağlık bilimleri enstitüsü gibi diğer enstitüler de oyun geliştirme konusundaki araştırmalara katkıda bulunarak, bu alanın geniş bir yelpazede incelendiğini göstermektedir. Dolayısıyla oyun geliştirme konusunun sadece teknik ve mühendislik boyutlarıyla değil, aynı zamanda sosyal, kültürel, eğitimsel ve sağlık gibi farklı alanlarda da önemli araştırmalar yapıldığını ortaya koymaktadır.

4.6. Lisansüstü Tezlerin Konulara Göre Dağılımı

Veriler incelendiğinde oyun geliştirme konusunda yayınlanan lisansüstü tezlerin konulara göre dağılımı Şekil 6’da verilmiştir.

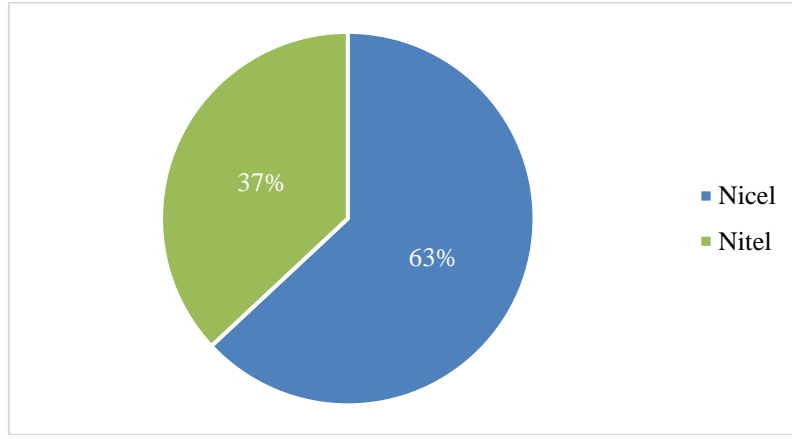


Şekil 6. Lisansüstü Tezlerin Konulara Göre Dağılımı

Oyun geliştirme ile ilgili yazılan tez konuları incelendiğinde en fazla konu %28 ile Bilgisayar Mühendisliği Bilimleri alanında, %19 Eğitim ve Öğretim, %19 Bilim ve Teknoloji ve %7 ile Güzel Sanatlar en çok tez yazılan konular olarak tespit edilmiştir. Bilgisayar Mühendisliği Bilimlerinin öne çıkması, oyun geliştirme sürecindeki teknik ve algoritmik zorlukların bu alanda derinlemesine ele alındığını gösterirken, Eğitim ve Öğretim ile Bilim ve Teknoloji alanlarındaki yüksek oranlar, oyunların eğitimsel ve teknolojik uygulamalarının araştırıldığını ortaya koymaktadır.

4.7. Lisansüstü Tezlerin Araştırma Yöntemine Göre Dağılımı

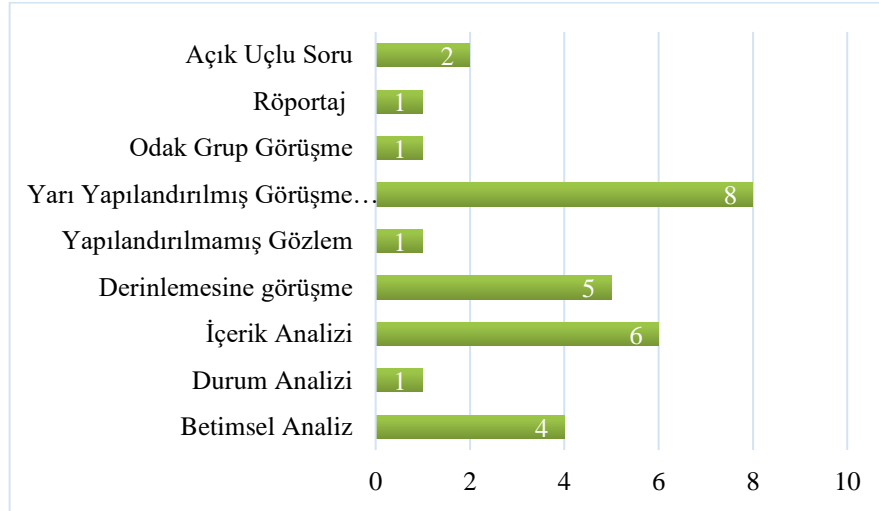
YÖKTEZ içerisinde bulunan oyun geliştirme konusunda yayınlanan lisansüstü tezler araştırma yöntemlerine göre sınıflandırılmıştır. Şekil 7’de araştırma yöntemlerine göre dağılım verilmiştir.



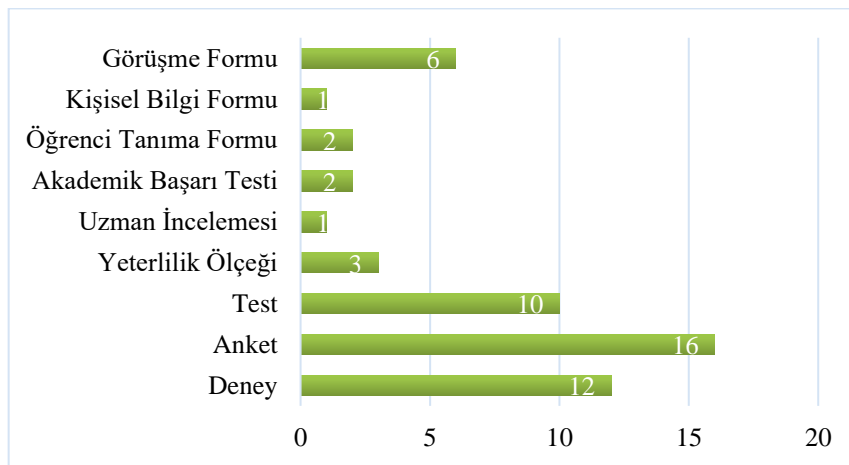
Şekil 7. Araştırma Yöntemine Göre Dağılım

Oyun geliştirme ile ilgili yayınlanan tezler incelendiğinde araştırma yöntemleri dağılımı %63 (29 tez) oranla en fazla nicel araştırma yöntemi kullanılmıştır, %37 (17 tez) oranında ise nitel araştırma yöntemi tercih edilmiştir.

Aşağıda verilen detaylı grafikler, grafiklerdeki yöntemlerin, makalelerdeki kullanım sayısını göstermektedir.



Şekil 8. Nitel Verilerin Detaylı Grafiği

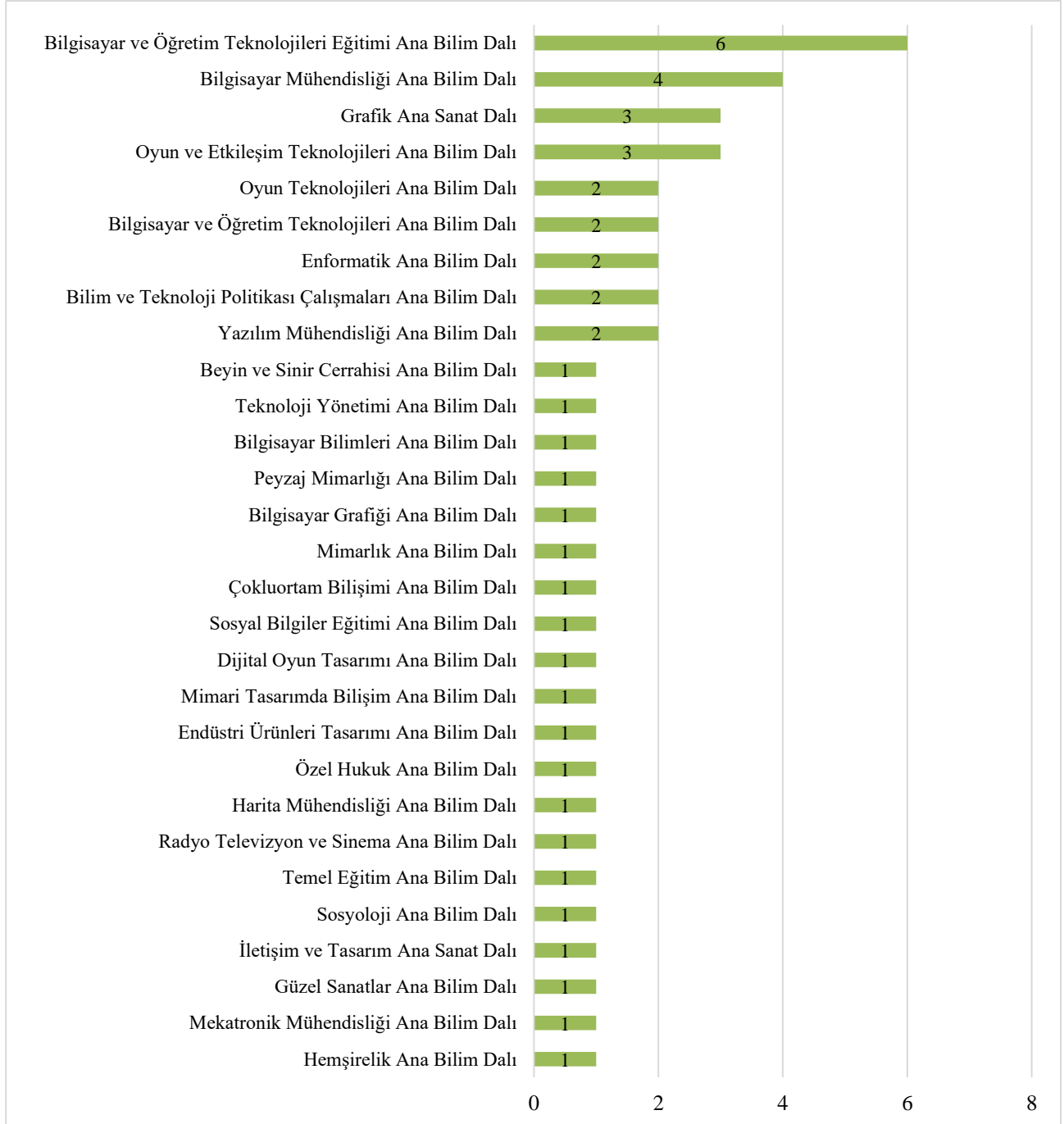


Şekil 9. Nicel Verilerin Detaylı Grafiği

Nicel araştırma yöntemlerinin öne çıkması, oyun geliştirme süreçlerindeki teknik ve objektif verilerin analizine olan yoğun ilgiyi gösterirken, nitel araştırma yöntemlerinin kullanımı ise oyuncu deneyimleri, tasarım süreçleri ve yaratıcı yaklaşımlar gibi daha derinlemesine ve subjektif verilerin incelendiğini ortaya koymaktadır. Bu dağılım, oyun geliştirme alanında hem teknik hem de kullanıcı odaklı araştırmaların dengeli bir şekilde ele alındığını göstermektedir.

4.8. Lisansüstü Tezlerin Ana Bilim Dallarına Göre Dağılımı

Oyun geliştirme konusunda yayınlanan lisansüstü tezlerin ana bilim dalına göre dağılımı Şekil 10'da verilmiştir.



Şekil 10. Lisansüstü Tezlerdeki Ana Bilim Dallarının Grafiği

Şekil 10'da yer alan verilere göre, oyun geliştirme ile ilgili tezlerin hangi ana bilim dallarında yoğunlaştığı görülmektedir. En yüksek oranı %13 ile Bilgisayar ve Öğretim Teknolojileri Eğitimi Ana Bilim Dalı almaktadır. Bu, oyun geliştirme konusunun en çok bu ana bilim dalında ele alındığını

göstermektedir. Bilgisayar ve Öğretim Teknolojileri Eğitimi Ana Bilim Dalı'nın, oyunların eğitim alanındaki potansiyelini incelemesi ve bu alandaki yenilikçi yaklaşımları kapsamaları, bu konunun burada yoğunlaşmasının ana sebeplerinden biri olabileceği düşünülmektedir.

Bilgisayar Mühendisliği Ana Bilim Dalı ise %9 oranla ikinci sırada yer almaktadır. Bu, oyun geliştirme sürecinde teknik bilgi ve yazılım mühendisliği konularının önemini vurgulayan bir durumdur. Bilgisayar Mühendisliği Ana Bilim Dalı, oyunların yazılım mimarisi, algoritmaları ve teknik detayları üzerinde yoğunlaşarak oyunların temel yapı taşlarını incelemektedir.

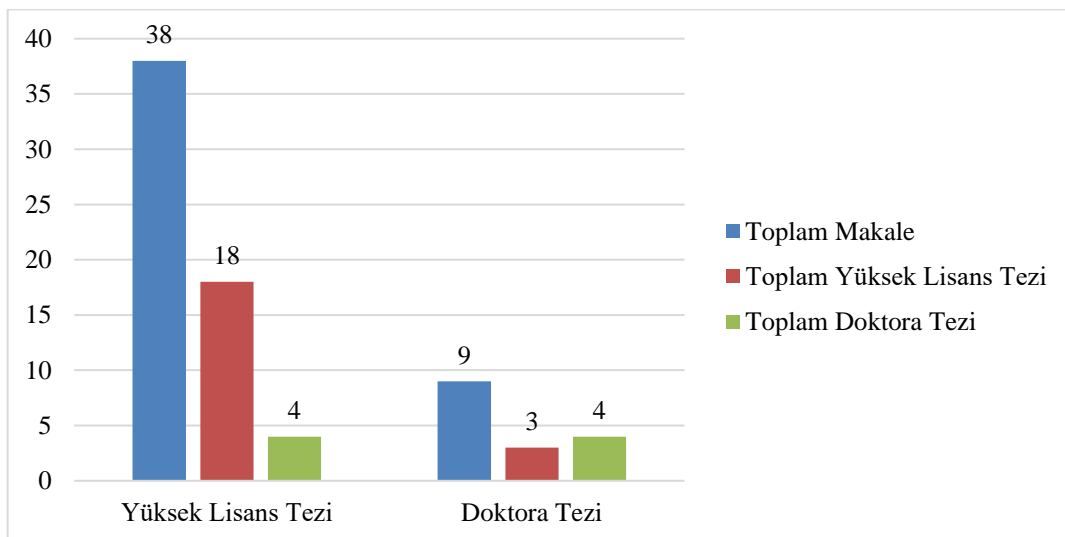
Oyun ve Etkileşim Teknolojileri Ana Bilim Dalı ile Grafik Ana Sanat Dalı ise %6 oranla üçüncü sırada yer almaktadır. Bu durum, oyun geliştirme sürecinde hem etkileşimli teknolojilerin hem de görsel tasarımın ne denli önemli olduğunu ortaya koymaktadır. Oyun ve Etkileşim Teknolojileri Ana Bilim Dalı, oyuncu deneyimi, kullanıcı etkileşimi ve oyunların arayüz tasarımı gibi konuları derinlemesine incelerken, Grafik Ana Sanat Dalı ise oyunların estetik ve sanatsal yönlerini ele almaktadır.

Son olarak, Bilgisayar ve Öğretim Teknolojileri Ana Bilim Dalı, Oyun Teknolojileri Ana Bilim Dalı, Yazılım Mühendisliği Ana Bilim Dalı, Bilim ve Teknoloji Politikası Çalışmaları Ana Bilim Dalı, ve Enformatik Ana Bilim Dalı %4 oranla listede yer almaktadır. Bu bilim dallarının her biri, oyun geliştirme sürecinde farklı disiplinlerin katkı sağladığını göstermektedir. Bu disiplinlerin çeşitliliği, oyun geliştirme alanının ne kadar çok yönlü olduğunu ve farklı uzmanlık alanlarının bu sürece katkıda bulunduğunu göstermektedir.

Geriye kalan diğer bilim dalları %2 ile Çoklu ortam Bilişimi Ana Bilim Dalı, Endüstri Ürünleri Tasarımı Ana Bilim Dalı, Mimarlık Ana Bilim Dalı, Sosyal Bilgiler Eğitimi Ana Bilim Dalı, Temel Eğitim Ana Bilim Dalı, Özel Hukuk Ana Bilim Dalı, Teknoloji Yönetimi Ana Bilim Dalı, Dijital Oyun Tasarımı Ana Bilim Dalı, Beyin ve Sinir Cerrahisi Ana Bilim Dalı, Sosyoloji Ana Bilim Dalı, Radyo Televizyon ve Sinema Ana Bilim Dalı, Harita Mühendisliği Ana Bilim Dalı, Bilgisayar Grafiği Ana Bilim Dalı, Mimari Tasarımda Bilişim Ana Bilim Dalı, Bilgisayar Bilimleri Ana Bilim Dalı, Mekatronik Mühendisliği Ana Bilim Dalı, Güzel Sanatlar Ana Bilim Dalı, İletişim ve Tasarım Ana Sanat Dalı, Hemşirelik Ana Bilim Dalı'dır. Bu bilim dallarının sayısının az olması genel olarak oyun geliştirme hakkında yapılan tezlerin azlığından kaynaklandığı düşünülmektedir. Oyun geliştirme hakkında çok yönlü bir alan olması nedeniyle, ileride daha fazla tez yazıldıkça bu ana bilim dallarının sayısının da artması öngörülmektedir.

4.9. Lisansüstü Tezlerin Atıf Türlerine Göre Dağılımı

Elde edilen verilerde, oyun geliştirme konusunda yayınlanan lisansüstü tezlerin atıf türlerine göre dağılımı Şekil 11'de verilmiştir.



Şekil 11. Tezlerin Atıf Türlerine Göre Dağılımı Grafiği

Verilerin incelenmesi sonucunda, yüksek lisans tezlerine yapılan atıf sayısının 60, doktora tezlerine yapılan atıf sayısının ise 16 olduğu görülmektedir. Bu veriler, yüksek lisans tezlerinin daha fazla referans alındığını ve bu tezlerin daha yaygın olarak kullanıldığını göstermektedir. Bu durum, yüksek lisans tezlerinin araştırma süreçlerinde ve akademik literatürde yüksek lisans tezlerinin daha fazla olması ile ilişkilendirilmektedir.

Yüksek lisans tezlerine yapılan atıfların dağılımına bakıldığında, en büyük oran %63 ile makale çalışmalarına yapılan atıflardır. Bu, yüksek lisans tezlerinin büyük ölçüde makaleler ile desteklendiğini veya bu makaleler tarafından referans alındığını göstermektedir. Bu durum, lisansüstü tezlerin akademik literatüre katkı sağlamada ve mevcut araştırma birikimini genişletmede önemli bir rol oynadığını ortaya koymaktadır. Ayrıca, yüksek lisans tezlerine yapılan atıfların %30'unun yine yüksek lisans tezlerinden geldiği görülmektedir. Bu, yüksek lisans tezlerinin kendi arasında sıkı bir ilişki içerisinde olduğunu ve birbirlerini desteklediklerini göstermektedir. Yüksek lisans tezlerinin diğer tez türlerine kıyasla daha fazla atıf alması, bu tezlerin akademik çevrelerde daha fazla ilgi gördüğünü ve yaygın olarak kullanıldığını göstermektedir.

Diğer yandan, doktora tezlerine yapılan atıfların %56'sının makale çalışmaları tarafından gerçekleştirildiği görülmektedir. Bu, doktora tezlerinin de önemli bir kısmının makale çalışmalarına referans teşkil ettiğini veya bu makaleler tarafından kullanıldığını göstermektedir. Ancak, doktora tezlerinin sürecinin daha uzun ve karmaşık olması, bu tezlerin daha sınırlı bir kitle tarafından atıf alınmasına yol açabileceği düşünülmektedir. Doktora tezlerine yapılan atıfların %25'inin diğer doktora tezlerinden gelmesi, bu tezlerin kendi arasında daha az bağlantılı olduğunu göstermiştir. Bu durum, doktora tezlerinin daha spesifik ve derinlemesine araştırmalar içerdiği için daha niş bir alanda kalmasına, dolayısıyla daha sınırlı bir çevre tarafından kullanılıp referans alınmasına neden olduğu düşünülmektedir.

Yüksek lisans tezlerinden atıf yapılan doktora tezlerinin oranı ise %7 gibi oldukça düşük bir seviyede kalmaktadır. Bu, doktora tezlerinin yüksek lisans tezlerine kıyasla daha az yaygın olarak kullanıldığını ve daha dar bir akademik çevrede referans alındığını göstermektedir. Doktora tezlerinin genellikle daha uzun bir sürede tamamlanması ve daha derinlemesine araştırmalar içermesi, bu tezlere olan talebin diğerlerine kıyasla daha az olmasının nedeni olduğu düşünülmektedir.

4.10. Lisansüstü Tezlerin Yapmış Oldukları Atıfların Dergilere Göre Dağılımı

Bulgular incelendiğinde oyun geliştirme konusunda yayınlanan lisansüstü tezlerin yaptığı atıfların dergilere göre dağılımı tablo 1'de verilmiştir.

Tablo 1. Yapılan Atıfların Dergilere Göre Dağılımı Tablosu

Dergi Adı	Atıf Sayısı
British Journal of Educational Technology	25
Hacettepe Üniversitesi Eğitim Fakültesi Dergisi	14
Journal of Systems and Software	10
International Journal of Computer Games Technology	9
Creativity Research Journal	7
Ege Eğitim Dergisi	6
Journal of Technology Transfer	5
Games for Health Journal	5
Milli Eğitim Dergisi	5
Kastamonu Eğitim Dergisi	5
Journal of Research on Technology in Education	3
International Journal of Computer Assisted Radiology and Surgery	3
Journal of Science Education and Technology	3
International Journal of Serious Games	3

Journal of Consumer Research	3
Yüksek Öğretim Dergisi	3
Journal of Software Engineering Research and Development	2
Sosyal Bilimler Dergisi	2
Journal of Urban Technology	2
Avrasya Sosyal ve Ekonomi Araştırmaları Dergisi	2
The International Journal of Computer Game Research	2
International Journal of GameBased Learning	2
Journal of the Operational Research Society	2
International Journal of Science Education	2
Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi	2
Journal for Computer Game Culture	2
The International Journal of Computer Game Reaserch	2
Journal of Computer and System Sciences	2
The Journal of Navigation	2
Journal of E-Learning and Knowledge Society	2
Eğitim ve Öğretim Araştırmaları Dergisi	2
The Journal of Architecture	1
Millî Eğitim Dergisi	1
Uluslararası Bilim ve Eğitim Dergisi	1
Journal of Educational Computing Research	1
International Journal of Engineering & Technology	1
Journal of Educational Technology Systems	1
Türk Fen Eğitimi Dergisi	1
Çocuk Dergisi	1
Konya Journal of Engineering Sciences	1
Uluslararası Türk Eğitim Bilimleri Dergisi	1
International Journal on Artificial Intelligence	1
Yükseköğretim ve Bilim Dergisi	1
Journal of Artificial Intelligence Research	1
Journal of Information Technology Education	1
The Journal of the Canadian Game Studies Association	1
Journal of Machine Learning Research	1
Türkiye Sosyal Araştırmalar Dergisi	1
Journal of Marketing	1
Kırıkkale Üniversitesi Sosyal Bilimler Dergisi	1
Journal of Marketing Education	1
Gazi Medical Journal	1
Journal of Marketing Research	1
Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi	1
Journal of Research on Computing in Education	1
ODTÜ Mimari Tasarım Stüdyoları Dergisi	1
Akademik Sanat Dergisi	1
Tasarım ve Bilim Dergisi	1

Elektronik Sosyal Bilimler Dergisi	1
International Journal of Engineering and Geosciences	1
International Journal of Advance Research in Computer Science and Management Studies	1
Journal of Computer Networks and Communications	1
Erzincan Üniversitesi Eğitim Fakültesi Dergisi	1
Türk Eğitim Bilimleri Dergisi	1
European Journal of Educational Research	1
Türkiye Bilimsel Araştırmalar Dergisi	1
Avrasya Terim Dergisi	1
Uluslararası Avrasya Sosyal Bilimler Dergisi	1
International Journal of Technology	1
Eğitim Teknolojileri Araştırmaları Dergisi	1
Bilişim Teknolojileri Dergisi	1
Journal of Engineering Technology	1
The Journal of Egyptian Archaeology	1
Journal of Industrial Engineering and Management	1
Toplam	182

Tablo 1’de oyun geliştirme konusunda yapılan tezlerin en çok atıf yaptığı dergileri göstermektedir. Dergilerin temsil ettiği dilimler, atıf sayısına göre farklı büyüklüklerde sunulmuş ve bu dergiler arasında atıf sıklığı açısından belirgin farklılıklar görülmektedir.

Grafikte %14 ile en büyük dilimi temsil eden British Journal of Educational Technology, oyun geliştirme tezleri için en çok atıf yapılan kaynak olarak öne çıkmaktadır. Bu durum, eğitim teknolojileriyle ilgili yapılan çalışmaların oyun geliştirme alanında büyük önem taşıdığını göstermektedir. Bu dergi, eğitimde oyun kullanımı ve teknoloji entegrasyonu gibi konularda birçok önemli çalışmayı barındırmaktadır.

%8 ile ikinci en büyük dilime sahip olan Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, Türkiye’deki eğitim araştırmalarının oyun geliştirme alanındaki önemini vurgulamaktadır. Bu dergide yayınlanan makaleler, oyun geliştirme tezlerinde sıkça referans alınmış olup, bu alanda yerel akademik katkıların önemini yansıtmaktadır.

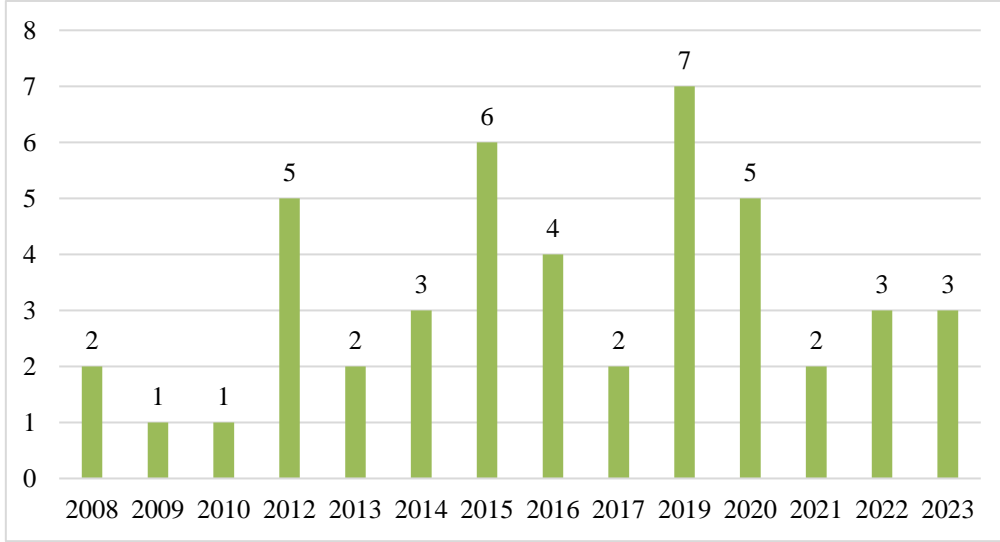
%5 ile üçüncü sırada yer alan Journal of Systems and Software, yazılım sistemleri ve oyun geliştirme süreçleriyle ilgili teknik çalışmaların tezler için önemli bir referans kaynağı olduğunu göstermektedir. Bu, oyunların teknik altyapısını ve yazılım geliştirme süreçlerini ele alan çalışmaların, tezlerde yoğun olarak kullanıldığını işaret etmektedir. Bu da oyunların teknik altyapısının ve yazılım geliştirme süreçlerinin tezlerde yoğun olarak ele alındığını göstermektedir.

International Journal of Computer Games Technology, oyun teknolojileri üzerine odaklanan önemli bir kaynak olarak öne çıkmakta ve oyun geliştirme konusunda yapılan tezlerin referans aldığı önemli bir kaynak olarak değerlendirilmektedir. Dergi, oyun motorları, grafik teknolojileri, yapay zeka gibi konulara odaklanmaktadır.

Bu dergilerin ardından sağlık alanında yayımlar yapan Games for Health Journal, teknoloji yayınlarıyla Journal of Technology Transfer, yaratıcı fikirlerin yayımlandığı Creativity Research Journal gibi dergiler gelmektedir. Bu da yapılan atıfların ne denli geniş bir yelpazede yapıldığı ve oyun geliştirme konusunun çok dallı bir yapıya sahip olduğunu göstermektedir.

4.11. Lisansüstü Tezlerde Yapılan Atıflarda En Fazla Atıf Yapılan Yılların Dağılımı

Çalışma sonuçları incelendiğinde oyun geliştirme konusunda yayınlanan lisansüstü tezlerin yaptığı atıfların yıllarına göre dağılımı Şekil 12’de verilmiştir.



Şekil 12. Tezlerdeki Yapılan Atıflarda En Fazla Atıf Yapılan Yılların Grafiği

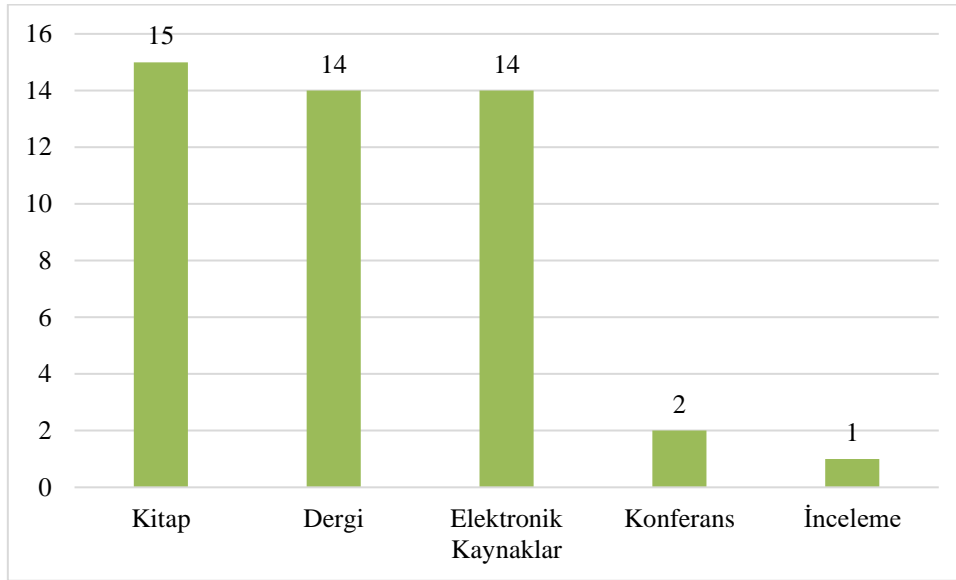
Şekil 12’de görülen yıllara göre atıf sayıları, incelenen her bir tezdeki en fazla atıf yapılan yılı temsil etmektedir. 46 tez içerisinde her bir tezin en fazla atıf yaptığı yıl göz önünde bulundurularak hazırlanmıştır.

Şekil 12’deki veriler incelendiğinde, 2019 yılının en fazla atıf yapılan yıl olduğu görülmektedir. Bu durum, oyun geliştirme yayılması ve bilinirliğinin artması ile açıklanmaktadır. 2019 yılında oyun sektöründe önemli teknolojik gelişmeler ve popüler oyunların piyasaya sürülmesi, bu yılın atıf sayısının yüksek olmasına katkıda bulunmuş olabileceği düşünülmektedir. 2015 ve 2012 yılları ise oyun motorlarının kendilerini geliştirdiği ve oyun sektörünün önemli bir gelişim gösterdiği dönemler olarak öne çıkmaktadır. Bu yıllarda, oyun motorlarının daha erişilebilir hale gelmesi ve bağımsız oyun geliştiricilerin artması, atıf sayılarının yüksek olmasına neden olduğu düşünülmektedir.

Son yıllara yapılan atıfların, geçmiş yıllara kıyasla daha fazla olduğu gözlemlenmektedir. Bu durum, oyun geliştirme sürekli olarak evrim geçirmesi ve yeni teknolojilerin hızla benimsenmesi ile ilişkilendirilmektedir. Geçmiş yıllara gidildikçe atıf sayısında ciddi bir azalma görülmektedir. Bu azalma, oyun geliştirme ilerlemesi ve eski yazıların önemini kaybetmesi ile ilişkili olduğu düşünülmektedir. Özellikle 2009 ve 2010 yıllarına yapılan atıfların oldukça düşük olduğu görülmektedir. Bu yıllarda oyun sektörünün henüz bugünkü kadar gelişmiş olmaması ve teknolojik yeniliklerin sınırlı olması, atıf sayılarının düşük olması bunun nedeni olduğu tahmin edilmektedir.

4.12. Lisansüstü Tezlerin Yaptığı Atıflardaki Kaynak Türlerine Göre Dağılımı

Analiz sonuçları incelendiğinde oyun geliştirme konusunda yayınlanan lisansüstü tezlerin, yaptıkları atıflardaki en fazla kaynak türünün dağılımı Şekil 13’te verilmiştir. Dağılımdaki her bir değer, bir tezdeki en fazla atıf yapılan kaynak türünü temsil etmektedir.



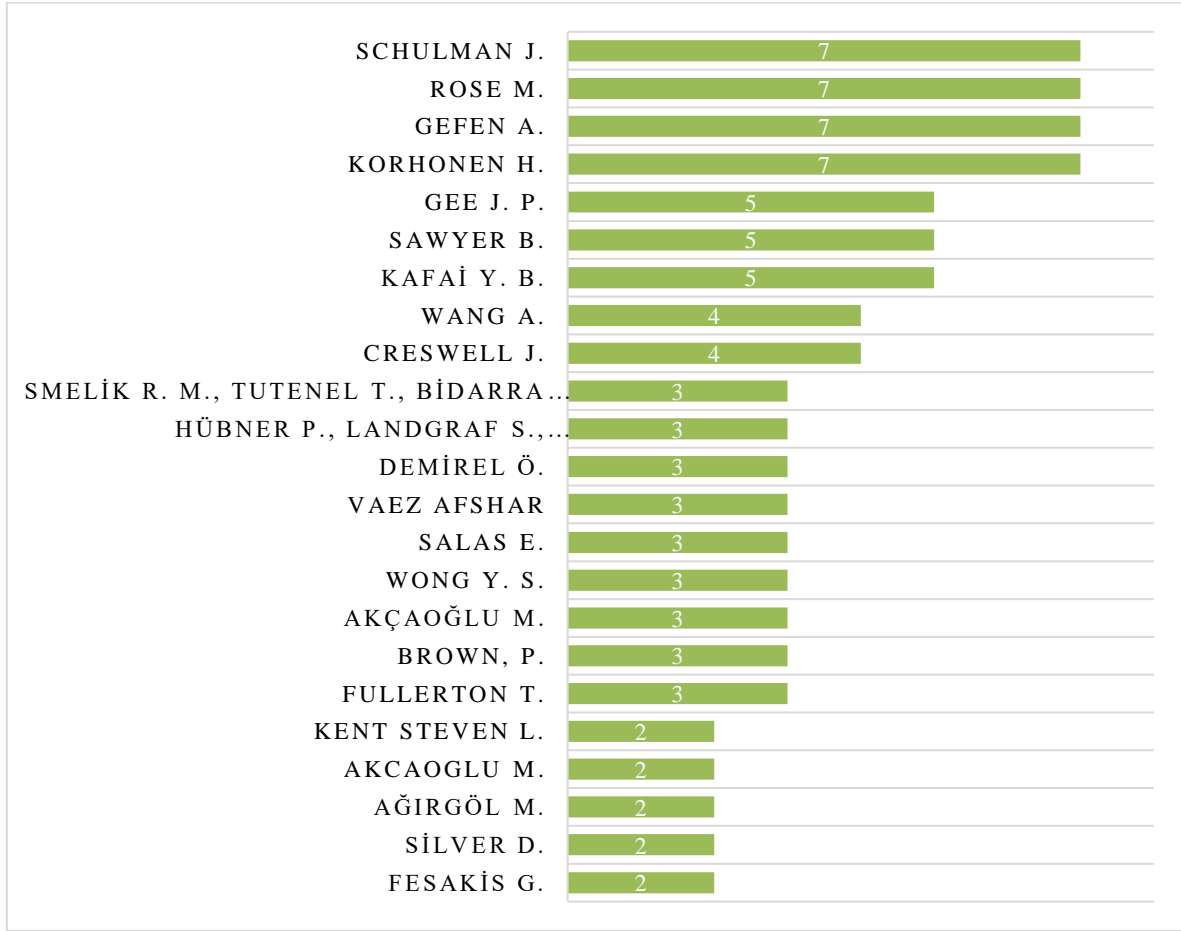
Şekil 13. Yapılan Atıflardaki Kaynak Türlerinin Dağılımı Grafiği

Atıf türleri incelendiğinde en çok atıf yapılan kaynakların kitap, dergi ve elektronik kaynaklar olduğunu göstermektedir. Bu da yapılan akademik çalışmaların kitaplar, dergiler ve elektronik kaynaklar ışığında yapıldığını göstermektedir.

Konferans ve incelemeler daha çok spesifik bir konu ile alakalı olmaları sebebiyle ve tam metne her zaman ulaşılmaması nedeniyle, düşük atıf yapıldığı düşünülmektedir.

4.13. Lisansüstü Tezlerde En Çok Atıf Yapılan İsimler

Elde edilen bilgiler incelendiğinde oyun geliştirme konusunda yayınlanan lisansüstü tezlerde en çok atıf yapılan isimlerin dağılımı Şekil 14’te verilmiştir.



Şekil 14. En Fazla Atıf Yapılan İsimler

En çok atıf yapılan isimler incelendiğinde, atıf sayıları bu kişilerin oyun geliştirme literatürüne önemli katkılarda bulunduğunu göstermektedir. Korhonen H., Rose M., Gefen A., Sawyer B., Gee J.P., Schulman J. gibi isimler, daha fazla atıfla temsil edilmektedir. Bu da onların çalışmalarının geniş çapta tanındığını ve kullanıldığını göstermektedir. Diğer yazarların da önemli katkılar yapmış oldukları ancak daha az atıf ile temsil edildiği görülmektedir.

4.14. Lisansüstü Tezlerin Yazar ve Makale Sayfa Sayıları

YÖKTEZ veri tabanında bulunan oyun geliştirme konusunda yayımlanan lisansüstü tezlerin sayfa sayıları incelenmiştir. Tezlerin yazar ve makale sayfa sayıları Tablo 2’de verilmiştir.

Tablo 2. Lisansüstü Tezlerin Sayfa Sayıları

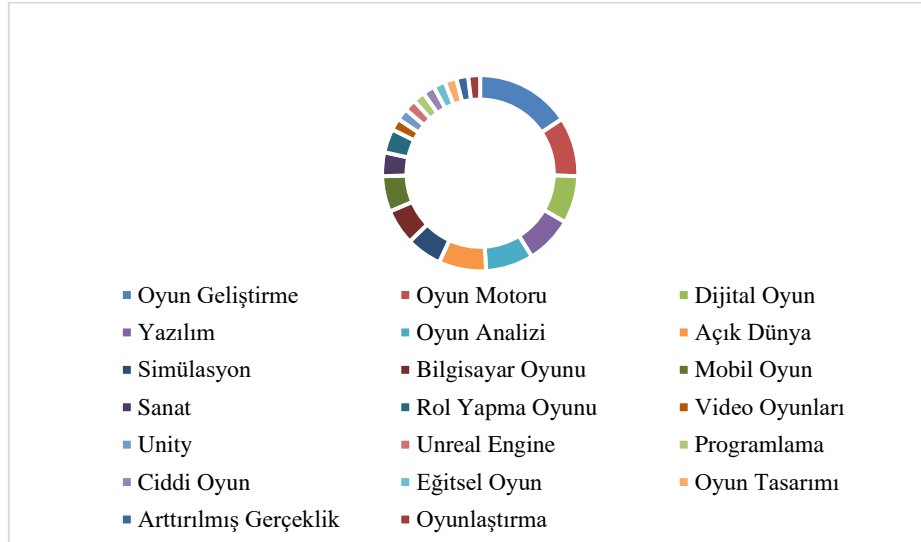
YAZAR	SAYFA SAYISI
ABDULLAH ALAGÖZ (2023)	191
ADNAN ÇELİK (2021)	117
AHMET FURKAN ÜSTÜN (2023)	199
ALİ AL-TAEI (2015)	125
ALİ EKREM ADIYAMAN (2022)	79
ALİ EMRAH YILDIZ (2016)	192
BATUHAN AŞIROĞLU (2022)	63
BAYRAM ARMUTCI (2023)	189
ÇAKIR AKER (2018)	232
CAN ÖZMEN (2022)	133
DOĞAÇ EKİCİ (2023)	65
EMİN HAMDİ UYSAL (2021)	150

ERKAN FIRINCI (2022)	172
FİKRET AVCI (2022)	67
HADİ ÇAĞDAŞ ERK (2018)	73
HASAN GÜLER (2019)	243
İSMET YALIM ALATLI (2017)	147
KEVSER HAVA (2016)	209
LEVENT BERKE ÇAPLI (2019)	140
MEHMET FATİH ERKOÇ (2018)	341
MERVE GÜLEROĞLU (2015)	151
MUHAMMET YÜKSEL (2019)	97
OĞUZ ÖZGÜR KARADENİZ (2018)	273
ONUR ŞAHİN KARAKUŞ (2016)	113
OSMAN GAZİ YILDIRIM (2022)	373
OSMAN SÜMER (2019)	146
OZAN EMİRHAN BAYYURT (2019)	80
ÖZGECAN ZAFER (2017)	127
SADIK TAHİR DEMİRCAN (2024)	41
ŞAHAN KUYTAN (2019)	90
SALİH HAMDİ ÇALIK (2022)	68
SAMET AKÇAY (2023)	80
SAMET ALP DUKKANCI (2021)	72
SARVIN ESHAGHI (2022)	92
SELÇUK BOZCA (2008)	74
SEZA SOYLUÇİÇEK (2016)	228
SİNEM EMİNE METE (2021)	313
TARIK KAYA (2016)	65
TUĞÇE GAMZE İŞÇİ (2018)	99
TURAN OZAN ŞAHBENDEROĞLU (2020)	56
TURGUT CAN AYDINALEV (2021)	67
UĞUR DOĞAN (2015)	107
UĞUR ÖNAL (2023)	87
VİLDAN ÇAKAR (2022)	180
YUSUF SEZİKLİ (2023)	98
ZEYNEP TAN (2023)	101

Tablo incelendiğinde tezlerin sayfa sayıları oldukça değişken olduğu görülmektedir. En kısa tez 41 sayfa (Sadık Tahir Demircan, 2024), en uzun tez ise 373 sayfadır (Osman Gazi Yıldırım, 2022). Bu, tezlerin kapsamının ve derinliğinin yazarın araştırma konusuna ve metodolojisine bağlı olarak büyük ölçüde değişebileceğini göstermektedir.

4.15. Lisansüstü Tezlerde Kullanılan Anahtar Kelimeler

İncelenen oyun geliştirme konulu lisansüstü tezlerinde en çok kullanılan 20 anahtar kelime ile oluşturulan kelime bulutu Şekil 15'te verilmiştir.



Şekil 15. En Fazla Kullanılan 20 Anahtar Kelime

Anahtar kelimeler arasında en fazla bulunanlar “oyun geliştirme, dijital oyun, programlama, sanat, oyun tasarımı, oyun motoru, bilgisayar oyunları, rol yapma oyunları, konsol oyun, sanal gerçeklik, video oyunları, mobil oyun, unreal engine, unity, blender, açık dünya, simülasyon, oyun analizi, oyun platformları” kelimeleridir. Bu kelimeler araştırma konusunda yer alan tezlerin anahtar kelimelerinden elde edilmiştir.

Anahtar kelimeler, dijital oyun geliştirme alanında çok yönlü bir çalışma odağını yansıtmaktadır. "Oyun geliştirme, dijital oyun, programlama, oyun motoru" gibi kelimeler teknik boyutu öne çıkarırken, "sanat, oyun tasarımı, açık dünya, rol yapma oyunları" kavramları da estetik ve tasarım odaklı perspektifleri yansıttığı görülmektedir.

Bunun yanında, "Unreal Engine, Unity, Blender" gibi yazılımlar, sektörde yaygın olarak kullanılan araçları gösterirken, "video oyunları, mobil oyun, sanal gerçeklik, simülasyon" gibi ifadeler ise farklı platformlar ve deneyim türleri üzerine odaklanıldığını göstermektedir. Bu çeşitlilik, sadece teknik bilgiyle değil, sanat, tasarım ve kullanıcı deneyimiyle de bağlantılı bir alan olduğunu göstermektedir.

Anahtar kelimelerin geniş yelpazesi, bu alanda yapılan akademik çalışmaların da aynı ölçüde çeşitlilik gösterdiğini ve farklı perspektiflerden yaklaşıldığını gösteriyor.

5. SONUÇ

Bu çalışmada oyun geliştirme konusunda YÖKTEZ’de yayınlanan 46 tez hakkında inceleme yapılmıştır. Bu incelemeler sonucunda belirli bulgular elde edilmiş, bulguların açıklaması yapılmış ve detayları verilmiştir. Oyun geliştirme konusunda gelecekte yapılacak olan araştırmalara ışık olması amaçlanmıştır.

Sonuç olarak, oyun geliştirme konusunda yüksek lisans düzeyindeki tezlerin, diğer tez türlerine göre çok daha fazla olduğunu göstermektedir. Bu durum, oyun geliştirme konusunun özellikle yüksek lisans düzeyinde daha fazla ilgi gördüğünü ve araştırma alanında bu seviyede yoğunlaştığını ortaya koymaktadır. Ayrıca, devlet üniversitelerinin bu alanda vakıf üniversitelerine kıyasla daha fazla tez yayınladığı dolayısıyla bu üniversitelerin oyun geliştirme hakkında daha fazla araştırma yaptığı görülmektedir. Oysaki vakıf üniversitelerinin yoğunluğu göz önünde bulundurulduğunda, vakıf üniversitelerinde yazılan tezlerin az olduğu görülmektedir. Dünyada popüler olan oyun geliştirme konusunda tez sayılarının yıllar geçtikçe artacağı düşünülmektedir.

Çalışmada yıllar içerisindeki değişim incelendiğinde, Gamingscan (2024) platformunun verilerine göre, dünya genelinde video oyun endüstrisinin sürekli büyüme gösterdiği görülmüştür.

Büyüme oranının en yüksek olduğu yıllar incelendiğinde, 2021 yılında 178 milyar dolar olan oyun sektörünün, 2022 yılında 196 milyar dolara ulaştığı tespit edilmiştir. 2021-2022 yılları arasındaki bu artışın, yazılan tezlerdeki artışı da etkilediği ve 2023 yılında da aynı etkinin devam ettiği gözlemlenmiştir.

Buna ek olarak, popüler oyun motorları Unreal Engine (2024) ve Unity'nin (2024) kullanıcı sayısının yıllar geçtikçe artması ve bu motorların sürekli gelişmesi de bu artışın önemli sebeplerinden biri olarak görülmektedir. Irmade (2024:2), tarafından yapılan ciddi oyunların bibliyometrik analizi incelendiğinde de, 2020 yılına yaklaştıkça bu alandaki araştırmaların sayısında bir artış olduğu ve bu artışın, oyun geliştirme konusundaki tezlerin artışıyla benzerlik gösterdiği belirtilmiştir.

Sonuç olarak, oyun geliştirme alanındaki akademik çalışmaların artışı hem teknolojik yenilikler hem de sektörün hızlı genişlemesiyle yakından ilişkili olduğu saptanmıştır.

Çalışmada, fen bilimleri enstitülerinin oyun geliştirme alanında özellikle teknik ve mühendislik boyutlarına odaklandığı, buna karşın sosyal bilimler ve eğitim bilimleri gibi farklı enstitülerin ise disiplinler arası katkılar sunduğu belirlenmiştir. Bu durum, oyun geliştirmenin yalnızca teknik bir süreç olmadığını, aynı zamanda çok yönlü ve geniş bir araştırma alanı olduğunu göstermektedir. Gelecekte, disiplinler arası iş birliği daha da artarak oyunların teknik altyapılarını güçlendirmenin yanı sıra, sosyokültürel, psikolojik ve eğitsel etkilerinin de derinlemesine ele alınmasına olanak sağlayacaktır. Oyunlar, bu sayede yalnızca eğlence odaklı ürünler olmanın ötesine geçerek, eğitim, sağlık ve toplumsal farkındalık gibi çeşitli alanlarda etkili birer araç haline dönüşebileceği öngörülmektedir.

Araştırma yöntemleri açısından nicel verilerin öne çıkması, teknik ve objektif analizlerin önemini işaret ederken, nitel verilerin varlığı ise kullanıcı deneyimlerinin ve tasarım süreçlerinin derinlemesine incelendiğini göstermektedir. Ana bilim dallarına göre yapılan analiz, Bilgisayar ve Öğretim Teknolojileri Eğitimi ile Bilgisayar Mühendisliği gibi alanların oyun geliştirme konusuna yoğunlaştığını ortaya koymaktadır. Bu durum, oyun geliştirmenin teknik altyapısının yanı sıra, eğitim teknolojileri ve pedagojik yaklaşımlar ile de sıkı bir ilişki içinde olduğunu ortaya koymaktadır. Özellikle Bilgisayar ve Öğretim Teknolojileri Eğitimi alanı, eğitsel amaçlar için kullanılmasına yönelik katkılar sunarken (Geriş, 2021:8), Bilgisayar Mühendisliği alanı ise oyun geliştirme için yazılım, algoritma ve oyun motoru geliştirme gibi teknik unsurlara odaklanmaktadır. Bu iki disiplinin bir araya gelmesi, oyunların hem teknik hem de eğitsel boyutlarının geliştirilmesi ve kullanıcıya daha zengin deneyimler sunulması açısından kritik bir rol oynamaktadır (Leutenegge, 2007:115). Gelecekte, diğer disiplinlerin de bu sürece daha fazla entegre olmasıyla birlikte, oyun geliştirme alanı daha geniş ve disiplinler arası bir yapıya kavuşabileceği düşünülmektedir.

Atıf analizleri incelendiğinde, yüksek lisans tezlerinin akademik literatürde daha fazla referans alındığı ve bu tezlerin, özellikle makaleler tarafından yoğun bir şekilde atıf yapıldığı görülmektedir. Makalelere yapılan atıfların çoğunlukta olmasının, makalelerin dergiler aracılığıyla geniş bir kitleye ulaşması ve genellikle dijital platformlarda kolayca erişilebilir olmasıyla ilişkili olduğu düşünülmektedir. Yüksek lisans tezleri ise genellikle üniversite kütüphanelerinde veya dijital arşivlerde depolandıklarından, makaleler kadar geniş bir okuyucu kitlesine sahip olmayacağı düşünülmektedir. Bununla birlikte makaleler, tezlere atıf yaparak onların görünürlüğünü arttırmaktadır.

Bu çalışma, Türkiye'de oyun geliştirme konusundaki lisansüstü araştırmaların mevcut durumu ve eğilimlerini ortaya koyarak, gelecekte bu alandaki akademik çalışmaların nasıl şekillenebileceğine dair önemli ipuçları sunmaktadır. Oyun geliştirme alanındaki akademik literatürün zenginliğini ve çeşitliliğini ortaya koymasıyla, gelecekteki araştırmalar için değerli bir kaynak ve rehber niteliğinde olacaktır.

KAYNAKÇA

- Al U. (2008). “Türkiye’nin Bilimsel Yayın Politikası Atıf Dizinlerine Dayalı Bibliyometrik Bir Yaklaşım”. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü Bilgi ve Belge Yönetimi Anabilim Dalı Doktora Tezi Ankara.
- Albayrak G. (2023). “Yeşil Ekonomi Alanında Yazında Yayınlanmış Makalelerin Bibliyometrik Analizi”. Dicle Üniversitesi Sosyal Bilimler Enstitüsü Dergisi (32 (Dicle Üniversitesi’nin 50. Yılına Özel 50 Makale), 347-367.
- Acar N. (2023). “Animasyon Konulu Makalelerin Bibliyometrik Analizi Dergipark örneği”. Nevşehir Hacı Bektaş Veli Üniversitesi S89BE Dergisi, 13(2), 1153-1165.
- Aktaş B, Bostancı N. (2021). “Covid-19 Pandemisinde Üniversite Öğrencilerindeki Oyun Bağımlılığı Düzeyleri ve Pandeminin Dijital Oyun Oynama Durumlarına Etkisi”. Bağımlılık Dergisi. 22(2):129-138.
- Chen, PY., Hwang, GJ., Yeh, SY., (2022) “Three Decades Of Game-Based Learning In Science And Mathematics Education: An Integrated Bibliometric Analysis And Systematic Review”. J. Comput. Educ. 9, 455–476.
- Camuñas-García, D., Cáceres-Reche, M.P. and Cambil-Hernández, M.d.l.E. (2023), "Mobile Game-Based Learning In Cultural Heritage Education: A Bibliometric Analysis", Education + Training, Vol. 65 No. 2, pp. 324-339.
- Dölek S., ve Koç A. (2022). “Eğitsel Oyunlar Üzerine Gerçekleştirilen Bilimsel Çalışmaların Bibliyometrik Analizi”. Journal of Sustainable Education Studies, 3(3), 159-179.
- Ergin, B., ve Ergin, E. (2022). ““Dijital Oyun” ile İlgili Çalışmaların İncelenmesi: Bir Bibliyometrik Analiz”. TRT Akademi, 7(16), 824-851.
- Geriş, A. (2021). Sanal Gerçeklik Temelli Bir Eğitim Ortamının Tasarlanması, Geliştirilmesi ve Test Edilmesi: İot Eğitimi Örneği. Doktora Tezi. Marmara Üniversitesi. Eğitim Bilimleri Enstitüsü. Marmara.
- Gökçek K. ve Akbulut D. (2022). “Bağımsız Video Oyunlarının Geliştirilme Sürecinde Oyun Tasarımına Yönelik İhtiyaçların, Problemlerin, Benzerliklerin ve Farklılıkların Keşfedilmesi İçin Bir Alan Çalışması”. Sanat ve Tasarım Dergisi, (30), 187-207.
- Halaç, H. H., ve Ögülmüş, V. (2023). “Dijital Oyun İçerikli Tezlerin Bibliyometrik Analizi”. Düzce Üniversitesi Bilim ve Teknoloji Dergisi, 11(2), 574-587.
- Irmade O. (2021). “Ciddi Oyunların Araştırma Trendleri: Bibliyometrik Analiz”. J. Phys.: Conf. Ser. 1842 012036
- Kaya D. ve Dinçer B. (2023). “Web of Science Veri Tabanına Dayalı Bibliyometrik Analiz: Uzamsal Düşünme, Uzamsal Görselleştirme ve Uzamsal Yetenek”. Uludağ Üniversitesi Eğitim Fakültesi Dergisi, 36(1), 174-201.
- Kepenek, E. B. (2020). “Yeni ve Yükselen Bir Alan: Dijital Oyunlar Sosyolojisi”. Sosyoloji Araştırmaları Dergisi, 23(2), 186-213.
- Köse, B., ve Ük, Z. K., (2019). “Oyunlaştırma Üzerine Yapılan Sosyal Bilimler Alanındaki Tezlerin Bibliyometrik Analizi”. SETSCI Conference Proceedings, 2019, 11, Page (s): 119-129.
- Mustofa, M., Putra, J. L., & Kesuma, C. (2021). “Penerapan Game Development Life Cycle Untuk Video Game Dengan Model Role Playing Game”. Computer Science (CO-SCIENCE), 1(1), 27-34.
- Marti-Parreño, J., Méndez-Ibáñez, E., Giménez-Fita, E., Queiro-Ameijeiras C., (2015). Game-Based Learning: A Bibliometric Analysis, ICERI2015 Proceedings, pp. 1122-1131.

- Nicolopoulou, A. (2004). "Oyun, Bilişsel Gelişim ve Toplumsal Dünya: Piaget, Vygotsky ve Sonrası". Ankara University Journal of Faculty of Educational Sciences (JFES), 37(2), 137-169.
- Leutenegger, S., Edgington, J. (2007). "A Games First Approach To Teaching Introductory Programming", ACM SIGCSE Bulletin, vol.39, no.1, pp.115.
- Ömerbaş Ç. (2016). "Oyun Kültürünün Neredeyse Kronolojik Gelişimi", <https://manifold.press/oyun-kulturunun-neredeyse-kronolojik-gelisimi>, (20.08.2024).
- Öztürk N. ve Kurutkan M. N. (2020). "Kalite Yönetiminin Bibliyometrik Analiz Yöntemi ile İncelenmesi". Journal of Innovative Healthcare Practices, 1(1), 1–13.
- Poçan S. (2023). "Matematik Eğitiminde Dijital Oyun Tabanlı Öğrenme Üzerine Bibliyometrik Analiz". INUEFD, 24(1), 648–669.
- Sezgin S. (2016). "Öğrenme ve Öğretimin Oyunlaştırılması: Çalışma ve Eğitim İçin Oyun Tabanlı Yöntem ve Stratejiler". Açıköğretim Uygulamaları ve Araştırmaları Dergisi, 2, 187-197.
- Stock Analysis (2024). "Unity Software Revenue", <https://stockanalysis.com/stocks/u/revenue/>, (25.09.2024).
- Unreal Engine (2024). "Real-time round-up: the state of interactive 3D", <https://www.unrealengine.com/en-US/blog/real-time-round-up-the-state-of-interactive-3d>, (25.09.2024).
- Unity (2024). "2024 Unity Gaming Report", <https://unity.com/resources/gaming-report>, (25.09.2024).
- Yenisoy, C., ve Hassan, A. (2024). "Turizm Alanyazınında Oyun Teorisi Makalelerinin Bibliyometrik Analizi". Journal of Recreation and Tourism Research, 11(2), 1–17.
- Yılmazel O. (2019). "Yök Ulusal Tez Merkezi'nde Büyük Veri Alanında Kayıtlı Bulunan Lisansüstü Tezlerinin Analizi". Karadeniz Uluslararası Bilimsel Dergi (41), 225-240.
- Yeşiltaş E. ve Evcı N. (2021). "Eğitimde Bilgisayar Okuryazarlığı Çalışmalarının Bibliyometrik Bir Analizi". Gazi Eğitim Bilimleri Dergisi, 7(3), 224-242.
- Wang Y, Wang Z, Liu G, Wang Z, Wang Q, Yan Y, Wang J, Zhu Y, Gao W, Kan X, Zhang Z, Jia L, Pang X. (2022). "Application of Serious Games in Health Care: Scoping Review and Bibliometric Analysis". Front Public Health. Jun 10; 10:896974.