



# JOURNAL OF INFORMATION TECHNOLOGIES

## BİLİŞİM TEKNOLOJİLERİ DERGİSİ

Volume / Cilt **16**

Number / Sayı **2**

Year / Yıl **2023**

Month / Ay **April / Nisan**





GAZİ ÜNİVERSİTESİ (GAZİ UNIVERSITY)  
BİLİŞİM ENSTİTÜSÜ (INSTITUTE OF INFORMATICS)



# BİLİŞİM TEKNOLOJİLERİ DERGİSİ

(JOURNAL OF INFORMATION TECHNOLOGIES)

ISSN: 1307-9697 e-ISSN: 2147-0715

Cilt (Volume): 16

Sayı (Issue): 2

Nisan (April) 2023

Sahibi (Owner)  
**Dr. Musa YILDIZ**  
Rektör (Rector)

**Genel Yayın Yönetmeni & Baş Editör**  
(General Publication Director & Editor in Chief)

**Dr. Aslıhan TÜFEKÇİ**

Bilişim Enstitüsü Müdürü  
Director of Institute of Informatics

**Yardımcı Editör**

(Associate Editor)

**Dr. Öner BARUT**

Bilişim Enstitüsü Müdür Yardımcısı  
Vice Director of Institute of Informatics

**Yardımcı Editör**

(Associate Editor)

**Dr. Murat YILMAZ**

Bilişim Enstitüsü Müdür Yardımcısı  
Vice Director of Institute of Informatics

## Editörler (Editors)

❖ <b>Dr. Ahmet KARAARSLAN</b>	Ankara Yıldırım Beyazıt Üniversitesi (Ankara Yıldırım Beyazıt University)
❖ <b>Dr. Hüseyin POLAT</b>	Gazi Üniversitesi (Gazi University)
❖ <b>Dr. İbrahim KÖK</b>	Pamukkale Üniversitesi (Pamukkale University)
❖ <b>Dr. Mahir DURSUN</b>	Gazi Üniversitesi (Gazi University)
❖ <b>Dr. Mehmet ŞİMŞEK</b>	Milli Savunma Üniversitesi (National Defence University)
❖ <b>Dr. M. Hanefi CALP</b>	Ankara Hacı Bayram Veli Üniversitesi (Ankara Hacı Bayram Veli University)
❖ <b>Dr. Mohammad Abdus SALAM</b>	Southern University and A&M Collage
❖ <b>Dr. Oktay YILDIZ</b>	Gazi Üniversitesi (Gazi University)
❖ <b>Dr. Olgun DEĞİRMENÇİ</b>	TOBB ETÜ (TOBB Economics and Technology University)
❖ <b>Dr. Recep BENZER</b>	Gazi Üniversitesi (Gazi University)
❖ <b>Dr. Serdar KULA</b>	Gazi Üniversitesi (Gazi University)

## Yayın Danışma Kurulu (Editorial Advisory Board)

<b>Dr. Ahmet COŞAR</b> Turkish Aeronautical Association University, Turkey	<b>Dr. Aslanbek NAZİEV</b> Ryazan State University, Russia	<b>Dr. Bogdan PATRUT</b> Alexandru Ioan Cuza University of Iasi, Romania
<b>Dr. Deepak GUPTA</b> Maharaja Agrasen Institute of Technology, India	<b>Dr. Jafar A. ALZUBİ</b> Al-Balqa Applied University, Jordan	<b>Dr. Jolanta SABAITYTĖ</b> Vilnius Gediminas Technical University, Lithuania
<b>Dr. İlya LEVİN</b> Tel Aviv University, Israel	<b>Dr. Pınar KARAGÖZ</b> Middle East Technical University, Turkey	<b>Dr. Ufuk ÇAĞLAYAN</b> Yaşar University, Turkey

<b>Dr. Veysi İŞLER</b> <i>Hasan Kalyoncu University, Turkey</i>	<b>Dr. Victor Hugo Costa DE ALBUQUERQUE</b> <i>Universidade de Fortaleza, Brazil</i>	<b>Dr. Vijender Kumar SOLANKİ</b> <i>CMR Institute of Technology, India</i>
<b>Dr. Ebrahim KHOSRAVI</b> <i>Clayton State University, United States</i>		

<b>Dil Editörü</b> <i>(Language Editor)</i> <b>Dr. Çağla Gizem AKKAŞ</b> <i>Bilişim Enstitüsü</i> <i>Institute of Informatics</i>
---

<b>Teknik Koordinatör</b> <i>(Technical Coordinator)</i> <b>Dr. Muhammed Ali KOŞAN</b> <i>Bilişim Enstitüsü</i> <i>Institute of Informatics</i>
---

<b>Teknik Sorumlu</b> <i>(Technical Assistant)</i> <b>Dr. Merve ÇELEBİ</b> <i>Bilişim Enstitüsü</i> <i>Institute of Informatics</i>
---

<b>Teknik Sorumlu</b> <i>(Technical Assistant)</i> <b>Tuana İRKEY</b> <i>Bilişim Enstitüsü</i> <i>Institute of Informatics</i>
--

<b>Teknik Sorumlu</b> <i>(Technical Assistant)</i> <b>Yasemin İÇTÜZER</b> <i>Bilişim Enstitüsü</i> <i>Institute of Informatics</i>
--

<b>Sekreterlik</b> <i>(Secretary)</i> <b>Bilişim Teknolojileri Dergisi</b> <i>Bilişim Enstitüsü</i> <i>Institute of Informatics</i>
---

<b>Bilişim Teknolojileri Dergisi uluslararası hakemli bir dergidir.</b> Journal of Information Technologies is an international refereed journal.
<b>Yazışma Adresi (Contact Address)</b> Tunus Cad. No: 35 Kavaklıdere Çankaya/ANKARA Telefon / Telephone: 0312 202 38 01 Faks / Fax: 0312 212 79 29
<b>Çevrimiçi Değerlendirme Sistemi (Online Evaluation System)</b> <a href="http://dergipark.gov.tr/gazibtd">http://dergipark.gov.tr/gazibtd</a> <b>E-posta (e-mail):</b> btd@gazi.edu.tr
<b>Bilişim Teknolojileri Dergisi 3 ayda bir (Ocak, Nisan, Temmuz, Ekim) yayınlanmaktadır.</b> Journal of Information Technologies is published every 3 months (January, April, July, October).

# An Expert System for Bolt Selection

*Araştırma Makalesi/Research Article*

 Kadir SARI\*,  Yunus KAYIR,  Hakan DİLİPAK

Manufacturing Engineering, Gazi University, Ankara, Türkiye

[kadirsarion@gmail.com](mailto:kadirsarion@gmail.com), [ykayir@gazi.edu.tr](mailto:ykayir@gazi.edu.tr), [hdilipak@gazi.edu.tr](mailto:hdilipak@gazi.edu.tr)

(Geliş/Received:26.10.2022; Kabul/Accepted:31.01.2023)

DOI: 10.17671/gazibtd.1195078

**Abstract**— Expert systems are one of the widely used artificial intelligence techniques and their use is increasing day by day. Expert systems are a technique that can use the knowledge and experience of experts, evaluate them at the decision-making stage and make inferences. Bolts are fasteners used to connect various parts and are within the scope of standard machine elements. There are different types of bolts in the industry and choosing the right bolt requires expertise. In this study, an expert system called Exbolt System, which selects bolts according to head types, was developed. Commonly used standard bolt types were based on bolt head types. Relevant rules for bolt types were established by collecting and analysing information about each bolt. While the rules were written, it was aimed to choose the truest bolt type. Criteria such as the need for fine-tuning, use in rotating parts, centring in the hole, flat surface requirement, mounting accuracy, system weight status, high force-holding and use in dirty environments were considered in the creation of the rules. The programme makes the best choice and recommends the bolt that can be used according to the head type to the user. CLIPS expert system programming language was used in the development of the Exbolt System. With the answers to the questions asked to the user by the programme, it was ensured that the truest bolt selection was made. The most accurate result was achieved by making faster, easier and more comprehensive decisions in the bolt selection, which requires expertise, and a more effective and efficient selection process was realised by saving time and labour.

**Keywords**— bolt head type, selection, design, expert system, artificial intelligence, clips

## Cıvata Seçimi için Bir Uzman Sistem

**Özet**— Uzman sistemler yaygın olarak kullanılan bir yapay zekâ tekniğidir ve kullanımı günden güne artmaktadır. Uzman sistemler, uzmanların bilgi ve tecrübelerini kullanabilen, bunları karar verme aşamasında değerlendirip çıkarım yapabilen bir tekniktir. Cıvatalar çeşitli parçaları birleştirme amaçlı kullanılan bağlantı elemanlarıdır ve standart makine elemanları kapsamındadır. Sektörde farklı tiplerde cıvatalar bulunmaktadır ve en doğru cıvata seçiminin yapılabilmesi uzmanlık gerektirmektedir. Bu çalışmada, baş tiplerine göre cıvata seçimi yapan Exbolt System adlı bir uzman sistem geliştirilmiştir. Cıvata baş tiplerine göre yaygın olarak kullanılan standart cıvata tipleri esas alınmıştır. Her bir cıvata ile ilgili bilgiler toplanıp analiz edilerek cıvata tipleri için ilgili kurallar oluşturulmuştur. Kurallar yazılırken, en doğru cıvata tipinin seçilmesi amaçlanmıştır. Kuralların oluşturulmasında ince ayar gereksinimi, dönen parçalarda kullanım, delikte merkezleme, düz yüzey gereksinimi, montajlama hassasiyeti, sistem ağırlık durumu, yüksek kuvvet tutabilme ve kirli ortamlarda kullanım gibi kriterler göz önüne alınmıştır. Program en iyi seçimi yaparak kullanıcıya baş tipine göre kullanılacak cıvatayı önermektedir. Exbolt System'in geliştirilmesinde CLIPS uzman sistem programlama dili kullanılmıştır. Program tarafından kullanıcıya yöneltilen sorulara alınan yanıtlar ile en doğru cıvata seçiminin yapılması sağlanmıştır. Uzmanlık gerektiren cıvata seçiminde daha hızlı, kolay ve kapsamlı bir şekilde karar verilerek en doğru sonuca ulaşılmış, zaman ve işgücünden tasarruf edilerek daha etkili ve verimli bir seçim süreci gerçekleştirilmiştir.

**Anahtar Kelimeler**— cıvata baş tipi, seçim, tasarım, uzman sistem, yapay zeka, clips

## 1. INTRODUCTION

The increase in the speed of technology development day by day has led to the widespread use of its applications in various fields. With the developments experienced, it is seen that these technologies are used by enterprises for purposes such as planning, selection, control, analysis, diagnosis and maintenance in areas such as design, supply and mass production when the engineering sector is considered. Efficiency and efficiency are increased by reducing time, cost and labour. In this way, targets can be reached more easily and quickly. With the acceleration of development and the transformation of needs into a necessity, systems called artificial intelligence (AI) that act, think, decide, learn and act like human beings are increasingly used in many fields for various purposes over time. Artificial Intelligence is defined as the branch of computer science that deals with equipping computers or machines with human mental states, such as acquiring information, perceiving, reacting, making decisions and thinking [1]. Computers greatly simplify and speed up some aspects of human thinking processes. Artificial intelligence technology applications allow complex manual operations to be performed automatically and quickly. Additionally, artificial intelligence technologies are rapidly increasing the capabilities and applicability of computers by integrating with other computer-based information systems [2]. These various applications can be realised by using artificial intelligence techniques. One of these artificial intelligence techniques is expert systems. Expert systems are computer systems that generally act as experts. Expert systems facilitate decision-making with their reasoning abilities. The most notable advantage of expert systems is that expert information is stored and made permanent so that information is not lost in any adverse situation. Other benefits are given below:

- *Productivity growth*; less labour and less cost,
- *Quality improvement*; produce consistent and appropriate results,
- *Consistency*; examining all the subject details together,
- *Flexibility*; updating knowledge bases,
- *Sophistication*; combining multiple expert knowledge,
- *Speed*; short decision-making time,
- *Reliability*; reach reliable results in a short time using the knowledge base. [3]

There are many studies on the subject in the literature. A user-interactive computer program was prepared in GW-Basic programming language in order to select and analyse standard bolts, washers, nuts, and some bolted connections under various loads using an IBM-XT compatible microcomputer [4]. An interactive online expert system was designed using the expert system development tool PC-PLUS for the selection of machine screws for the design of industrial products. The knowledge base for this system includes rules on type classification, mounting techniques and material properties of machine screws. In addition, fixing strength, corrosion resistance, magnetic properties, electrical conductivity, thermal conductivity, thermal expansion, safety, cost and appearance factors

were also taken into account to improve the quality of the selection. [5]. An intelligent software was developed that can automatically classify and select machines and machine elements, develop and implement classification algorithms using the current knowledge base [6]. By taking the expert system approach, bolt and gear wheel selections were made according to their usage areas, and the selection rules were analysed in the prepared program [7]. A study was conducted in which the application of internet technologies was discussed during the optimal selection of typical or standard machine elements in the design process, computer tools for creating engineering databases and internet applications were discussed, and optimization criteria were discussed [8]. A method was developed for the selection of the nominal diameter of ISO metric screws that can be used in high-duty bolted connections [9]. For ease of disassembly, a multi-criteria decision-making model was developed that can assist in the selection of fasteners with the PROMETHEE method within the scope of the Design for Disassembly (DfD) concept [10]. A model based on the DfD concept and Analytical Network Process (ANP) was developed for the selection of fasteners, which is one of the most influential points in the disassembly of products, among various alternatives [11]. Using the KAPPA-PC expert system package program, an expert system was developed that enables a non-expert in drilling operations to easily, quickly and accurately select insert drills and cutting parameters. [12]. An expert system called ExpertTS was developed to enable the selection of inserts for external diameter and internal threading operations [13]. An expert system was developed that can select bearings for the required features [14].

In general, expert systems were used in the development of computer applications involving other machine elements such as bolts, washers, nuts or gear wheels and their classification, selection and analysis. Additionally, computer tools for the development of applications on such subjects in the internet environment and the creation of databases were examined and optimization issues were discussed. Additionally, studies were carried out to make the most accurate selection of fasteners at the design stage to make the disassembly process based on various concepts and methods more effective and efficient.

As a result, there are very few studies with an expert system approach on the selection of bolts that can be classified under various headings such as head, body, screw and material types. Based on this, this study, unlike other studies, it is aimed to make the most accurate selection of commonly used bolt head types by considering various criteria such as the purpose of use of bolts, places of use and working conditions. In this way, a more effective and efficient selection process experience is aimed by making selections more easily, comprehensively and in a short time.

## 2. ARTIFICIAL INTELLIGENCE

Artificial intelligence (AI) is the engineering or science of making intelligent machines, especially intelligent

computer programmes, performing behaviours that are called intelligent when done by humans [15,16].

Computer programmes work with algorithms. The algorithm is the step-by-step planning of the solution path to reach the result. Intelligence is the ability to produce algorithms. AI, on the other hand, is an automatic system that can produce algorithms [17].

Artificial intelligence is a science associated with computer science and statistics. It includes topics such as expert systems, artificial neural networks, fuzzy logic, genetic algorithms, voice recognition, image processing, robotics and learning. Expert systems are one of the artificial intelligence techniques that can be applied and widely used in every field where the concept of expertise is mentioned (Figure 1).

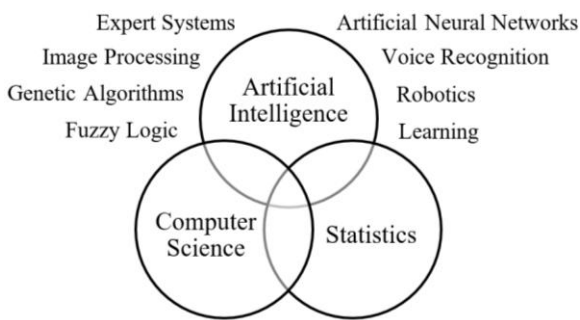


Figure 1. Artificial intelligence approach

2.1. Expert Systems

Expert systems (ES) are one of the artificial intelligence techniques that make the knowledge and experience of an expert permanent, make inferences using this data, solve problems and produce results. Expert systems are knowledge-based systems. Evaluates problems comprehensively and imitate human intelligence in solving them. In this context, it works by using algorithms and inference mechanisms interactively with each other. While establishing the ES structure, the information engineer collects the necessary information from the relevant field expert and stores it by adapting it to the knowledge base. While solving problems, ES chooses rules related to content knowledge, processes them in an inference engine, and performs thinking and analysis like an expert. It learns the solved problem and adds it to the knowledge store. ES uses can be listed as interpretation, estimation, selection, diagnosis, design, planning, imaging, debugging, repair, training and control [18].

2.1.1. General Structure of Expert Systems

Expert systems consist of three main modules as knowledge base, inference engine and user interface. During the creation and use of the expert system, it also interacts with the user and the experts (Figure 2).

- *User*: The person who uses the expert system.
- *Expert*: It is the person who forms the basis of the system and has the most knowledge and experience in the field of work.
- *Knowledge Base*: It is a structure of expert knowledge and consists of rules, facts, networks and frameworks.
- *Inference Engine*: It is the mechanism that can reach conclusions with the rules, facts and all other information in the knowledge base.
- *User Interface*: It is a component that uses the expert system and constantly interacts with it. [19,20].

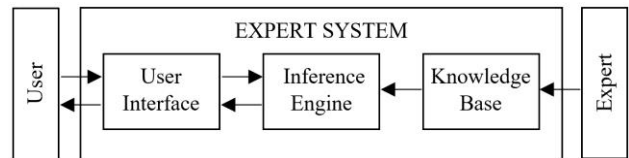


Figure 2. The general structure of expert systems [21]

The reasoning of knowledge and rules in expert systems is carried out by two different inference methods. These methods are classified as forward (data-based) and backward (goal-based).

In the forward chaining method, the reasoning unit starts from the beginning and reaches the conclusion. In this type, the "if" refers to starting over, and the "then" part refers to the result. In backward chaining, it starts from the "after" part of the result and reaches the solution by processing the "if" conditions (Figure 3). A mixture of both forward and backward reasoning can also be used.

$$\text{Rule 1} > \text{Rule 2} > \text{Rule 3} > \text{Rule n} > \text{Conclusion} \quad (a)$$

$$\text{Rule 1} < \text{Rule 2} < \text{Rule 3} < \text{Rule n} < \text{Conclusion} \quad (b)$$

Figure 3. a) Forward chaining, b) Backward chaining

The expert system development process consists of 6 basic stages given in Figure 4.

1.	Identifying and Defining the Problem
2.	Determining Implementation Criteria
3.	Creating Alternatives
4.	Solution Search and Evaluation
5.	Making Choices and Making Recommendations
6.	Application

Figure 4. Expert system development process [22]

3. BOLTS AND CRITERIA FOR BOLT SELECTION

Bolts are mainly used to connect parts detachably together. They are standard machine elements with a screw profile drilled according to certain rules on a cylindrical body and the heads of which are shaped in different ways. The part of the bolt or screw that is connected to the head of the screw and partially or completely threaded on it is called a screw. The body end is shaped as dish or conical for easy

threading of screw threads. The conically shaped screw tip is called the chamfer. Bolts are used as fastening and movement elements in the technique. Bolts are made in standard sizes based on their nominal diameters and head shapes according to their intended use [23-26]. For example, in Figure 5, the standard sections of the Hexagon head bolt are given.

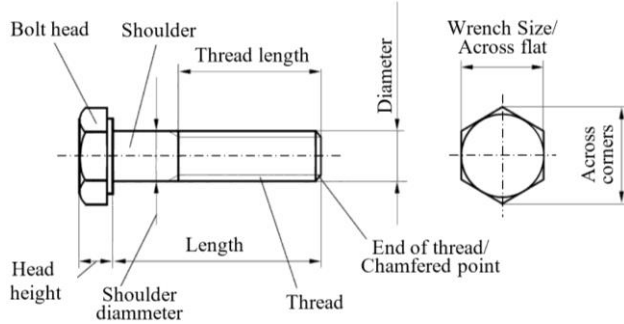


Figure 5. Parts of hex head bolt [26]

Bolted joints are used as an economical fastening method in many application areas due to their advantages such as requiring only one hole drilling or additional screwing in the parts to be connected, standardization of bolts, low-cost production thanks to mass production possibilities, and ease of assembly and disassembly. Commonly used bolt joints are loose connections made with the help of two elements, bolt and nut. They provide a highly reliable connection [27]. Bolted connections should be configured most accurately according to different criteria to be used efficiently and safely [28].

### 3.1. Classification And Characteristics of Bolts

There are classifications for bolts under various headings such as head shapes, body shapes and strength properties. In this study, there is a classification of bolts according to their head types. In order to determine the scope of the bolts, commonly used catalogues were taken as a basis. A common conclusion was reached among the catalogues and the most commonly used bolt types were determined. As a result, 5 bolt types as hexagon head bolt, cylinder head bolt, countersunk bolt, stud and setscrew were included in the study.

Information about bolts was compiled based on information and data obtained through scientific studies, technical documents and experience. Some bolt information contains generally accepted knowledge. The collected information was compared with the literature and its validity was checked. The information that was corrected at the end of the process was included in the current study. [23-27].

The features of these bolts, which are visualized below, are presented in order to create information and criteria for the expert system [23-27].

**Hexagon Head Bolt:** They are used for general-purpose applications. It is available in various sizes and dimensions. The head shape is hexagonal. It can be assembled with a wrench or socket wrench. Thanks to the hexagonal head shape, a strong tightening force can be obtained by providing high torque. These bolts are not preferred if there is any risk of leaving the bolt head out.

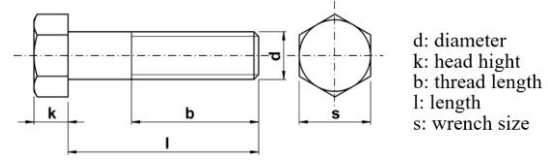


Figure 6. Hexagon head bolt

**Cylinder Head Bolt:** The head is cylindrical and the wrench flat is socket-shaped on the inside of the head. The geometry of these sockets can be in various shapes such as hexagonal, flat or star. These bolts can be mounted with tools such as hex wrenches, flat or Phillips screwdrivers according to these sockets. Thanks to its socket head, it provides ease of assembly in narrow spaces. These bolts are beneficial to be used in situations where lightness is aimed, since they have sockets and because the bolt head can be recessed into the connected part, some mass will be removed from the assembled part, and they can do the same job with a smaller number of bolts compared to other bolts.

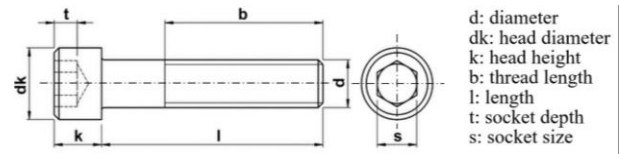


Figure 7. Cylinder head bolt

**Countersunk Head Bolt:** The head is the countersunk type and has a socket inside the head like cylinder head bolts. Depending on the socket type, they can be assembled with various wrenches, screwdrivers, etc. These bolts are inserted into the countersunk hole in the part to be assembled as recessed. These bolts can be preferred if the head of the bolt is not wanted to be outside and a flat surface is desired on the part. Due to these situations, it can be applied in situations where lightness is desired. It can be used for centring in the hole. The prices of these bolts are relatively lower in line with the catalogues.

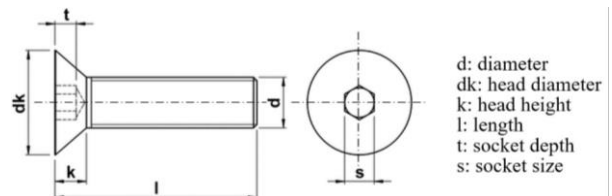


Figure 8. Countersunk bolt

**Stud Bolt:** These bolts have no heads and are threaded at both ends. One end is connected to the part to be screwed and a nut is attached to the other end. They can also be mounted using nuts for both ends. Stud bolts can be used in cases where other bolts are insufficient in terms of

length. It is generally produced in 1000-2000 mm lengths and can be used according to the need by cutting it to the desired lengths. These bolts are preferred in order not to damage the parts that are assembled due to frequent disassembly and assembly. It is used where high assembly precision is required. On the other hand, assembly operations are more difficult due to the absence of a bolt head or internal socket.

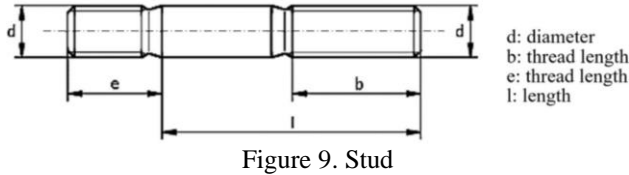


Figure 9. Stud

**Setscrew:** Headless bolts with screw threads on the body. For this reason, setscrews are used to fix rotating parts such as flywheels, gear wheels or pulleys on shafts and axles and to make fine tunings such as wear and gap adjustment. Another name is the adjustment bolt. It has sockets in various shapes to be mounted with allen wrench, etc. tightening tools at the end. Setscrews can be used in situations where a flat surface and lightness are required on the part since they do not have a head and have sockets at the end. They provide centring in the hole. If they are assembled, the level of disassembly is high. Their prices are higher compared to other bolts. For these reasons, it is recommended to be used only in special-purpose situations provided by this bolt.

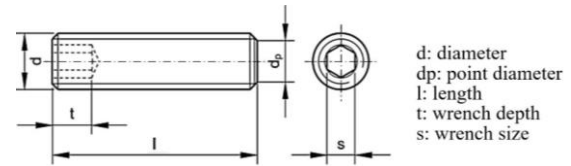


Figure 10. Setscrew

### 3.2. Bolt Selection Criteria

The criteria for each bolt were defined by creating a knowledge base based on numerical, technical and experience-based data, including the features mentioned in the previous title. In this way, the bolts were analysed according to their unique features and usage purposes, taking into account their possibilities, and separated according to head types.

For the developed expert system, these criteria, which will be applied during the selection of the most correct bolt according to the system, and the characteristics of each bolt, were brought together and correlated, and the answers to these were defined. The information obtained as a result of these transactions was summarized in Table 1 below. This information was used in the configuration of the expert system. Then, an example system for the use of bolt types is given in Figure 11.

Table 1. Bolt types and their relations with criteria

CRITERIA	BOLT TYPES				
	Hexagon H. B.	Cylinder H. B	Countersunk H. B	Stud	Setscrew
Abrasion, clearance etc. fine-tuning requirement	no	no	no	no	yes
Flywheel, pulley, gear etc. use in rotating parts	no	no	no	no	yes
Centring in hole	no	no	yes	no	yes
Flat surface requirement on the part	no	yes	yes	no	yes
Mounting accuracy	middle	middle	middle	high	low
Product/system weight status	high	low	low	high	low
High force-holding	yes	yes	yes	yes	no
Dusty, dirty etc. use in environments	yes	no	no	yes	no

In Figure 11 below, an example system showing the usage areas and purposes of the bolts is given. The details are explained on the basis of this figure, on the places where the bolts are connected. In this system, the hexagon head bolt was preferred because it does not require a special requirement in its location and there is no limitation in the area. The cylindrical head bolt was used because the mounting space is limited and it can be connected more easily, for example, with a allen key. The countersunk bolt was preferred

in order to provide a flat surface and centering in the hole. Stud bolt was preferred because it can be found in long lengths and it is advantageous for mounting accuracy, and it was mounted in the position where it is seen in the system. Because there is a need for long bolts in order to make the connection in the region where it is located and a sensitive assembly is required. Setskur bolt was used to fix the circular part to the shaft and to secure its position.



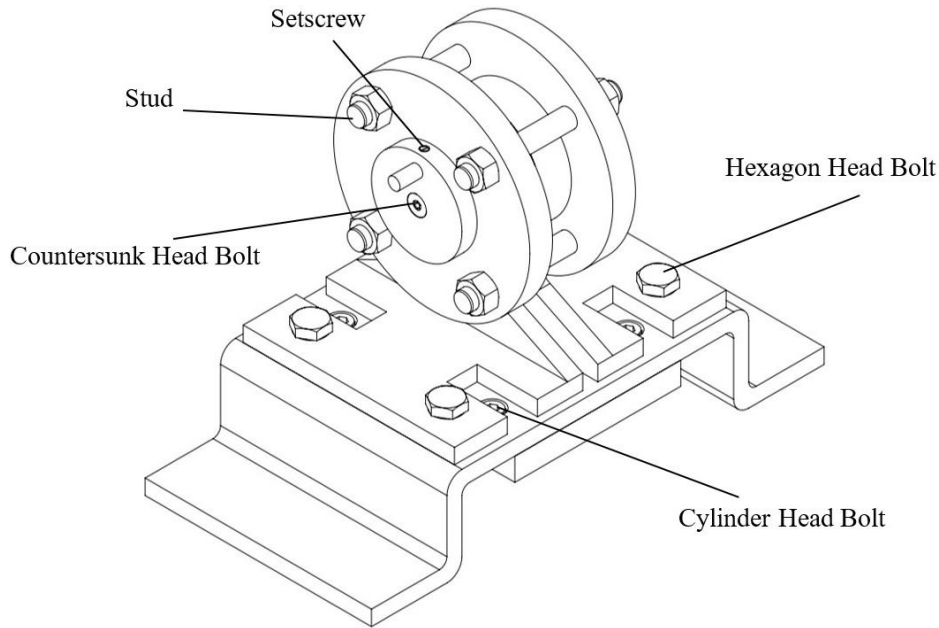


Figure 11. An example system for the use of bolt types

3.3. Design Rules and Criteria

The intended use of bolt head types is from general to more specific. As one enters from general use to more specific use, the availability of bolts according to head type decreases and their costs increase. According to the head types, the use of bolts from general to more specific can be listed as hexagon head bolt, cylinder head bolt, countersunk bolt, stud and setscrew. This situation was determined through data obtained from catalogues widely used in the market.

User responses to bolt selection may require bolts with more specific uses. In this case, the selection process tends towards these bolts. If user responses do not require special

use bolts, the trend in the selection process is towards the more general use bolt head type.

With this expert system, questions are asked to the user about situations such as usage purposes, usage areas and working conditions. Questions are asked to the user from more specific to more general situations. The responses received are controlled by criteria related to these situations. Matching bolt head types are determined as a result of the control. If only one bolt head type is specified, it is reflected to the user as the most correct choice. If more than one bolt head type is specified, the general and special usage situation is taken into consideration and the more general use bolt head type is recommended to the user. The flow of design rules with Figure 12 is given below with the general programming structure.

**Bolt types according to head shapes:**

- Hexagon Head Bolt (HHB)
- Cylinder Head Bolt (CyHB)
- Countersink Head Bolt (CHB)
- Stud (STD)
- Setscrew (SSW)

**User questions and possible answers:**

- [1] Abrasion, clearance etc. fine-tuning requirement: (y) yes, (n) no
- [2] Flywheel, pulley, gear etc. use in rotating parts: (y) yes, (n) no
- [3] Centring in hole: (y) yes, (n) no
- [4] Flat surface requirement on the part: (y) yes, (n) no
- [5] High mounting accuracy requirement: (y) yes, (e) either
- [6] Intended system weight status: (l) light, (h) heavy, (e) either
- [7] High force-holding: (y) yes, (n) no
- [8] Dusty, dirty etc. use in environments: (y) yes, (n) no

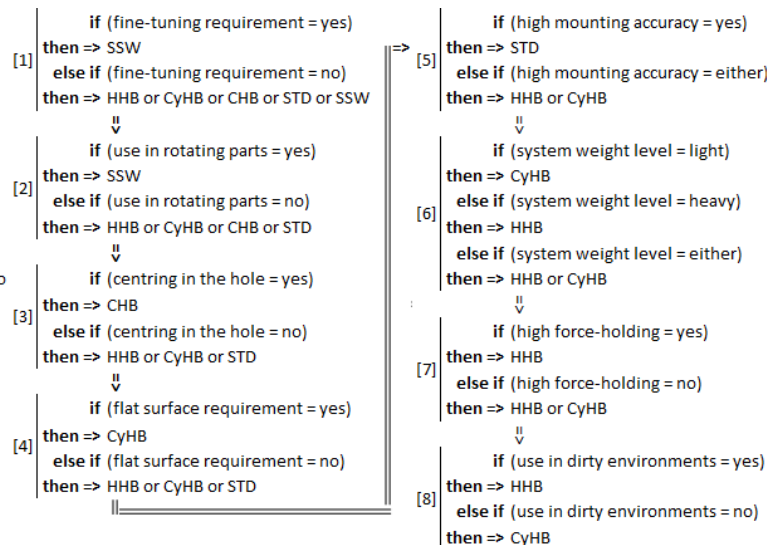


Figure 12. Design rules

#### 4. CLIPS EXPERT SYSTEM PROGRAMMING LANGUAGE AND TOOL

In this study, CLIPS expert system programming language and tool was used.

CLIPS (C Language Integrated Production System): CLIPS is an expert system programming language and expert system creation tool. Developed at NASA's Johnson Space Center from 1985 to 1996, CLIPS is a rule-based programming language useful for building expert systems and other programmes where a heuristic solution is easier to implement and maintain than an algorithmic solution. Written in C for portability, CLIPS can be installed and used on a wide variety of platforms. Since 1996 CLIPS has been available as public domain software [29].

The CLIPS program is widely used due to its features such as being rule-based, using the forward chaining method, being open-source software, being easily available, being able to run on many systems, interacting with other programs, ease of use, and having comprehensive user guides. Because of these features, the use of CLIPS was preferred.

##### 4.1. CLIPS Structure and Images

Some rule structures of the CLIPS programming language and then the interface and usage of the CLIPS 6.4 expert system tool are presented.

##### 4.1.1. CLIPS Rule Structure

The "defrule" structure, which is one of the CLIPS structures, is used for rule definition. The part before the "=>" statement is an "if" statement. The part after "=>" indicates a "then" structure. However, conditions can be created for desired situations with the expressions "or" and "and". It is sufficient to meet one of the conditions for the expression "or". For the "and" statement, all the conditions specified must be met.

```
(defrule rule1 ""
(or (criteria1 answer 1)
(and (criteria1 2 answer2) (criteria1 3 answer 3)))
=> (assert (conclusion inference)))
```

Figure 13. CLIPS rule structure

##### 4.1.2. CLIPS Expert System Tool

The operating system in which the CLIPS programme is installed has two launcher applications, CLIPSDOS, which uses the DOS environment, and CLIPSIDE, which uses its editor. This study was carried out on CLIPSIDE.

The environment tab is actively used on the CLIPSIDE window. The developed expert system is loaded into the programme with the "Load Construction" option under this

tab. When the programme is wanted to be run, the "Run" option is clicked to run the expert system. If the system is wanted to be restarted after the expert system is started and terminated, the "Reset" option is selected to delete the remaining data from the system that was previously operated and the system is reset. If desired, all texts on the editor can be cleared with the "Clear Scrollback" option. After these steps, the expert system can be "Run" again (Figure 14).

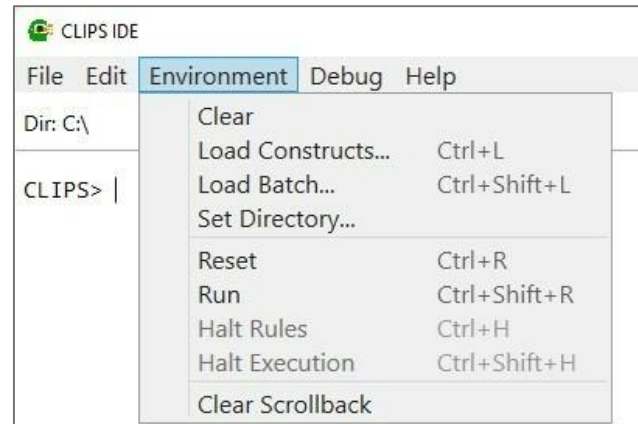


Figure 14. CLIPSIDE editor and Environment tab

#### 5. DEVELOPED EXPERT SYSTEM

An expert system was developed to select the bolt type and to make this selection most correctly. CLIPS 6.4 expert system tool and programming language were used in the development of this system called *Exbolt System*.

*Exbolt System*: It is an expert system developed with the CLIPS programming language in order to select the most correct bolt type to be applied to a system. In this expert system, a total of 8 criteria was determined for 5 bolt types. At least 40 rules were developed for the criteria. Relevant rules were defined by using information and data obtained through scientific studies, technical documents and experience. In the query rules, questions were written in order to determine the selection criteria by taking answers from the user. The number of questions may vary depending on the answers given by the user. A series of scope narrowing processes are carried out according to the answers given to the questions. Thus, a selection process was structured to reach the most accurate bolt type. The selected bolt is printed on the program screen and recommended to the user.

After the program is run, it returns results once. If the result is to be obtained again, the system must be reset. Thus, it can be restarted and the result can be obtained. Since the selection process is based on clear cause-effect relationships, it can produce the most accurate results every time. In this way, results can be achieved in a logical way.

*Exbolt System* rules was created with codes written with criteria that can provide bolt selection.

Figure 15 and example result rules of the *Exbolt System* are given below. In the "if" structure of the "hbbconc" rule, which expresses the Hexagonal Head Bolt type, the "weight", "forceholding" and "dirtyenvironment" criteria are checked in the "or" bracket. When the user answers the questions "heavy", "yes" and "yes" to the criteria of "weight", "holding strength" or "dirtyenvironment", respectively, if any of these conditions are met, the "if" part of the rule is passed to the "then" part. With the "assert" statement in this section, "Hexagon Head Bolt" information is assigned to the result variable and the most accurate bolt-type result is reached. Other rules are checked similarly according to the answers given by the user to the questions. Thanks to this system, the relations of the answers with the criteria are checked, the conclusion rule that meets the conditions is run and the result is reached.

```
(defrule hhbconc ""
(or (weight heavy) (forceholding evet) (dirtyenvironment yes))
=> (assert (conclusion "Hexagon Head Bolt")))

(defrule chbconc ""
(or (flatsurface yes) (weight light) (dirtyenvironment no))
=> (assert (conclusion "Cylinder Head Bolt")))

(defrule choconc ""
(centring yes)
=> (assert (conclusion "Countersunk Head Bolt")))

(defrule studconc ""
(mountingaccuracy yes)
=> (assert (conclusion "Stud")))

(defrule setscrewconc ""
(or (finetuning yes) (rotating yes))
=> (assert (conclusion "Setscrew")))
```

Figure 15. A sample rules part of the expert system

### 5.1. Application Example

The images of the process of operating the expert system called *Exbolt System* through the CLIPSIDE programme are given below with their explanations. These explanations will be based on an example flanged pipeline system with bolted joints.

This flanged system is widely used and this system mostly uses the hexagon head bolt type. Because of these situations, it was seen as a reliable sample and was preferred.

The developed expert system and established rules were tested on the already existing bolted flange connection. In order for the expert system to produce accurate results, the working environment and conditions of the bolted flanged connection were analysed in detail. In this direction, the expert system test process was carried out. The results show that the rules work correctly in the selection of the bolt head type and thus give a logical output. In this section, a detailed explanation of the case study is given.

The purpose of use, places of use and working conditions of the flanged pipeline system and bolts, which are considered as an example for the control of selection criteria, are examined. Accordingly, fine-tuning conditions such as wear and clearance are not required. It is also not used for fixing rotating parts such as flywheel, pulley and gear wheel. High centring is not intended in the holes where the bolt connection will be made. It is not a problem to have the bolt head and the nut on the surface of the part, so there is no requirement for a flat surface. Connections do not require high assembly precision. It is not a question of aiming to make the system light or heavy. Additionally, this system does not require high-force holding. However, the system is used in an environment with dust, dirt, etc. (Figure 16).



Figure 16. Bolt-on flanged pipeline system

Considering these working conditions, the process of running the expert system called the *Exbolt System* was carried out. Accordingly, the CLIPS programme was opened via the IDE, and the user is greeted by an editor that included the programme's version number and date information. The expert system file with the name and extension *ExboltSystem.clp*, which was developed with the CLIPS programming language, was loaded into the CLIPSIDE environment by specifying the location with the "Load Construction" option under the "Environment" tab (Figure 17).

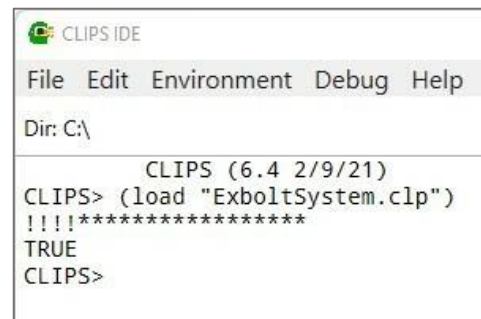


Figure 17. Installing the expert system called Exbolt System

Following the installation process, the "Run" statement was selected on the Environment tab on the CLIPS window, and the expert system was run. Then the first question was directed to the user by CLIPS (Figure 18).



*Exbolt System* has reached the right result in the selection of bolt type that requires expertise.

## 6. DİCCİSSION

Artificial intelligence is a technology used to enable a computer system to mimic real human thought and behaviour. By using artificial intelligence, machine learning, data mining, neural networks and other technologies, it enables systems to learn, understand and be understood and continuously improved.

Artificial intelligence can be applied in various fields. For example, it can be used in speech recognition, image recognition, language processing and automatic machine learning. Artificial intelligence is also used in various industries such as machinery, robotics, automatic control, finance and e-commerce.

Artificial intelligence has developed rapidly in recent years and is finding more applications thanks to technologies such as larger datasets, more powerful processors and more complex algorithms. In the coming years, artificial intelligence will improve people's lives and will be used more and more in various sectors.

In the machinery sector, it is used for selection, diagnosis, maintenance and even more different purposes. When the studies carried out for the purpose of selection of artificial intelligence are examined, it was seen that there are studies on the selection of machinery and machine elements. One of these machine elements is bolts. Studies was carried out on the selection of bolts under various classifications. In the present study, it was aimed to choose the most correct bolt type according to the head types. For this purpose, an expert system was developed. The results of this developed system were checked over an existing system. Program outputs and obtained results were compared with previous studies and their aims, differences and superiority over each other were evaluated.

The *Exbolt System* test process was carried out taking into account the place of use and operating conditions of the already existing flanged connection, which is considered as an example. The bolt type obtained at the end of the test process was compared with the bolt type used on the flanged connection and it was seen that a correct selection process was carried out (Figure 20).

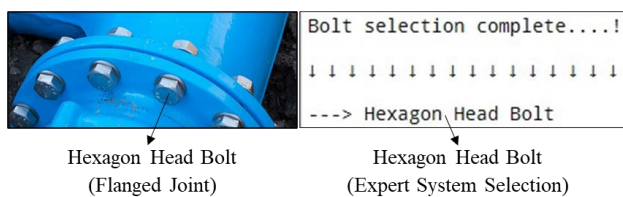


Figure 20. Comparison of existing flanged joint and developed expert system

In the closest study to this study in the literature, an expert system was developed for the selection of machine screws for the design of industrial products. This system was created using the PC-PLUS expert system tool [5].

The current study, which is compared with this study in the literature, considers bolts and bolt head types instead of screws. Different head shapes were used in the studies. Therefore, the scope of bolt head types in this study was formed in a different way. In the selection of bolt type, more places of use, purposes and working conditions were taken into account. Thanks to these features, the *Exbolt System* has become a more targeted study.

The work in the literature includes rules on assembly techniques and material properties. In addition, fixing strength, corrosion resistance, magnetic properties, electrical conductivity, thermal conductivity, thermal expansion, safety, cost and appearance factors were also taken into account to improve the quality of the selection. The *Exbolt System*, on the other hand, is based on the selection of bolt head types. For this, the place and purpose of use, working conditions, including criteria such as the need for fine-tuning, use in rotating parts, centring in the hole, flat surface requirement, mounting accuracy, system weight status, high force-holding and use in dirty environments are taken into account.

## 7. CONCLUSIONS

The selection of bolts, which are used for various purposes in almost every field from past to present, is complex, thought-provoking and therefore a time-consuming process. To improve this situation and to select the truest bolt, an expert system called *Exbolt System* was developed with the CLIPS expert system programming language. For this developed bolt selection system, a total of 8 bolt selection criteria was determined, including concepts such as working conditions and mounting situations, and the number of which may vary according to the answers given by the user. Catalogues that are widely used in the sector were examined and a common conclusion was reached and the most commonly used bolt types were determined. As a result, 5 bolt types were reached and an expert system was designed accordingly. The programme and the questions directed to the user was structured in a clear, clear and understandable way. Time is saved by asking only the necessary questions depending on the answers. The developed expert system evaluates the answers to the questions according to the determined criteria and suggests the user by reflecting the truest bolt for the relevant situation on the programme screen. With the expert system called *Exbolt System*, which has an effective structure, the most accurate result was achieved in a simpler and shorter time without the need for an expert in the bolt selection process, which requires expertise, is complex and takes time. In this way, a more effective and efficient selection process was passed.

With the expert system developed within the scope of this study, the head type selection of the bolts was handled in a different dimension compared to the literature and various criteria such as the purpose of use of the bolts, the places of use and working conditions were taken into consideration. This study is based solely on the head type selection of the bolts. In this direction, subjects such as bolt threads and profiles, material information are not included in the scope of the study. By using this expert system at the beginning of the design process, the user can prevent unforeseen limitations and errors that may be created by future usage conditions before they occur. In this way, labour and time savings can be achieved by minimizing design repetitions and tests.

Based on this study, the scope of bolts can be expanded by including other bolt types in future studies, and a more detailed selection process can be carried out by increasing the number of criteria. The material properties of bolted connection systems can be included in the study. A more extended expert system can be written by conducting various analyses such as static and dynamic within the programme. By developing a user interface and providing expert system integration, an interactive visual expert system that can provide visual content and technical information to the user can be developed.

## REFERENCES

- [1] G. Güzeldere, "Yapay Zekanın Dünü Bugünü Yarını", *Cogito*, (13), 27-42, 1998.
- [2] İnternet: A. Kebude, Yapay Zekâ, İnternet: <https://books.google.com.tr/books?id=FNZcEAAAQBAJ>, 03.04.2023.
- [3] D. İçen, S. Günay, "Uzman Sistemler ve İstatistik", *İstatistikçiler Dergisi: İstatistik ve Aktüerya*, 7 (2), 37-45, 2014.
- [4] Y. Levent, **Interactive computer aided selection of standard bolts and analysis of bolted joints**, Yüksek Lisans Tezi, Orta Doğu Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, 1990.
- [5] J. L. Chen, H. S. Yan, G. Y. Shine, "An expert system for machine screws selection in engineering design", *Expert Systems with Applications*, 4(1), 141-146, 1992.
- [6] E. S. Alp, **Design and development of an expert system shell spesific to selection of machines and machine elements**, Yüksek Lisans Tezi, Orta Doğu Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, 1997.
- [7] M. T. Özkan, M. Gülesin, "The Selection of Bolts and Gears Through Expert System Approach", *Turkish Journal of Engineering and Environmental Sciences*, 25 (3), 169-177, 2001.
- [8] J. Wróbel, S. Tomaszek, R. Raczko, "İnternet technologies in optimal selection of typical or standardized machine elements", *Sbornik vědeckých prací Vysoké školy báňské - Technické univerzity Ostrava: Řada hutnická*, 51(1), 167-171, 2008.
- [9] D. Croccolo, M. De Agostinis, N. Vincenzi, "A contribution to the selection and calculation of screws in high duty bolted joints", *International Journal of Pressure Vessels and Piping*, 96, 38-48, 2012.
- [10] R. A. R. Ghazilla, Z. Taha, S. Yusoff, et al., "Development of decision support system for fastener selection in product recovery oriented design", *Int J Adv Manuf Technol*, 70, 1403-1413, 2014.
- [11] Jeandin T., Mascle C., **A New Model to Select Fasteners in Design for Disassembly**, *Procedia CIRP*, 40, 425-430, 2016.
- [12] Y. Kayır, E. Demirer, S. Güneş, "Takma Uçlu Matkap ve Kesme Parametrelerinin Seçimi İçin Bir Uzman Sistem". *El-Cezeri*, 5(3), 797-806, 2018.
- [13] Y. Kayır, S. Güneş, E. Demirer, "Vidaların Açılmasında Kesici Uç Seçimi İçin Bir Uzman Sistem", *Gazi Mühendislik Bilimleri Dergisi*, 4(3), 176-182, 2018.
- [14] Ö. Asal, Y. Kayır, R. Mergen, "Rulman Seçimi İçin Bir Uzman Sistem", *Gazi Mühendislik Bilimleri Dergisi (GMBD)*, 5(3), 216-226, 2019.
- [15] A. G. H. Pirim, "Yapay Zekâ", Yaşar Üniversitesi, *Yaşar Üniversitesi E-Dergisi*, 1(1), 81-93, 2006.
- [16] K. K. Çevik, E. Dandıl, "Yapay Sinir Ağları İçin .NET Platformunda Görsel Bir Eğitim Yazılımının Geliştirilmesi", *Bilişim Teknolojileri Dergisi*, 5(1), 19-28, 2012.
- [17] Y. Köroğlu, "Yapay Zekâ'nın Teorik ve Pratik Sınırları", *Boğaziçi Üniversitesi Yayinevi*, İstanbul, Türkiye, 2017.
- [18] A. Kastal, A. O. Köse, "Yapay Zekâ", **XIV. Türkiye'de İnternet Konferansı**, Bilgi Üniversitesi, Aralık 12-13, 2009.
- [19] D. Yıldız, "Bilgi Yönetiminde Kural Tabanlı Uzman Sistem Geliştirme Adımları ve Başarı Faktörleri", *Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi*, 5 (1), 28-43, 2021.
- [20] M. Bozdemir, "Mekanik Tasarım Eğitimi İçin Bir Uzman Sistem Uygulaması", *Uludağ Üniversitesi Mühendislik Fakültesi Dergisi*, 24 (1), 219-230, 2019.
- [21] B. Başoğlu, M. Bulut, "Kısa dönem elektrik talep tahminleri için yapay sinir ağları ve uzman sistemler tabanlı hibrit sistem geliştirilmesi", *Journal of the Faculty of Engineering & Architecture of Gazi University*, 32(2), 2017.
- [22] İnternet: İstanbul Teknik Üniversitesi. Uzman Sistemler Giriş. [https://web.itu.edu.tr/~sonmez/lisans/es/uzman\\_sistemler\\_giris.pdf](https://web.itu.edu.tr/~sonmez/lisans/es/uzman_sistemler_giris.pdf), 03.04.2023.
- [23] D. Çoban, B. S. Ünlü, "Civata Ve Somunların Çeşitleri Ve Üretimi", *Soma Meslek Yüksekokulu Teknik Bilimler Dergisi*, 2 (20), 36-48, 2015.
- [24] İnternet: Karadeniz Teknik Üniversitesi. Civata Somun Bağlantıları. <https://avesis.ktu.edu.tr/resume/downloadfile/kucuncu?key=42326233-22d8-427b-a057-aab99291edf2>, 03.04.2023.
- [25] İnternet: MEGEP: Mesleki Eğitim ve Öğretim Sistemini Güçlendirme Projesi. Birleştirme Elemanları. [http://www.megep.meb.gov.tr/mte\\_program\\_modul/moduller\\_pdf/Birle%C5%9Firme%20Elemanlar%C4%B1.pdf](http://www.megep.meb.gov.tr/mte_program_modul/moduller_pdf/Birle%C5%9Firme%20Elemanlar%C4%B1.pdf), 03.04.2023.
- [26] İnternet: İTÜ. Civatalar. <https://web.itu.edu.tr/temizv/Sunular/Civata.pdf>, 03.04.2023.
- [27] Bursa Teknik Üniversitesi. Civata Bağlantıları. [https://sayfam.btu.edu.tr/upload/dosyalar/14549712667.BOLUM\(CIVATA\\_BAGLANTILARI\)\\_4.12.2014.pdf](https://sayfam.btu.edu.tr/upload/dosyalar/14549712667.BOLUM(CIVATA_BAGLANTILARI)_4.12.2014.pdf), 03.04.2023.

- [28] İ. Saraç, “Cıvatalı bağlantılarda delik toleransı ve cıvata sıkma torku değişiminin bağlantı dayanımına etkilerinin sayısal olarak incelenmesi”, *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 37 (2), 685-696, 2022.
- [29] Internet: CLIPS. CLIPS: A Tool for Building Expert Systems, <https://www.clipsrules.net>. 03.04.2023.

# Yapay Sinir Ağları Kullanılarak Protein Katlanması Tanıma

*Araştırma Makalesi/Research Article*

 Sena DİKİCİ\*,  Volkan ALTUNTAŞ

Bilgisayar Mühendisliği, Bursa Teknik Üniversitesi, Bursa, Türkiye

[sena.dikici@btu.edu.tr](mailto:sena.dikici@btu.edu.tr), [valtuntas@gmail.com](mailto:valtuntas@gmail.com)

(Geliş/Received:06.07.2022; Kabul/Accepted:06.03.2023)

DOI: 10.17671/gazibtd.1141468

**Özet**— Proteinler uzun aminoasit zincirlerinden oluşur ve vücut kimyasını düzenlemekle birlikte hücrelerin yapısı ve aralarındaki iletişim için öneme sahiptir. Bir proteinin hücre bazındaki görevini gerçekleştirebilmesi için, molekülü hücredeki hedefiyle etkileşime girebilecek üç boyutlu yapıya dönüştüren bir bükülme süreci olan katlanma işlemini gerçekleştirmesi gerekir. Sıcaklık, ağır metaller veya kimyasal durumlar gibi etkenler proteinlerin yanlış katlanmasına sebep olabilir. Yanlış katlanan proteinler, vücuttaki görevini yerine getiremez. Alzaymır, kistik fibrozis, deli dana hastalığı gibi hastalıklara sebep olabilir. Protein katlanması tanıma işlemi, biyologlar açısından bir problem olarak değerlendirilir. Literatürde yer alan şablon tabanlı yaklaşımlara karşın yapay sinir ağları, protein katlanması probleminin çözümüne yönelik yüksek başarımlar gösterir. Yapay sinir ağları, ele alınan problemin çözümü için geniş veri kümelerinde yer alan ve problemin çözümüne katkı sağlayacak bilgi kazancı yüksek özellikleri kullanan bir hesaplama tekniğidir. Bu çalışmada SCOPe 2.06, SCOPe 2.07, SCOPe 2.08 veri setleri kullanılarak şablon tabanlı yaklaşımlardan elde edilen sonuçların yapay sinir ağı yöntemi ile birleştirilerek protein katlanması tanıma işlemi gerçekleştirilmiştir. Gerçekleştirilen deneyler sonucunda yapay sinir ağı yönteminin katkısı ile literatürde yer alan sonuçların iyileştirildiği görülmüştür. Bu çalışma ile biyoinformatik alanında protein katlanması tanıma probleminin çözümüne yeni bir yaklaşım sunularak literatüre katkı sağlanması amaçlanmıştır.

**Anahtar Kelimeler**— protein katlanması, yapay sinir ağları, protein katlanması tanıma, biyoinformatik

## Protein Folding Recognition by Artificial Neural Networks

**Abstract**— Proteins are made up of amino acid chains and are important for the structure of the cells and their communication with each other, while regulating body chemistry. For a protein to perform its function on a cell basis, it must perform the folding process that converts the molecule into a three-dimensional structure that can interact with its target in the cell. Factors such as temperature, heavy metals or chemical conditions may cause proteins to be folded incorrectly. Incorrectly folded proteins cannot perform their role in the body. Protein folding recognition is considered a problem for biologists. Despite the template-based approaches in the literature, artificial neural networks show high performance in solving the protein folding problem. Artificial neural networks are a computational technique that uses high-specification knowledge in large data sets to help solve the problem that is being addressed and will contribute to the solution of the problem. In this study, protein folding recognition was performed by combining the results from template-based approaches with artificial neural network method using the Scope 2.06, Scope 2.07, Scope 2.08 datasets. The results in the literature were improved by the contribution of the artificial neural network method because of the experiments conducted. This study aims to contribute literature by introducing an innovative approach to the solution of the problem of protein folding recognition in the field of bioinformatics.

**Keywords**—protein folding, artificial neural network, protein folding recognition, bioinformatics



## 1. GİRİŞ (INTRODUCTION)

Proteinler, organizmaların faaliyetlerinde önemli rol oynayan bileşenlerdir. Bir proteinin işlevi, diğer proteinlerle etkileşime girmesi ve katlanmasıyla belirlenir. Proteinlerin görevleri ve birbirleriyle etkileşimleri, proteinlerin yapılarının sınıflandırılması açısından önemlidir. Proteinlerin üçüncül yapılarının tanımlanması, protein fonksiyon gösterimi, ilaç tasarımı, moleküler biyoloji gibi alanlar için en önemli görevlerden biridir. RNA (Ribonükleit asit) 'ya kopyalanan genetik bilginin bir protein veya protein zinciri haline dönüşmesi sırasınad, her protein lineer bir amino asit zinciri veya sabit bir üç boyutlu yapısı olmayan rastgele bir bobin olarak sentezlenir. Zincirdeki amino asitler, sonunda iyi tanımlanmış, katlanmış bir protein oluşturmak için birbirleriyle etkileşime girer. Aminoasit dizileri proteinlerin üç boyutlu yapılarını belirler. Protein katlanması, proteinlerin dizileri ile yapıları arasındaki ilişkiyi açıklar. Bir polipeptit zincirinin biyolojik olarak aktif bir protein haline gelmesi için katlandığı sürece protein katlanması denir. Proteinlerin katlanma yapıları, fonksiyonlarının tanınması için önemli rol oynar. Amino asit dizilerindeki geniş çeşitlilik ise protein yapısındaki farklı konformasyonları açıklar. Mevcut durumdaki konformasyonda proteinler kararlılığını sürdürdüğü takdirde kendilerine ait görevleri yerine getirebilmektedir. Bu görevler biyolojik işlemlerin doğru şekilde devamlılığını sağlar.

Bir proteinin genel katlanma türünün karakterize edilmesi önemli bir etkidir. Proteinin yapısal sınıfının ilk tanımı 1976 yılında Levitt ve Chothia tarafından yapılmıştır. Birinci aşama ( $\alpha$ ), az miktarda iplikten oluşur. İkinci aşamada ( $\beta$ ), proteinler dört yapısal sınıftan ve az sayıda sarmaldan oluşur. Üçüncü aşama ( $\alpha/\beta$ ), sarmallardan ve paralel dizilerden oluşur. Son aşama ( $\alpha+\beta$ ) ise sarmallardan ve paralel olmayan dizilerden oluşur [1]. Proteinlerin yapısal sınıfı veya katlanma türü, moleküler biyoloji alanında önem taşır. Proteinlerin, dizi kimliklerinden bağımsız olarak katlanma türü sayesinde üç boyutlu yapısı tespit edilebilmektedir [2].

Sıcaklık, ağır metaller veya kimyasal olaylar gibi etkenler proteinlerin yanlış katlanmasına sebep olmaktadır. Proteinlerin yanlış katlanması, hastalıklara sebep olabilmektedir. Örneğin, bazı kanser türleri p53 ismi verilen proteinin, bu protein kanser proteini olarak da adlandırılmaktadır, yanlış katlanması sebebiyle meydana gelmektedir [3]. Beyin hücrelerindeki proteinlerin yanlış katlanması ve kontrol edilemez boyuta ulaşması sonucunda alzaymır ve deli dana hastalığı meydana gelmektedir. Protein katlanmasının doğru tahmini, yanlış katlanma sonucu meydana gelen hastalıkların erken ve hızlı teşhis edilmesi açısından önem taşımaktadır.

Protein katlanması tanıma, hedef proteinlerin sadece sekans bilgilerine dayanarak bilinen protein yapı şablonlarına göre sınıflandırmayı amaçlayan bir problemdir. Şablon tabanlı yaklaşımlardan elde edilen sonuçların iyileştirilmesi amacıyla bu yaklaşımlar makine

öğrenmesi ya da derin öğrenme yöntemleri ile birleştirilebilmektedir. Çalışmamızda da protein katlanması türlerinin, yapay sinir ağları kullanılarak tespit edilmesi ile literatürde yer alan sonuçların iyileştirilmesi amaçlanmıştır. Çalışmamızın protein katlanması tanıma için zaman ve doğruluk değerleri kriterleri açısından olumlu yönde katkı sağlayacağı öngörülmektedir. Hedef ve sorgu protein dizileri arasında şablon tabanlı yaklaşımlardan elde edilen üç farklı benzerlik puanı yöntemine ait sonuçlar kullanılmıştır. Benzerlik puanları, yapay sinir ağları ile eğitilerek elde edilen sonuçlar literatüre sunulmuştur.

## 2. LİTERATÜR (RELATED WORK)

Biyoloji ve tıp alanında yapılan çalışmaların günümüze yaklaştıkça bilgisayar bilimleri yöntemleri ile birleştirildiği çalışmaların yaygınlaştığı görülmüştür [4,5,6,7]. Protein katlanması tanıma problemi için bilgisayar bilimleri ile birleştirilerek yapılan çok sayıda çalışma mevcuttur. Protein katlanması tanıma işlemi proteinlerin 3 boyutlu yapılarıyla ilişkilendirilmektedir. Bu doğrultuda protein yapısının protein sekansları üzerinden tahmin edilmesi, problemin önemi ve iyi tanımlanmış hesaplamalı temelleri sebebiyle uzun süredir araştırmacılar tarafından incelenmektedir. Bu çalışmalar bilgisayar bilimleri ile birleştirilerek yeni bir çalışma alanı oluşturmuştur. Makine öğrenmesi yöntemleri ile protein yapısının tahmin edilmesi ile ilgili çalışmalar yapılmıştır. Derin artık ağlar protein yapılarının tahmini için yaygın olarak kullanılan yöntemlerden biridir. Örneğin, ResNet mimarisi birçok çalışmada yer alan sistemler için taban kabul edilen mimari olmuştur. Çalışmalar sonucunda ResNet mimarisinin güçlü olduğu ve sonuçları iyileştirdiği literatüre sunulmuştur [8]. Protein katlanması tanıma problemi için matematiksel temellere dayalı geleneksel yöntemler literatürde yer almaktadır. Belirli algoritmalar kullanılarak şablon tabanlı yaklaşımlar ile katlanma türü tanıma işlemi gerçekleştirilmiştir. Bu yaklaşıma sahip çalışmalardan birinde yerel uyarlanabilir komşu bağlantısı (local adaptive neighbor connection) yöntemi kullanılmıştır. 23 farklı protein kullanılan çalışmada yerel uyarlanabilir komşu bağlantısı yönteminin 18 protein için en yüksek kalite değerine sahip olduğu, 12 protein için ise ya en hızlı ya da en hızlı ikinci performansla sahip olduğu görülmüştür [9]. Makine öğrenmesi yöntemleriyle gerçekleştirilen çalışmaların başarısı sayesinde biyoinformatik alanı literatürde ve araştırmalarda güncelliğini korumaktadır. Protein katlanması tanıma problemi de bu kapsamda yer almaktadır. Protein katlanması ile ilgili çalışmalardan birinde SPARKS-X (Diziden Yapı Tahmini - Structure Prediction from Sequence), HHblits (HMM-HMM tabanlı ışık hızında yinelemeli dizi arama - HMM-HMM-based lightning-fast iterative sequence search) ve DeepFR (Derin Katlanma Tanıma - Deep Fold Recognition) şablon tabanlı yaklaşımlarından benzerlik puanları elde edilmiştir. SPARKS-X, sorgunun tahmin edilen tek boyutlu yapısal özellikleri ile şablonların karşılık gelen doğal özellikleri arasında olasılığa dayalı eşleştirmeyi kullanarak gelişen bir protein katlama tanıma ve şablon tabanlı modelledir. HHblits ise protein yapısının ve işlevinin diziyeye dayalı

doğru tahmini için geliştirilmiş Gizli Markov modeli tabanlı yinelemeli bir dizi arama yöntemidir. DeepFR ise süper aile ve katlanma seviyelerinde protein katlanması tanıma için derin evrişimli sinir ağı kullanan bir yaklaşımdır [10]. Elde edilen benzerlik puanları, global bir özellik matrisinde birleştirilmiştir. Proteinlerin dizi yapılarından elde edilen benzerlik matrisi, destek vektör makinesine eğitim verisi olarak verilmiştir. Bu yaklaşıma TSVM-fold ismi verilmiştir. Destek vektör makinesi yöntemi için farklı çekirdek fonksiyonları kullanılarak yapılan çalışmada en yüksek doğruluk değeri %63 olarak elde edilmiştir [11]. Şablon tabanlı yaklaşımlarla protein katlanması tanıma konulu başka bir çalışmada ise SPARKS, FOLDpro, SPARKS-X ve BoostThreader yöntemleri kullanılmıştır. Çalışmada kullanılan şablon tabanlı yaklaşımlardan elde edilen doğruluk değerleri sırasıyla %47.7, %48.8, %67, %57.4 olarak elde edilmiştir [12]. Makigaki ve Ishida'nın gerçekleştirmiş olduğu çalışmada FFAS (Katlama ve İşlev Atama Sistemi - The Fold and Function Assignment System), HHPRED ve SPARKS-X şablon tabanlı yaklaşımlar kullanılarak sırasıyla %47, %47, %52 benzerlik puanı elde edilmiştir [13]. HHPRED, protein homoloji tespiti ve yapı tahmini için etkileşimli sunucudur [14]. Bin Liu ve diğerlerinin çalışmasında ise çoklu dizi hizalaması yaklaşımı, destek vektör makinesi yöntemi ile birleştirilmiştir. SCOP veri tabanının 1.67 versiyonu kullanılmış ve %78 doğruluk değeri elde edilmiştir [15] Sudha ve diğerleri EDD, DD, TG, RDD veri setleri üzerinde yapay sinir ağı yöntemini kullanarak protein yapısal sınıflandırması ve katlanması tanıma için yeni bir yöntem önermişlerdir. Daha önce yapılan çalışmalardan elde edilen özneliklerin farklı kombinasyonları kullanılarak yapay sinir ağı yöntemini, destek vektör makinesi ve Bayes yöntemi ile elde ettikleri deneysel sonuçlarla karşılaştırarak yapay sinir ağı yönteminin yüksek başarımla elde ettiği sonucunu sunmuşlardır. EDD, DD, TG ve RDD veri setlerine ait veriler yapay sinir ağları ile eğitildiğinde elde edilen doğruluk değerleri sırasıyla %83, %76.6, %76, %73.3 olmuştur [2]. Wei ve Zou'nun çalışmasında n adet özellik temsili için n adet tekli sınıflandırıcı eğitilerek topluluk sınıflandırıcı stratejileriyle birleştirilmiştir. Çalışmada 5 farklı veri seti topluluk sınıflandırıcı yöntemi ile protein katlanma sınıflarına ayrılarak elde edilen sonuçlar karşılaştırılmıştır. Çalışma sonucunda DD veri setinin dengesiz olduğu, SCOP veri tabanının ise fazla sayıda katlanma türü içermesi sebebiyle kısıtlar oluşturduğu sunulmuştur. ProFold sistemi, topluluk sınıflandırıcı yöntemi kullanılarak %76.2 değer ile en yüksek doğruluk değerine sahip olmuştur [16]. Başka bir çalışmada ise proteinlere ait aminoasit özelliklerinin farklı kombinasyonları kullanılarak çalışmanın başarımla değerlendirilmiştir. Bu çalışmada göreceli temas sırası modeli ve sabit parametreler kullanılarak, mutlak temas sırası modeli ve sabit parametreler kullanılarak, bağıl temas sırası modeli ve değişken parametreler kullanılarak, mutlak temas sırası modeli ve değişken parametreler kullanılarak farklı deneyler gerçekleştirilmiştir. Bu deneyler sonucunda en yüksek korelasyon değerine sahip deney mutlak temas sırası modeli ve değişken parametreler kullanılarak elde edilmiştir [17]. Protein katlanması ve dinamiği ile ilgili çalışmalardan bazıları

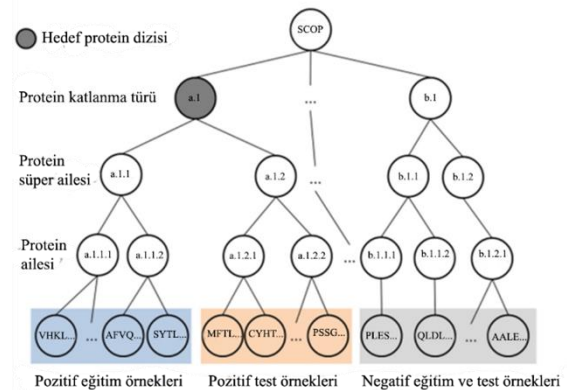
verilerden elde edilen bilgilerin makine öğrenmesi yöntemleri kullanılarak simüle edilebilmesine yöneliktir. Proteinlerdeki yarı kararlı durumların ve bunların istatistiklerinin çıkarılması için Markov Durum Modelleri (Markov State Models) kullanılmaktadır. Protein katlanması ve açılmasının hızlı simülasyonu, protein dinamikleri hakkında da hızlı bir şekilde bilgi edinilmesini sağlamaktadır. Yapılan çalışma makine öğrenmesi yöntemlerinin sıklıkla kullanılmasına da çalışma sonuçlarında iyileştirme sağladığını göstermiştir [18].

### 3. MATERYAL VE METODLAR (MATERIAL AND METHODS)

Bu bölümde çalışmada kullanılan veri seti, veri setlerinde mevcut olan ve çalışmada kullanılan öznelikler, çalışmada kullanılan yapay sinir ağları yöntemi açıklanmıştır. Yapay sinir ağları yöntemine ait fonksiyonlar ve algoritmalar açıklanmıştır. Modelin başarımla ölçütlerine yer verilmiştir.

#### 3.1. Veri Seti (Dataset)

Çalışmamızda SCOPE veri tabanında yer alan veri setleri kullanılmıştır. SCOPE, Berkeley Laboratuvarlarında geliştirilen bir veri tabanıdır. Proteinlerin yapısal sınıflandırılması amacıyla oluşturulmuş olup proteinlerin yapısal ilişkilerine ait veriler yer almaktadır. SCOPE veri tabanı, protein yapılarını ve dizilerini analiz etmek için oluşturulmuştur. Proteinler arasındaki yapısal ve evrimsel ilişkilerin doğru, ayrıntılı ve kapsamlı bir tanımını sağlamayı amaçlamaktadır. Protein akrabalıkları ve protein kıvrımlarına ilişkin geniş bir araştırma sonucu oluşturulan SCOPE veri tabanı, özellikle protein dizilerinden tanınmayacak kadar eski olan veriler hakkında güvenilir bilgiler sağlayarak, araştırmalar ve sınıflandırma için bir çerçeve oluşturmaktadır [19]. Bu çalışmada SCOPE veri tabanının 2.06, 2.07 ve 2.08 versiyonu kullanılmıştır. SCOPE 2.06 veri setinde 244.332 adet veri ve 1431 katlanma türü, SCOPE 2.07 veri setinde 276.231 adet veri ve 1457 adet katlanma türü, SCOPE 2.08 veri setinde 344.848 adet veri ve 1485 adet katlanma türü mevcuttur.



Şekil 1. SCOP veri setinin sınıf yapısı [11].  
(The structure of SCOP dataset)

Şekil 1’de SCOPe veri setinin sınıf yapısına göre protein katlanma türü, protein süper ailesi ve protein ailesi sınıflandırmasının nasıl yapıldığına ilişkin gösterim yer almaktadır.

### 3.2. Özellik Seçimi (Feature Selection)

Özellik seçimi mevcut veri kümesinin en iyi temsilini oluşturan alt kümenin belirlenmesi olarak tanımlanmaktadır. Veri tabanındaki özellik sayısı azaltılarak veri boyutunun azaltılması sağlar. Özellik seçimi için bilgi kazanımı (information gain) yöntemi kullanılmıştır. Karar ağaçları, kök düğümden başlanarak aşağıya doğru karar verici yönlendirmeler ile oluşturulan ağaçlardır. Ağacı inşa etmek için entropi kullanılmaktadır. Aslında kullanılan yöntem entropi oluşturmaktadır. Entropiyi kullanan bu yöntem bilgi kazanımı denmektedir. Entropi, tanımlandığı her alandaki düzensizliğin ölçülmesi için kullanılmaktadır. Denklem 1’de entropi formülü verilmiştir. Entropi H sembolüyle gösterilmektedir. H, bir bilginin bir ortamdaki düzensizliğini belirlemek için kullanılmaktadır. Bu noktada bilginin aynı tek düze yapıda akması düzensizliği ortadan kaldırır. Entropi formülünde n verinin miktarını ve  $p_i$  ise ilgili verinin olma olasılığını ifade etmektedir. Bilgi kazanımı ise durumlar arası entropi farkından oluşmaktadır. Denklem 2’de ise bilgi kazanımına ait formül verilmiştir.

$$H = - \sum_{i=1}^n p_i \log(p_i) \quad (1)$$

$$\begin{aligned} \text{Bilgi Kazanımı (S,A)} &= H(S) - \sum_{t \in T} p(t) H(t) \\ &= H(S) - H(S | A) \end{aligned} \quad (2)$$

#### 3.2.1. DM Puanı (DM Score)

Çalışmada eğitilmek üzere, SCOPe veri setlerinde yer alan proteinlere ait özelliklerden bilgi kazanımı sonucunda yüksek puana sahip özellikler seçilmiştir. Bu özelliklerden biri DM puanıdır.

Protein diziliminin analizi ve yapılarının karşılaştırılması ile protein katlanma türü sınıflandırılır [20]. Katlanma türü, proteinlerin ikinci aşamada karşılık gelen yapılarının kararlılığına göre sınıflandırılmıştır. Korunan ortak çekirdeğin çoklu yapısal hizalaması ile karşılık gelen dizilerin korelasyonu tartışılmıştır.

Bir proteinin yapısı, proteinlere ait moleküllerin farklı kısımlarında katlanma sonucu göstereceği değişikliğe ait atomik koordinatların listesi sayesinde elde edilmektedir [21]. Bir proteine ait yüzeyin bağlı konumu, proteinin göreviyle yakından ilişkilidir; yüzeye ait parçalar diğer moleküllerle doğrudan etkileşime girer. Proteinlerin yapısı itibarıyla birbirlerinden farklı görevleri gerçekleştiren

grupların vermiş olduğu kimyasal tepkimeler, proteinlerin tepkimeye girdiği arayüzle ilişkilerine bağlıdır. Lee ve Richard algoritmasında protein yapısında bulunan her bir atom veya atom grubu, tanımlandığı boyutta (çözücü veya çözünen) bir moleküle erişebilme yeteneğinin listesi sayesinde düzensiz yüzeyin açıklaması yapılır [21]. Hesaplamalar baz alındığında titreşimler ya da esneklik hakkında hiçbir bilgi yansıtmayan proteinin koordinat listesinde, türetilen sayılar statik erişilebilirliğe sahiptir. Bu tür sayıların, kimyasal verilerin yorumlanması için koordinatların kendisinden daha yararlı olacağı öne sürülmüştür.

Geliştirilen bilgisayar programları ile bir protein molekülünün içinde en az bir çözücü molekülü barındıracak büyüklükteki boşluklar tespit edilebilmektedir. Bu, protein bölümlerinde içbükey eğriliğe sahip tüm erişilebilirlik yüzeylerini yalıtarak ve bu içbükey yüzeyin içinden molekülün dışına giden bir kanala sahip olanları ortadan kaldırarak yapılmaktadır [19]. Aminoasit kalıntılarının temsil ettiği kümenin bir alt kümesini belirlemek için, her bir koordinat veri kümesinde analiz gerçekleştirilmiştir. Kimyasal çözücünün erişebildiği temas yüzey alanı, Richmond ve Richards’ın (1978) algoritması kullanılarak, o katlama birimi çevresindeki her atom için hesaplanmıştır. Katlama ve işlev için gerekli olan ligandlar, merkezi bir metale bağlanan bir atom veya molekül, erişilebilirlik hesaplamasına dahil edilip gruplara bağlı inhibitörler dahil edilmemiştir [21].

Her atom için angstrom karesi cinsinden erişilebilirlik hesaplaması elde edildikten sonra, ana zincir ve yan zincir için kalıntılar toplanmıştır. Yan zincir boyutundaki varyasyon ve dolayısıyla yüzey alanındaki varyasyonlar, genişletilmiş bir konformasyon için erişilebilir yüzey alanı teorik olarak hesaplanmıştır ve toplamlar normalleştirilmiştir [21]. Daha sonra çözücü molekülle etkileşime giren yan zincirin yüzdelik değerleri elde edilmiştir [22].

Lee ve Richards algoritması ile Hubbard ve Blundell yaklaşımı birleştirilerek proteinlere ait DM puanı elde edilmiştir [22].

#### 3.2.2. SP Puanı (SP Score)

Bilgi kazanımı denklemi (Denklem 2) ile hesaplama yapıldıktan sonra yüksek puana sahip olarak seçilen diğer özellik ise SP puanıdır. Bu yaklaşımda protein dizilerinin karşılaştırılması için uzak homolog proteinlerin, protein veri bankasında yer alan; moleküllerin yapısal olarak bilinen protein alanları ve zincirleri için hesaplanan yapısal hizalamalarının tutulduğu DASH veri tabanı kullanılmıştır. Ölçüm sistemleri analizi kullanılarak SP puanı hesaplanıp literatürde yer alan diğer ölçümlere göre %10 civarında iyileşme elde edildiği görülmüştür [23]. DASH hizalamaları önceden hesaplandığından ek hesaplama maliyeti ve ek ağ yükünden tasarruf sağlanmıştır. MAFFT-DASH yaklaşımı, son kullanıcı yükünü azaltarak sıralama ve yapısal hizalama bilgilerini entegre etmenin oldukça

uygun ve verimli bir yolunu sunmaktadır. Ayrıca düşük hesaplama maliyetleri ile doğru hizalamalar sağlamaktadır.

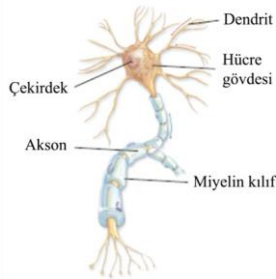
### 3.2.3. CL Puanı (CL Score)

Bilgi kazanımı yöntemi ile en yüksek puana sahip olan üçüncü özellik CL puanı olmuştur. CL puanı çalışmada kullanılmak üzere seçilmiştir. DM puanı, SP puanı ve CL puanları kullanılarak protein katlanması tanıma işlemi gerçekleştirilmiştir.

Biyokimya alanında bir biyomoleküle bağlanarak kompleks yapı oluşturan bileşiklere ligand denir. Ligand moleküller genellikle hedef proteinlerdeki bağlanma yerlerine iyonik bağlar, hidrojen bağları veya Van der Waals güçleri ile bağlanır [24]. Protein-ligand kompleksleri ise kristal yapılarındaki artış ile kanonik olmayan etkileşimlerin anlaşılmasına yardımcı olmuştur. Bu etkileşimler sayesinde kristal yapı veri tabanlarının kapsamlı analizi ile kanonik olmayan etkileşimlerinin yeni bir puanlama sistemiyle yeni bilgiler edinilmesi ve fonksiyonlarının geliştirilmesi sağlanmıştır. Yapılan çalışma sayesinde ligandlarda Cl grubunun aromatik halkaya bağlandığı bulunmuştur [25]. Protein-ligand komplekslerinin Cl etkileşimleri analiz edilerek protein veri bankası için yeni bir puanlama yöntemi ortaya konmuştur. Bu sayede etkileşimlerin doğası ve Cl etkileşimlerinin geometrisi anlaşılır hale gelmiştir [25]. Cl puanı sayesinde protein katlanması sırasında oluşan kompleks yapıların anlaşılır kılınması amaçlanmıştır.

### 3.3. Yapay Sinir Ağları (Artificial Neural Network)

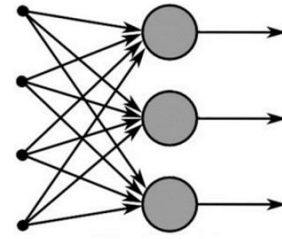
Yapay sinir ağları, insan beyninin çalışma şeklinin bilgisayar ortamına uyarlanması olarak tanımlanır. İnsanlara benzer şekilde deneyim kazanarak ve uygun metotlarla öğrenir. Yapay sinir ağları, verilerdeki kalıpları ve ilişkileri tespit ederek bilgileri toplar [26]. Nöronlar, hücre işleyişinin kontrolünü sağlayan çekirdek, hücreye bilgi taşıyan dendrit ve sinyali uzağa taşıyan akson olarak adlandırılan hücre gövdesinden oluşmaktadır (Şekil 2). İmpuls akson boyunca sinapsa, bir nöron ile bir sonraki arasındaki bağlantıya geçer ve sinyaller birinden diğerine ya hep ya hiç şeklinde iletilir [26]. Nöronlar, birbirlerine tamamen bağlıdır ve öğrenme, tahmin etme ve tanıma yeteneğine sahiptir [26].



Şekil 2. Nöron yapısı [27].  
(Structure of a neuron)

Yapay sinir ağı, sinir yapısını oluşturan katsayılar (ağırlıklar) ile bağlantılı yüzlerce tek birimden, yapay nöronlardan oluşan biyolojik olarak esinlenilmiş bir hesaplama modelidir. Yapay sinir ağındaki bulunan her bir verinin kendisine ait ağırlıkları, transfer edilme süreci ve ürettiği sonuç (çıkıtı) değeri vardır. Yapay sinir ağları girdi katmanında bulunan verilere bağlı olarak çıktı üreten bir denklem olarak da tanımlanır. Bağlantı ağırlıkları sistemin hafızasını temsil ettiği için yapay sinir ağlarına bağlantıcı modeller de denir. Tasarlanmış birçok sinir ağı türü vardır ve hepsi nöronlarının transfer fonksiyonları, öğrenme kuralı ve bağlantı formülü ile tanımlanabilir [26].

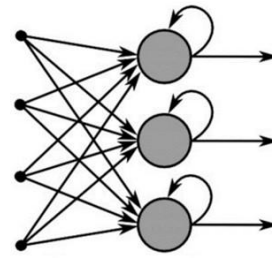
Yapay nöronlar uyarıcı veya engelleyici girdiler alabilir. Uyarıcı girdiler, bir sonraki nöronun toplama mekanizmasının eklenmesine neden olurken, engelleyici girdiler çıkarılmasına neden olur. Bir nöron aynı katmandaki diğer nöronları da inhibe edebilir. Buna lateral inhibisyon denir. Ağ, en yüksek olasılığı seçmek ve diğer olasılıkları engeller [28].



Şekil 3. İleri beslemeli ağ [29].  
(Feedforward network)

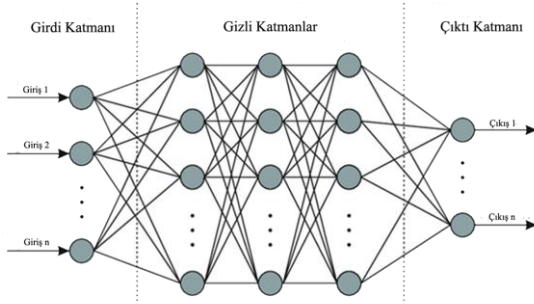
Geri bildirim, bir katmanın çıktısının bir önceki katmanın girişine veya aynı katmana geri yönlendirildiği başka bir bağlantı türüdür. İleri beslemeli mimaride ise girdi katmanından çıktı katmanına doğru ileri bağlantılar mevcuttur. Geri dönüş yapılmaz. Bu nedenle önceki çıktı değerlerinin kaydı tutulmaz (Şekil 3) [29].

Geri besleme mimarisi ise ileri besleme mimarisinin aksine çıktı katmanından girdi katmanında bulunan nöronlarla bağlantı kurabilir. Her nöronun eğitim sırasında belirli bir hata değerine sahip olması olasıdır. Geri beslemeli ağlar önceki durumda elde edilen sonuçları hafızasında tutar (Şekil 4). Önceki sonuçların hafızada tutulması sebebiyle bir sonraki durum mevcut giriş sinyallerinin yanı sıra ağın önceki durumlarına bağlıdır [29].



Şekil 4. Geri beslemeli ağ [29].  
(Feedback network)

Çalışmamızda kullanılan yapay sinir ağı yapısı bir adet giriş katmanı, 3 adet gizli katman ve bir adet çıktı katmanından oluşmaktadır. Şekil 5'te temsili gösterimi verilmiştir. Nöron ve parametre sayıları Tablo 1'de verilmiştir.



Şekil 5. Yapılan çalışmanın yapay sinir ağı mimarisinin şematik gösterimi [30]  
(Schematic representation of the artificial neural network architecture of the study)

### 3.3.1. Aktivasyon Fonksiyonu (Activation Function)

Tarihine bakıldığında, sigmoid fonksiyonu, çok katmanlı algılayıcı ve Boltzmann Makinesi gibi yapay sinir ağlarında nöronların aktivasyon işlevi olarak sıklıkla kullanılmıştır [30]. Sigmoid fonksiyonu veya tanjant hiperbolik fonksiyon gibi aktivasyon fonksiyonlarında hata değeri minimum hale getirilmek istendiğinde gradyan değeri sıfır olabilmektedir. Gradyan değerinin sıfır olması problemine, kaybolan gradyan problemi denmektedir. Kısıtlı Boltzmann Makinesini yöntemine ait bu problemin çözümüne yönelik olarak Nair ve arkadaşları doğrultulmuş doğrusal birimleri (ReLU) tanıttı [31].

Glorot, gizli katmanlar için uygulanan aktivasyon fonksiyonunun ReLU fonksiyonu olarak tercih edildiğinde, fonksiyonun kullanıldığı derin sinir ağlarının öğrenme hızının arttığını söylemiştir. ReLU, derin sinir ağları için standart olarak kullanılır [32]. Kaybolan gradyan sorununun ReLU aktivasyon fonksiyonu kullanılarak önenebileceği görülmüştür [32].

ReLU fonksiyonunun temel avantajı aynı anda tüm nöronları aktive etmemesidir. Bir nöron, negatif değer ürettiğinde ReLU fonksiyonu nöronun aktive edilmemesini sağlamaktadır. Bu sebeple çok katmanlı sinir ağlarında ReLU aktivasyon fonksiyonu kullanılabilir. Çalışmamızda geliştirilen model ise giriş katmanı, 3 adet gizli katman ve çıktı katmanından oluşması sebebiyle ReLU aktivasyon fonksiyonu kullanılmıştır.

### 3.3.2. Kayıp Fonksiyonu (Loss Function)

Yapay sinir ağlarının başarımını değerlendirme yöntemlerinden biri kayıp fonksiyonunun değeridir. Eğitilen modelin tahmin ettiği değerler ile gerçekte bilinen değerler arasındaki fark ile ifade edilir. Elde edilen değer sayesinde modelin ne kadar geliştirebileceği hakkında bilgi sahibi olunmasını sağlar [33].

Çalışmamızda kayıp fonksiyonu olarak kategorik çapraz düzensizlik fonksiyonu (categorical cross entropy) kullanılmıştır. Çapraz entropi hata fonksiyonu, eğitim esnasında oluşabilecek yavaşlamaların önüne geçmek amacıyla kullanılan bir fonksiyondur.

Aktivasyon fonksiyonlarının ardından kayıp fonksiyonları kullanılmaktadır. Çalışmamızda çoklu sınıflandırma gerçekleştirildiği için kategorik çapraz entropi (KÇE) kayıp fonksiyonu kullanılmıştır. Aktivasyon fonksiyonları çıktı olarak her bir sınıfın olasılık değerini verir. Fonksiyon çıktısı sıfır ile bir arasında değer almaktadır. Bu eşitlik kullanılarak sıfırdan büyük ve ağ çıktısının beklenen değere yakın olduğu durumlarda entropinin sıfıra yakın olduğu görülmektedir [34].

$$KÇE = - \sum_{q=1}^l \sum_{k=1}^p d_{qk} \log(y_{qk}) \quad (3)$$

Denklem 3'te kategorik çapraz entropi kayıp fonksiyonu verilmiştir. Denklemde yer alan p, sınıf etiketi sayısını; l ise veri sayısını temsil etmektedir. Fonksiyon çıktıları 0 ile 1 aralığında olduğundan  $y_{qk}$ , [0,1] aralığındadır.  $d_{qk}$  değerleri gerçek değerleri,  $y_{qk}$  değerleri tahmin değerlerini temsil etmektedir.

### 3.3.3. Optimizasyon Algoritması (Optimization Algorithm)

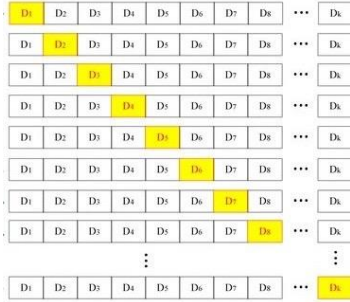
Yapay sinir ağlarının hata oranını en aza indirmek için optimizasyon algoritmaları kullanılmaktadır. Öğrenme işlemi, filtre katsayıları ve katmanlar arası ağırlık değerlerinin veriyi temsil eden uzayında optimum olarak belirlenmesiyle gerçekleşmektedir. Çalışmamızda, optimum çözüme ulaşmada iyi performans gösteren Adam optimizasyon algoritması, öğrenme algoritması olarak seçilmiştir [35].

### 3.4. K-Katlamalı Çapraz Doğrulama (K-Fold Cross Validation)

K-Katlamalı Çapraz doğrulama yöntemi modelin değerlendirilmesinde kullanılan yöntemlerden biridir. Aşırı öğrenme (over fitting) ve eksik öğrenmenin (under fitting) tespit edilmesini sağlamaktadır. Aşırı öğrenme meydana geldiğinde model eğitim veri seti içerisindeki fonksiyonel örüntünün yanı sıra gürültülü verileri de öğrenir. Buna ezberleme de denmektedir. Ezberleme ise modeli eğitim veri setine bağlı kılar bu sebeple model yeni veriler üzerinde başarılı tahminde bulunamaz [36]. Eksik öğrenme ise modelin veri setindeki örüntüleri eksik öğrenmesinden kaynaklanmaktadır. Eksik öğrenme modelin genelleme yapmasına sebep olur. Aşırı öğrenme ve eksik öğrenme zayıflıklarının önüne geçmek için k-katlamalı çapraz doğrulama yöntemi kullanılmaktadır [37].

Şekil 6'daki akış şeması, k-katlı çapraz doğrulamanın, verileri rastgele K gruplara bölerek başladığını ve ardından her grup için aşağıdaki işlemlerin gerçekleştirildiğini göstermektedir [38].

Öncelikle veri seti üzerinde K adet parçaya bölünmüş veriler içerisinde bir adet alt küme belirlenerek test kümesi oluşturulur. Kalan K-1 parça ise eğitim kümesi olarak kullanılır [38]. Bu işlem K kez tekrarlanır. Her iterasyonda sıradaki alt küme test verisi olarak kullanılır.



Şekil 6. K-Katlamalı çapraz doğrulama süreci gösterimi (Sarı renkli kareler test veri setini, beyaz renkli kareler eğitim veri setini temsil etmektedir.) [38]  
(The main process of K-fold cross-validation)

### 3.5. Başarım Ölçütü (Performance Measure)

Yapay sinir ağlarının başarımını değerlendirmek için kullanılan ölçütlerden biri de doğruluk değeridir. Doğru sınıflandırılan örnek sayısının toplam örnek sayısına bölünmesi ile doğruluk değeri elde edilir. Yanlış sınıflandırılmış örnek sayısının toplam örnek sayısında bölünmesi ile de hata oranı elde edilir.

$$\text{Doğruluk} = \frac{DP+DN}{DP+YP+DN+YN} \quad (4)$$

$$\text{Hata oranı} = \frac{YP+YN}{DP+YP+DN+YN} \quad (5)$$

Denklem 1 ve Denklem 2’de yer alan DP, doğru pozitif; DN, doğru negatif; YP, yanlış pozitif ve YN, yanlış negatif anlamlarını taşır.

## 4. BULGULAR (RESULTS)

Çalışmamızda SCOPe 2.06, SCOPe 2.07 ve SCOPe 2.08 veri setleri yapay sinir ağı yöntemleri kullanılarak eğitilmiştir. Değerlendirme ölçütü olarak K-katlamalı çapraz doğrulama yöntemine ait doğruluk değerleri alınmıştır. Katlama sayısı için 5, 10, 15 ve 20 değerleri ile deneyler gerçekleştirilmiştir. Gerçekleştirilen deneyler sonucunda doğruluk değeri en yüksek K değeri seçilmiştir.

Çalışmamızda kullanılan modelde yapay sinir ağlarına ait nöron sayısı, parametre sayısı, kayıp fonksiyonu, aktivasyon fonksiyonu ve optimizasyon algoritması Tablo 1’de verilmiştir.

Tablo 1. Scope 2.06 veri seti için çalışmada kullanılan modele ait parametreler  
(Parameters of the model used in the study for the SCOPe 2.06 dataset)

	Giriş Katmanı	Birinci Gizli Katman	İkinci Gizli Katman	Üçüncü Gizli Katman	Çıktı Katmanı
Nöron Sayısı	7	32	64	128	1430
Parametre Sayısı	0	256	2112	8320	184470
Kayıp Fonksiyonu	Kategorik Çapraz Düzensizlik Fonksiyonu (Categorical Cross Entropy)				
Aktivasyon Fonksiyonu	ReLU Fonksiyonu				
Optimizasyon Algoritması	Adam Optimizasyon Algoritması				

Her bir veri seti için yapay sinir ağının eğitiminde kullanılan veri sayısı, veri setine ait sınıf sayısı ve parametre sayıları Tablo 5’te verilmiştir. SCOPe 2.06, SCOPe 2.07 ve SCOPe 2.08 veri setleri sırasıyla genişletilmiş veri setleridir. Veri sayısı ve sınıf sayısının arttığı görülmektedir.

Tablo 2. Farklı veri setlerine ait veri, sınıf ve parametre sayısı  
(Number of data, classes and parameters for different data sets)

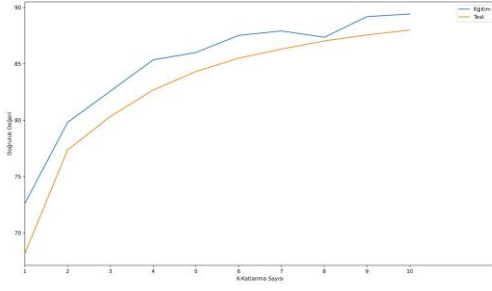
Veri Seti	Veri Sayısı (Eğitim)	Sınıf Sayısı	Parametre Sayısı
SCOPe 2.06	171028	1431	195287
SCOPe 2.07	193361	1457	198641
SCOPe 2.08	241395	1485	202253

SCOPe 2.06 veri seti ile yapılan çalışmada, katlama sayılarına ait doğruluk ve kayıp değerleri Tablo 3’te verilmiştir.

Tablo 3. SCOPe 2.06 veri seti için katlama numaralarına ait doğruluk ve kayıp değerleri  
(Accuracy and loss values for fold numbers for SCOPe 2.06 dataset)

Katlama Sayısı	Doğruluk Değeri	Kayıp Değeri
1	72.6189	1.0517
2	79.8059	0.7604
3	82.5768	0.6191
4	85.3476	0.5192
5	86.0025	0.5192
6	87.5209	0.4938
7	87.9133	0.4472
8	87.3526	0.4274
9	89.1822	0.4256
10	89.4155	0.3754
<b>Ortalama</b>	84.7737	0.5685

SCOPe 2.06 veri seti ile katlama sayısı 10 alınarak gerçekleştirilen deneyde ortalama doğruluk değeri 84.77 elde edilmiştir. Literatürde yer alan diğer çalışmalara kıyasla iyi bir sonuç elde edilmiştir. SCOPe 2.06 veri setine ait katlama sayılarına göre eğitim ve test doğruluk değerlerinin grafiksel gösterimi Şekil 7’de verilmiştir.



Şekil 7. SCOPE 2.06 veri seti ile yapılan deneysel çalışmanın grafiksel gösterimi

(Graphical representation of the experimental work with the SCOPE 2.06 dataset)

Şekil 7’de görüldüğü üzere katlama sayılarının her biri için eğitim doğruluk değerleri test doğruluk değerlerinden yüksek elde edilmiştir. SCOPE 2.07 veri seti ile yapılan çalışmada, katlama sayılarına ait doğruluk ve kayıp değerleri Tablo 4’te verilmiştir.

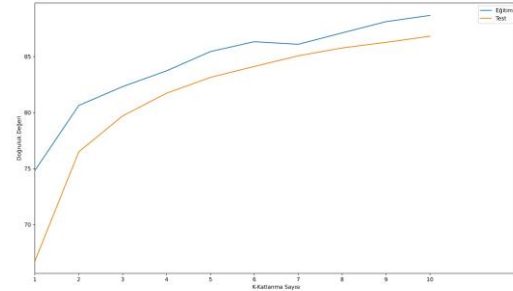
Tablo 4. SCOPE 2.07 veri seti için katlama numaralarına ait doğruluk ve kayıp değerleri

(Accuracy and loss values for fold numbers for SCOPE 2.07 dataset)

Katlama Sayısı	Doğruluk Değeri	Kayıp Değeri
1	74.8298	1.0474
2	80.6320	0.7370
3	82.314	0.6558
4	83.7237	0.5792
5	85.4432	0.5496
6	86.3229	0.4906
7	86.0912	0.4863
8	87.1085	0.4560
9	88.1113	0.4210
10	88.6689	0.4229
<b>Ortalama</b>	<b>84.3247</b>	<b>0.5846</b>

SCOPE 2.07 veri seti ile katlama sayısı 10 alınarak gerçekleştirilen deneyde ortalama doğruluk değeri 84.32 elde edilmiştir. Literatürde yer alan diğer çalışmalara kıyasla iyi bir sonuç elde edilmiştir. Ancak SCOPE 2.06 veri seti ile kıyaslandığında daha düşük doğruluk değeri elde edilmiştir. SCOPE 2.07 veri setine ait katlama sayılarına göre eğitim ve test doğruluk değerlerinin grafiksel gösterimi Şekil 8’de verilmiştir.

Şekil 8 incelendiğinde doğruluk değerlerinin katlama sayılarına göre sapmalar içerdiği görülse de değerlerin başarılı olduğu söylenebilmektedir. SCOPE 2.08 veri seti ile yapılan çalışmada, katlama sayılarına ait doğruluk ve kayıp değerleri Tablo 5’te verilmiştir. SCOPE 2.08 veri setine ait katlama sayılarına göre eğitim ve test doğruluk değerlerinin grafiksel gösterimi Şekil 9’da verilmiştir.



Şekil 8. SCOPE 2.07 veri seti ile yapılan deneysel çalışmanın grafiksel gösterimi

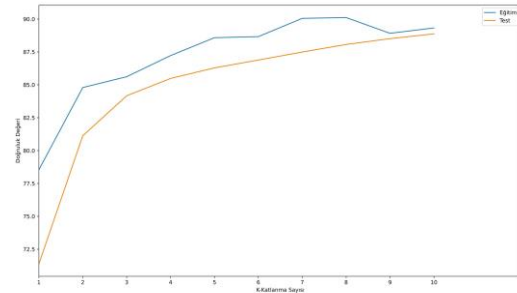
(Graphical representation of the experimental work with the SCOPE 2.07 dataset)

Tablo 5. SCOPE 2.08 veri seti için katlama numaralarına ait doğruluk ve kayıp değerleri

(Accuracy and loss values for fold numbers for SCOPE 2.08 dataset)

Katlama Sayısı	Doğruluk Değeri	Kayıp Değeri
1	78.5391	0.8482
2	84.7904	0.5661
3	85.6140	0.5088
4	87.2205	0.4507
5	88.5805	0.4208
6	88.6173	0.4103
7	90.0507	0.3539
8	90.1116	0.3628
9	88.9140	0.3915
10	89.3228	0.3668
<b>Ortalama</b>	<b>87.1806</b>	<b>0.4680</b>

SCOPE 2.08 veri seti ile gerçekleştirilen deneyde ortalama doğruluk değeri 87.18 elde edilmiştir. Çalışmamızda kullanılan diğer veri setleri ile kıyaslandığında en iyi sonuç SCOPE 2.08 veri seti ile elde edilmiştir.



Şekil 9. SCOPE 2.08 veri seti ile yapılan deneysel çalışmanın grafiksel gösterimi

(Graphical representation of the experimental work with the SCOPE 2.08 dataset)

Şekil 9’da yer alan grafik için de katlama sayıları aralarında aşağı ya da yukarı yönlü sapmalar olduğu görülmektedir. Ancak eğitim ve test doğruluk değerleri arasındaki sapmalar doğruluk değerleri ile değerlendirildiğinde uygun olduğu söylenebilmektedir. Tüm veri setlerine ait ortalama kayıp ve doğruluk değerleri varyansları ile hem eğitim hem de test sonuçları için Tablo 6 ve Tablo 7’de verilmiştir.

Tablo 6. Veri setlerine ait eğitim sonuçları  
(Train results for data sets)

Veri Seti	Ortalama Kayıp Değeri	Ortalama Doğruluk Değeri	Varyans
SCOPE 2.06	0.5685	84.7737	± 4.9408
SCOPE 2.07	0.5846	84.3247	± 3.9687
SCOPE 2.08	0.4680	87.1806	± 3.3379

Veri setlerinin veri ve sınıf sayısı bakımından genişletilmesinin deney sonuçlarında olumlu yönde katkı sağlayacağı öngörülmüştür. Eğitim sonuçları incelendiğinde SCOPE 2.06 veri setine kıyasla daha güncel bir versiyon olan SCOPE 2.07 veri setinin, doğruluk değeri düşük çıkmıştır. Varyans değerleri ile karşılaştırma yapıldığında SCOPE 2.07 veri setinin SCOPE 2.06 veri setinden daha tutarlı olduğu söylenebilmektedir. SCOPE 2.06 ve SCOPE 2.07 veri setleri ile elde edilen sonuçlardaki varyans farkı, ortalama doğruluk değerleri karşılaştırıldığında kabul edilebilir görülmüştür.

Model üzerinde eğitim işlemi gerçekleştirilirken çalışmada kullanılan üç veri seti için de her katlamaya ait döngü sayısı (epoch) 300 olarak alınmıştır. Deneysel açıdan karşılaştırılabilir olması açısından döngü sayısı, döngü boyutu, katman sayısı, model yapısı gibi özellikler her veri seti için eşit alınmıştır. Bu doğrultuda en başarılı eğitim sonucu 87.18 doğruluk değeri ile SCOPE 2.08 veri setine ait olmuştur.

Tablo 7. Veri setlerine ait test sonuçları  
(Test results for data sets)

Veri Seti	Ortalama Kayıp Değeri	Ortalama Doğruluk Değeri	Varyans
SCOPE 2.06	0.5947	83.3690	± 5.9900
SCOPE 2.07	0.5593	83.4462	± 5.6564
SCOPE 2.08	0.4912	86.8238	± 4.1862

Veri setlerine ait test doğruluk değerleri incelendiğinde, veri setinin iyileştirilmesi ile doğruluk değerlerinin iyileşmesi arasında pozitif ilişki olduğu görülmektedir. Çalışmada kullanılan her üç veri seti için de literatürde yer alan çalışmalara kıyasla daha iyi sonuçlar elde edilmiştir.

Bu çalışmada en iyi sonuç hem eğitim hem de test sonuçları için sırasıyla 87.18 ve 86.32 değerleri ile SCOPE 2.08 veri setine ait olmuştur.

Elde edilen benzerlik değerleri sonuçları kullanılarak eğitilen, yalnızca yapay sinir ağları kullanılan çalışmamız, literatürde yer alan birden fazla yöntemin birleştirildiği çalışmalara göre daha yüksek doğruluğa sahip sonuç elde etmiştir. Birden fazla yöntemin birleştirildiği çalışmalarda sistem performansı negatif olarak etkilenebilmektedir. Verilerde katlanma türü sayısının artışı protein katlanması sınıflandırması için ise pozitif yönde katkı sağlamaktadır.

Scop ve Scope veri tabanlarında bulunan veri setleri kullanılarak yapılan çalışmalarda kullanılan yöntem ve bu yöntemlerden elde edilen doğruluk değerleri Tablo 8'de verilmiştir.

Tablo 8. Protein katlanması tanıma için literatürde yer alan çalışmaların karşılaştırması  
(Comparison of studies in the literature for protein folding recognition)

Çalışma	Kullanılan Yöntem	Test Doğruluk Değeri
Levitt [1]	Destek Vektör Makinesi	%63
Suddha [2]	Destek Vektör Makinesi ve Yapay Sinir Ağları birleştirilmiş	%83
Xu [39]	ResNet	%57.7
Liu [15]	Destek Vektör Makinesi	%78
Liu [40]	Evrişimli Sinir Ağı (CNN)	%72.55
Morcillo [41]	Evrişimli Sinir Ağı (CNN), Kapılı Yinelemeli Üniteler (GRU) ve Karar Ağaçları birleştirilmiş	%76.3
Bu çalışma	Yapay Sinir Ağları	%86.32

Literatürde yer alan çalışmalar incelendiğinde protein katlanması tanıma işlemi için Destek Vektör Makinesi yönteminin sıklıkla kullanıldığı ve bu yöntemin yapay sinir ağı yöntemiyle birleştirildiği görülmüştür. SCOPE veri tabanlarındaki veri setlerinin güncellenmesiyle birlikte, proteinlere ait aile ve süper aile bilgisi, katlanma türü sayısının artması gibi etkenler doğruluk değerlerini artırmıştır. Yapay sinir ağlarının diğer yöntemlerle birleştirildiği çalışmalarda doğruluk değerlerinin önemli ölçüde arttığı görülmüştür. Destek vektör makinesi ile yapay sinir ağları yöntemlerini birleştirilen çalışma [2], doğruluk değerleri karşılaştırmasına göre modelimize en yakın değeri elde etmiştir. Ancak iki farklı yöntemin kullanıldığı (destek vektör makinesi ve yapay sinir ağları) çalışmaya kıyasla yalnızca yapay sinir ağı kullanılan yaklaşımımızın performans açısından üstünlük sağladığı söylenebilir.

Çalışmamızda SCOPE veri tabanlarındaki 2.06, 2.07 ve 2.08 veri setlerinde bulunan proteinler ile ilgili özelliklerin tamamı kullanılmamıştır. Bilgi kazanımı hesaplanarak, en yüksek değere sahip DM, SP ve CL özellikleri kullanılmıştır. Bu sayede çalışmamızda kullanılan modelin performansında olumlu etki görülmüştür. Bu çalışmada elde edilen en yüksek test doğruluk değeri SCOPE 2.08 veri seti ile %86.82 olmuştur. Şablon tabanlı yöntemler kullanılarak elde edilen benzerlik değerleri sonuçları kullanılarak eğitilen, yalnızca yapay sinir ağları kullanılan çalışmamız, literatürde yer alan birden fazla yöntemin birleştirildiği çalışmalara göre daha yüksek doğruluğa sahip sonuç elde etmiştir. Birden fazla yöntemin birleştirildiği çalışmalarda sistem performansı negatif olarak etkilenebilmektedir. Verilerde katlanma türü



sayısının artışı protein katlanması sınıflandırması için ise pozitif yönde katkı sağlamaktadır.

## 5. TARTIŞMA (CONCLUSION)

Protein katlanması tanıma işlemi, protein yapılarının tahmini için gereklidir. Bu çalışmada yeni bir yaklaşım önerilmiştir. Literatür çalışmaları incelendiğinde protein katlanması tanıma yöntemi için farklı bilim dallarından faydalanılarak moleküler biyoloji alan için önem taşıyan protein katlanması tespiti ya da protein katlanması tahmini sayesinde protein yapılarının belirlenmesi işleminin gerçekleştirildiği görülmüştür. Özellikle proteinlerin kimyasal yapılarından faydalanılarak yapılan analizler sonucu katlanma türü tahmini ve sınıflandırması için farklı puanlama yöntemleri oluşturulmuş ve sınıflandırma işleminin iyileştirilmesi amacıyla protein veri bankalarında paylaşılmıştır. Bu çalışmada da protein dizilerinden, proteinlerin kimyasal bağ yapılarından elde edilen puanlar, yapay sinir ağı yöntemi ile birleştirilerek performansın iyileştirilmesi sağlanmıştır. Sunulan yeni yaklaşım, üç farklı veri seti açısından kıyaslanmıştır. Protein katlanması tanıma işlemi için literatürde yer alan diğer tahmin edici yöntemlerden, destek vektör makinesi, Bayes yöntemi, K en yakın komşu yöntemi gibi yaygın kullanılan yöntemlerden önemli ölçüde daha iyi performans gösteriyor.

## TEŞEKKÜR (ACKNOWLEDGEMENTS)

Yazarlar, TÜBİTAK ULAKBİM, Yüksek Başarım ve Grid Hesaplama Merkezine (TRUBA), Bursa Teknik Üniversitesi Yüksek Performanslı Hesaplama Laboratuvarına teşekkürlerini sunar.

## KAYNAKLAR (REFERENCES)

- [1] M. Levitt, C. Chothia, "Structural patterns in globular proteins." Nature, 261(5561), 552-558, 1976.
- [2] P. Sudha, D. Ramyachitra, P. Manikandan, "Enhanced artificial neural network for protein fold recognition and structural class prediction", Gene Reports, 12, 261-275, 2018.
- [3] J. S. Butler, S. N. Loh, "Folding and misfolding mechanisms of the p53 DNA binding domain at physiological temperature", Protein science, 15(11), 2457-2465, 2006.
- [4] Y. Kaya, R. Tekin, "Epileptik nöbetlerin tespiti için aşırı öğrenme makinesi tabanlı uzman bir sistem", Bilişim Teknolojileri Dergisi, 5(2), 33-40, 2012.
- [5] A. Haltaş, A. Alkan, "Medline veritabanı üzerinde bulunan tıbbi dokümanların kanser türlerine göre otomatik sınıflandırılması", Bilişim Teknolojileri Dergisi, 9(2), 181, 2016.
- [6] G. Akgül, A.A. Çelik, Z.E. Aydın, Z. K. Öztürk, "Hipotiroidi Hastalığı Teşhisinde Sınıflandırma Algoritmalarının Kullanımı", Bilişim Teknolojileri Dergisi, 13(3), 255-268, 2020.
- [7] A. Şenol, Y. Canbay, M. Kaya, "Makine Öğrenmesi Yaklaşımlarını Kullanarak Salgınları Erken Evrede Tespit Etme Alanındaki Eğilimler", Bilişim Teknolojileri Dergisi, 14(4), 2021.
- [8] M. AlQuraishi, "Machine learning in protein structure prediction", Current opinion in chemical biology, 65, 1-8, 2021.
- [9] C. Ekenna, S. Thomas, N.M. Amato, "Adaptive local learning in sampling based motion planning for protein Folding", BMC systems biology, 10(2), 165-179, 2016.
- [10] J. Zhu, H. Zhang, S.C. Li, C. Wang, L. Kong, S. Sun, D. Bu, "Improving protein fold recognition by extracting fold-specific features from predicted residue-residue contacts", Bioinformatics, 33(23), 3749-3757, 2017.
- [11] K. Yan, J. Wen, J. X. Liu, Y. Xu, B. Liu, "Protein fold recognition by combining support vector machines and pairwise sequence similarity scores", IEEE/ACM Transactions on Computational Biology and Bioinformatics, 18(5), 2008-2016, 2020.
- [12] Y. Yang, E. Faraggi, H. Zhao, Y. Zhou, "Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates", Bioinformatics, 27(15), 2076-2082, 2011.
- [13] S. Makigaki, T. Ishida, "Improvement of template-based protein structure prediction by using chimera alignment", In Proceedings of the 2018 8th International Conference on Bioscience, Biochemistry and Bioinformatics, Tokyo, Japonya, 32-37, Ocak 2018.
- [14] J. Söding, A. Biegert, A.N. Lupas, "The HHpred interactive server for protein homology detection and structure prediction", Nucleic acids research, 33(2), 244-248, 2005.
- [15] B. Liu, Y. Zhu, "ProtDec-LTR3.0: protein remote homology detection by incorporating profile-based features into learning to rank", Ieee Access, 7, 102499-102507, 2019.
- [16] L. Wei, Q. Zou, "Recent progress in machine learning-based methods for protein fold Recognition", International journal of molecular sciences, 17(12), 2118, 2016.
- [17] M. Corrales, P. Cusco, D. R. Usmanova, H.C. Chen, N.S. Bogatyreva, G.J. Filion, D.N. Ivankov, "Machine learning: how much does it tell about protein folding rates?", PloS one, 10(11), 2015.
- [18] F. Noé, G. De Fabritiis, C. Clementi, "Machine learning for protein folding and Dynamics", Current opinion in structural biology, 60, 77-84, 2020.
- [19] Internet: SCOPe: Structural Classification of Proteins — extended, <https://scop.berkeley.edu>, 15.05.2022.
- [20] D. M. Halaby, A. Poupon, J. P. Mornon, "The immunoglobulin fold family: sequence analysis and 3D structure comparisons", Protein engineering, 12(7), 563-571, 1999.
- [21] T. J. Richmond, F. M. Richards, "Packing of  $\alpha$ -helices: Geometrical constraints and contact area", Journal of molecular biology, 119(4), 537-555, 1978.
- [22] T. J. P. Hubbard, T. L. Blundell, "Comparison of solvent-inaccessible cores of homologous proteins: definitions useful for protein modelling. Protein Engineering", Design and Selection, 1(3), 159-171, 1987.
- [23] J. Rozewicki, S. Li, K. M. Amada, D. M. Standley, K. Katoh, "MAFFT-DASH: integrated protein sequence and structural alignment", Nucleic acids research, 47(W1), W5-W10, 2019.

- [24] V. Adar, Protein-ligand etkileşimleri, <http://www.magum.hacettepe.edu.tr/MMKurs/KURS1Proteinligand.pdf>, 24.06.2022.
- [25] Y. N. Imai, Y. Inoue, L. Nakanishi, K. Kitaura, “*Cl- $\pi$  interactions in protein-ligand complexes*”, *Protein Science*, 17(7), 1129-1137, 2008.
- [26] R. Rojas, **Neural Network A Systematic Introduction**, Springer, Heidelberg, Almanya, 1996.
- [27] M. M. Yılmaz, **Periferik sinir defekt onarımında biyolojik kondüit modeli: de-epitelize insan amniyotik membranı ve adipoz kökenli mezankimal kök hücre tabakası içeren sinir kondüit modelinin sinir iyileşmesine etkisinin değerlendirilmesi**, Uzmanlık Tezi, Hacettepe Üniversitesi, Tıp Fakültesi, 2020.
- [28] J. Ma, J. Tang, “*A review for dynamics in neuron and neuronal network*”, *Nonlinear Dynamics*, 89(3), 1569-1578, 2017.
- [29] A. Eliasy, J. Przychodzen, “**The role of AI in capital structure to enhance corporate funding strategies**”. *Array*, 6, 2020.
- [30] I. H. Sarker, “*Deep cybersecurity: a comprehensive overview from neural network and deep learning perspective*”, *SN Computer Science*, 2(3), 1-16, 2021.
- [31] V. Nair, G. E.Hinton, “*Rectified linear units improve restricted boltzmann machines*”, *Icml*, 2010.
- [32] X. Glorot, A. Bordes, Y. Bengio, “*Deep sparse rectifier neural networks*”, **In Proceedings of the fourteenth international conference on artificial intelligence and statistics**, Fort Lauderdale, A.B.D, 315-323, Nisan 2011.
- [33] I. Goodfellow, Y. Bengio, A. Courville, “*Deep learning (adaptive computation and machine learning series)*”, Cambridge Massachusetts, 321-359, 2011.
- [34] M. Bağ, **Derin öğrenme kullanarak IP üzerinden ses hizmeti veren şebekelerde sahtekarlığa yönelik çağrılarının tespiti**, Yüksek Lisans Tezi, Ankara Üniversitesi, Fen Bilimleri Enstitüsü, 2019.
- [35] Y.N.Fu’adah, N.K.C. Pratiwi, M.A. Pramudito, N. İbrahim, “*Convolutional neural network (CNN) for automatic skin cancer classification system*”, **IOP Conf. Ser. Mater. Sci. Eng.**, 982, 12005, 2020.
- [36] G. Korkmaz, E. Eroğlu, “*Model karmaşıklığının kontrolü*”, *İktisadi ve İdari Yaklaşımlar Dergisi*, 2(2), 146-162, 2020.
- [37] B. Ö. Başer, M. Yangın, E.S. SARIDAŞ, “*Makine öğrenmesi teknikleriyle diyabet hastalığının sınıflandırılması*”, *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 25(1), 112-120, 2021.
- [38] Z. Lyu, Y. Yu, B. Samali, M. Rashidi, M. Mohammadi, T.N. Nguyen, A. Nguyen, “*Back-propagation neural network optimized by K-fold cross-validation for prediction of torsional strength of reinforced Concrete beam*”, *Materials*, 15(4), 1477, 2022.
- [39] J. Xu, “*Distance-based protein folding powered by deep learning*”, *Proceedings of the National Academy of Sciences*, 116(34), 16856-16865, 2019.
- [40] C. Li, B. Liu, “*MotifCNN-fold: protein fold recognition based on fold-specific features extracted by motif-based convolutional neural networks*”, *Briefings in Bioinformatics*, 21(6), 2133–2141, 2020.
- [41] A. Villegas-Morcillo, V. Sanchez, A.M. Gomez, “*FoldHSphere: deep hyperspherical embeddings for protein fold Recognition*”, *BMC bioinformatics*, 22(1), 1-21, 2021.

# Kentsel Trafik Tahminine Yönelik Derin Öğrenme Tabanlı Verimli Bir Hibrit Model

*Araştırma Makalesi/Research Article*

 Anıl UTKU

Computer Engineering Department, Munzur University, Tunceli, Turkey

[anilutku@munzur.edu.tr](mailto:anilutku@munzur.edu.tr)

(Geliş/Received:26.08.2022; Kabul/Accepted:16.03.2023)

DOI: 10.17671/gazibtd.1167140

**Özet**— Trafik yoğunluğu problemi, kentsel hayatın en önemli sorunlarından biri haline gelmiştir. Trafik yoğunluğu sebebiyle harcanan zaman ve yakıt, araç kullanıcıları ve ülkeler için önemli bir kayıptır. Trafikte geçen zamanı azaltmak amacıyla geliştirilen uygulamalar, uzun vadeli trafik yoğunluğu hakkında başarılı tahminlerde bulunamamaktadır. Kameralar, sensörler ve mobil cihazlar üzerinden elde edilen trafik verileri, trafik yönetimi sorununu çözebilmek amacıyla yapay zekâ teknolojilerinin kullanımını ön plana çıkarmaktadır. Bu çalışmada, trafik yoğunluk tahminine yönelik Convolutional Neural Network (CNN) ve Recurrent Neural Network (RNN) modelleri kullanılarak hibrit bir tahmin modeli geliştirilmiştir. Çalışmada, CNN ve RNN'in öne çıkan özelliklerinden faydalanmak amaçlanmıştır. CNN, özellik çıkarma aşamasında, RNN ise sıralı zaman serisi verileri üzerinde öğrenme ve tahmin için etkili bir modeldir. Bu yöntemler hibrit bir şekilde kullanılarak tahmin doğruluğunun artırılması amaçlanmıştır. İstanbul Büyükşehir Belediyesi tarafından sunulan saatlik trafik yoğunluğu veri seti kullanılmıştır. Kullanılan veri seti 2321 farklı nokta için 2020 Ocak ile 2020 Aralık tarihleri arasındaki trafik yoğunluk bilgisini içermektedir. Geçen araç sayısı, Bağcılar Avrupa Otoyolu kavşağında daha yüksek olduğu için bu konum deneysel çalışmalarda kullanılmıştır. Seçilen konum için 9379 satır araç bilgisi bulunmaktadır. Geliştirilen hibrit model Linear Regression (LR), Random Forest (RF), Support Vector Machine (SVM), Multilayer Perceptron (MLP), CNN, RNN ve Long-Short Term Memory (LSTM) ile İstanbul'un 2020 yılına ait trafik verileri kullanılarak test edilmiştir. Deneysel sonuçlar, önerilen hibrit modelin karşılaştırılan modellere göre daha başarılı sonuçlara sahip olduğunu göstermiştir. Önerilen model kavşaktan geçen araç sayısı tahmininde 0,929 R2 değerine, kavşaktan geçen araçların ortalama hızlarının tahmininde ise 0,934 R-Squared (R2) değerine sahip olmuştur.

**Anahtar Kelimeler** — trafik tahmini, makine öğrenmesi, derin öğrenme, CNN, RNN

## Deep Learning Based an Efficient Hybrid Model for Urban Traffic Prediction

**Abstract**— The traffic density problem has become one of the most important problems of urban life. The time and fuel spent due to traffic density is a significant loss for vehicle users and countries. However, applications developed to reduce the time spent in traffic cannot make successful predictions about long-term traffic density. Traffic data obtained from cameras, sensors and mobile devices highlight the use of artificial intelligence technologies to solve the traffic management problem. In this study, a hybrid prediction model was developed using Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) models for traffic density prediction. It was aimed to benefit from the prominent features of CNN and RNN. CNN is an effective model in the feature extraction phase, while RNN is an effective model in the learning and prediction phase on sequential time series data. It was aimed to increase the prediction accuracy by using these methods in a hybrid way. The hourly traffic density dataset provided by the Istanbul Metropolitan Municipality was used. The dataset used includes traffic density information for 2321 different points between January 2020 and December 2020. Since the number of passing vehicles is higher at the Bağcılar European Motorway junction, this location was used in the experimental studies. There are 9379 lines of vehicle information for the selected location. The developed hybrid model was tested using Linear Regression (LR), Random Forest (RF), Support Vector Machine (SVM), Multilayer Perceptron (MLP), CNN, RNN and Long-Short Term Memory (LSTM) with 2020 Istanbul's traffic data. Experimental results showed that the proposed hybrid model was more successful results than the compared models. The proposed model has 0.929 R-Squared (R2) in the predicting the number of vehicles passing through the junction, and 0.934 R2 in the predicting the average speed of the vehicles passing through the junction.

**Keywords**— traffic prediction, machine learning, deep learning, CNN, RNN

## 1. INTRODUCTION

Today, with the increase in urbanization, the growth in the urban population causes an increase in traffic density [1]. Traffic density not only causes loss of time, but also causes loss of labour, income, health and social life, and results in environmental and noise pollution [2]. Therefore, meeting the need for mobility, especially in big cities, eliminating traffic-related problems, or keeping them under control is only possible with the development of intelligent systems.

The traffic density, which has become the biggest problem of Istanbul, is increasing daily. Especially due to the pandemic, people's use of their own vehicles instead of public transportation has increased traffic density [3]. As a result, there is a severe density at almost every point of Istanbul traffic at the beginning and end of working hours.

Traffic congestion in Istanbul increases considerably on weekdays, especially during commuting and commuting hours [4]. City governments use sensors and traffic cameras at critical junctions to collect mobility data. However, due to high deployment and maintenance costs, the applicability of these technologies across the city is limited. With the development of intelligent transportation systems, vast amounts of mobility data can be collected from mobile devices [5]. These obtained data make large-scale and reliable traffic predictions applicable. Increasing traffic data provides potential new perspectives for solving the traffic management problem. Traffic prediction has become an important research area in recent years. In the literature, studies use artificial intelligence methods for traffic density prediction.

Du et al. developed a hybrid model using CNN and LSTM for short-term traffic forecasting [6]. With the developed model, it was aimed to learn the temporal and spatial properties. By using LSTM, temporal dependencies and local trends were extracted. Experimental results showed that the developed model was more successful than classical deep learning models.

Mohammed and Kianfar presented a comparative analysis of Deep Neural Networks (DNN), RF, Gradient Boosting and Generalized Linear Model (GLM) for traffic density and traffic rate prediction [7]. In the study, traffic data from Interstate 64 in St. Louis Missouri is used. Experimental results have shown that RF gives better results than the compared models.

Zhang et al. developed a CNN-based short-term traffic prediction model [8]. Selected features are forwarded to CNN for prediction. The developed model is compared with SVM, Seasonal Autoregressive Integrated Moving Average (SARIMA), k-Nearest Neighbours (kNN), Artificial Neural Network (ANN) and CNN. Experimental results showed that the developed model was more successful than other models.

Gu et al. proposed the Fusion Deep Learning (FDL) model to predict lane-based traffic speed [9]. A dual-layer deep learning model was created using LSTM and Gated Recurrent Unit (GRU). The developed model was tested using Beijing traffic data with ARIMA, Lighthill-Whitham-Richards (LWR), Multi-Layer Perceptron (MLP), Kalman Filter (KF), Radial Basis Function Neural Network (RBFNN), CNN and LSTM. Experimental results showed that dual-layer deep learning model was more successful than the compared models.

Wang et al. proposed an LSTM-based model for traffic speed estimation [10]. In the study, it is aimed to determine the traffic flow mechanism by dividing the road network into critical points. Dividing roads are modelled using LSTM. Spatial-temporal features are used in traffic forecasting. Experimental studies using ANN, kNN and CNN showed that the proposed model is more successful.

Taş and Müngen developed an ANN based model for the prediction of regional traffic density [11]. In the study, it is aimed to increase the success of the forecast model by using environmental factors such as weather conditions. Experimental results showed that the developed model has an R-Squared ( $R^2$ ) value of 0.88.

Takak et al. visualized traffic density and predicted traffic speed in the short, medium and long term [12]. Traffic speed data obtained from sensors at many points in Istanbul are visualized according to time and location. Experimental studies using Autoregressive Integrated Moving Average (ARIMA) and regression models showed that the regression model predicts short-term traffic flow with 23.99% Mean Absolute Percentage Error (MAPE) and 8.5 MAE.

Essien et al. propose a deep learning-based traffic forecasting model that combines information extracted from tweets with traffic and weather information [13]. The developed LSTM based model was tested using the Manchester traffic dataset. The developed model is compared with SVM, eXtreme Gradient Boosting (xGBoost) and RF. Experimental results showed that the developed model was more successful than the compared models.

Wang et al. developed a hybrid model based on CNN-LSTM for short-term traffic flow prediction [14]. CNN was used to extract temporal and spatial features from the data. It was aimed to learn the long-term dependencies by presenting the features extracted by CNN as input to LSTM. Experimental results showed that the developed model was more successful than the classical neural network models.

Zheng et al. developed a hybrid deep learning model using CNN and LSTM modules to extract temporal and spatial properties of traffic flow data [15]. An attentional mechanism was developed by assigning different weights to recipe flow data at different times. It was aimed to

extract the daily and weekly characteristics of traffic flow data using the Bi-LSTM model. Experimental results showed that the developed model was more successful than the other models compared.

The studies in the literature generally use iterative neural network models for time-dependent data such as traffic flow data. Iterative neural network models are helpful because they can extract time dependencies from time series data. In the studies examined, it was seen that hybrid models were developed using CNN and LSTM models. In this study, a hybrid model was developed using CNN and RNN models, similar to the studies in the literature. The developed model consists of a 1D CNN layer, similar to the models used in the studies in the literature. It was aimed to extract temporal features from the data using 1D CNN. RNN was used to learn the features extracted by CNN and to increase the prediction accuracy.

The main contributions of this study to the literature can be summarized as follows:

- This is the first study using this dataset for traffic density prediction.
- Deep learning-based hybrid prediction model was proposed using CNN and RNN models.
- The proposed model was comparatively analysed with popular machine learning and deep learning models such as LR, RF, SVM, MLP, CNN, RNN and LSTM.

## 2. MATERIAL AND METHOD

In this study, a hybrid deep learning model was proposed for Istanbul by using CNN and RNN models to predict average traffic speed and vehicle density. The proposed model aims to predict the average speed and traffic density at the Bağcılar European Motorway junction, one of Istanbul's critical junction points. The proposed hybrid prediction model was compared with LR, RF, SVM, MLP, CNN, RNN and LSTM.

### 2.1. Dataset

This study used the hourly traffic density dataset between January 2020 and December 2020 presented by Istanbul Metropolitan Municipality [16]. The dataset consists of date time, longitude, latitude, geohash, minimum speed, maximum speed, average speed and number of vehicle attributes. Date time represents time in Year/Month/Day and time format. Longitude represents the longitude value and Latitude represents the latitude value. Geohash refers to the geolocation code obtained according to the latitude and longitude values. Minimum, maximum, and average speed attributes represent the minimum, maximum, and average speed values, respectively. Finally, the number of vehicles represents the hourly number of vehicles passing a specific location. The dataset used is shown in Table 1.

Table 1. Hourly traffic density dataset

Date time	Longitude	Latitude	Geohash	Minimum speed	Maximum speed	Average speed	Number of vehicle
2020-11-30 23:00:00	29.2950439453125	41.1026000976562	sxkcf	41	137	80	130
2020-11-30 23:00:00	29.0863037109375	41.0092163085938	sxk9mc	6	122	49	77
2020-11-30 23:00:00	28.0975341796875	41.1904907226562	sx7fr6	51	129	86	65
2020-11-30 23:00:00	29.0753173828125	41.0256958007812	sxk9ms	9	89	56	29
2020-11-30 23:00:00	29.1412353515625	40.9927368164062	sxk9pn	14	124	75	36

The dataset contains traffic density information for 2321 different points. The locations with the highest number of vehicles were determined to select where the traffic density and speed prediction would be made. As seen in Figure 1, Bağcılar European Motorway junction, one of the points with the highest number of vehicle passes, was chosen for experimental studies.

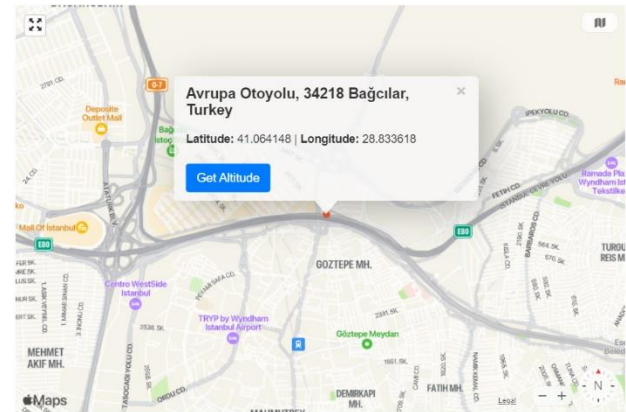


Figure 1. Bağcılar European Motorway junction

Figure 2 shows the number of vehicles passing through the Bağcılar European Motorway junction per hour.

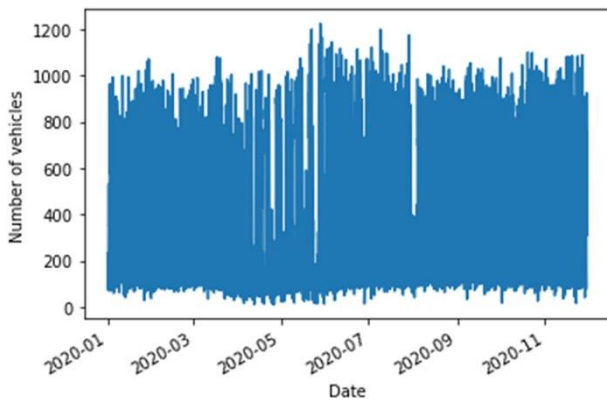


Figure 2. Number of vehicles per hour

Figure 3 shows the average speed of vehicles passing through the Bağcılar European Motorway junction per hour.

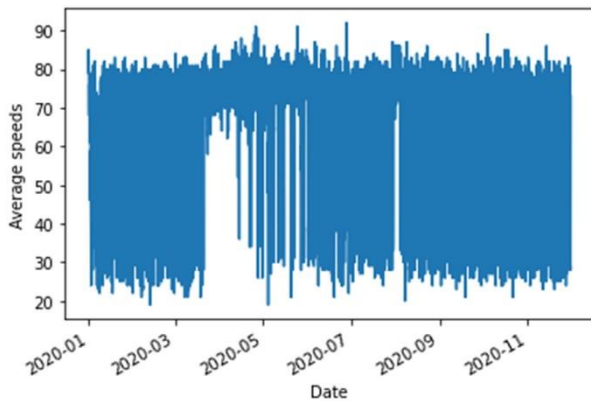


Figure 3. Average speeds of passing vehicles per hour

## 2.2. Baseline Models

This section explains LR, RF, SVM, MLP, CNN, RNN and LSTM models used in the study.

LR is a method used for linear and continuous variables [17]. When a linear relationship is observed between the variables, LR is used to predict the future and examine how the variables affect each other [18]. LR creates a straight line that minimizes discrepancies between predicted and actual values [19]. Then, the value of the dependent variable is predicted using the independent variable.

RF is used to make accurate predictions by producing more compatible models using multiple decision trees [20]. RF trains decision trees on different datasets and combines the prediction results. Then, it votes on the prediction results and uses the prediction with the highest votes [21].

SVM aims to find the hyperplane in the feature space that can optimally separate the two classes from each other. It is a supervised learning and classification method. The distance between the positive and negative samples closest to the hyperplane is called the margin [22]. SVM tries to

find support vectors that make this distance furthest. Finding the maximum margin for linearly separable data is easier. However, while classifying the nonlinearly separated data, the data are transferred to a different space where they can be linearly separated and classified in this new space [23].

MLP has a structure in which many neurons with non-linear activation functions are hierarchically connected. MLP consists of one or more layers. The input layer receives the incoming data and sends it to the middle layer. Incoming information is transferred to the next layer. The number of intermediate layers varies according to the problem, at least one, and is adjusted according to the need. The output of each layer becomes the input of the next layer. Thus, the output is reached. Neurons are connected to all neurons in the next layer. The output layer processes data from previous layers and determines the network's output. The output number of the system is equal to the number of elements in the output layer [24].

CNN comprises multiple layers and connects the input data to the output data. It generally performs operations such as scanning objects, clustering objects, and finding similar ones over a picture or a video. In order to activate this network, various parameters and libraries that contribute to the formation of neural networks are used [25]. After the input data is created, filtering is carried out by various layers. Filters are expressed with matrix calculation, and a single matrix is obtained while generating the output data. Data are analysed by creating 5 different layers. A convolutional layer is created to detect the features on the image. Non-linearity is introduced to the system with the Non-Linearity Layer. Thanks to the Pooling Layer, the weight in the image is tried to be reduced, and the compatibility of the image with the parameters is tested in line with this reduced weight. Flattening layer is the layer where the data required to create the output data is prepared. Finally, Fully-Connected Layer is used for classification.

RNN is a class of neural networks in which node connections form a directed loop [26]. RNN has an input layer, hidden layers, and an output layer. All of these layers work independently [27]. Structures in each layer have weights and layer-specific thresholds. As a result of these recurrent steps, the previous input state is stored and combined with the newly obtained input value so that the relationship of the newly obtained input with the previous input is provided [28].

LSTM is a model that can learn long-term transactions. LSTM is an advanced variant of RNN used for modeling sequential data [29]. LSTM consists of forget it, input and output layers. The forget layer decides whether the incoming information will be forgotten or not. The input layer decides which information is stored in memory or not. Finally, the output layer decides whether or not information will be output [30].

### 2.3. Data Pre-processing

In this study, data pre-processing was performed on the dataset before the models were applied. Training, testing and validation data were selected. Using the GridSearchCV library, the parameters of the models were optimized on the validation data. Models were created by selecting the parameters with the lowest MSE values. Then the number of vehicles and the average speed of the vehicles were predicted. MSE, RMSE and R<sup>2</sup> values were calculated according to the estimation results obtained. The flowchart of the developed system is presented in Figure 3.

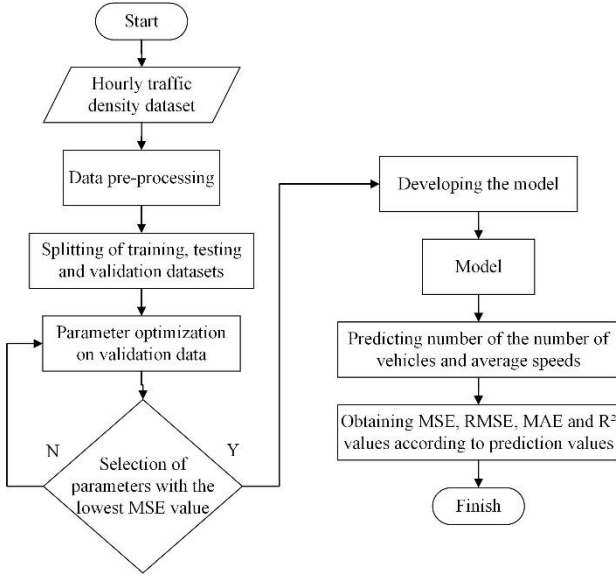


Figure 4. The flowchart of the developed system

The data used in this study is a time series dataset consisting of hourly indexed data points. In order to apply machine learning methods to time series data, these data should be structured as supervised learning problems using the sliding window method. In the sliding window method, historical observation data equal to the specified window size is placed as input into the sliding window. The value in the next time step to be predicted will be the output of the sliding window. In this study, the sliding window size was chosen as 3 according to the experimental studies. As seen in Figure 5, time series data is structured as a supervised learning problem, with the data in time steps  $t_1$ ,  $t_2$ , and  $t_3$  as the input of the sliding window, and the data in time step  $t_4$  as the output of the sliding window.

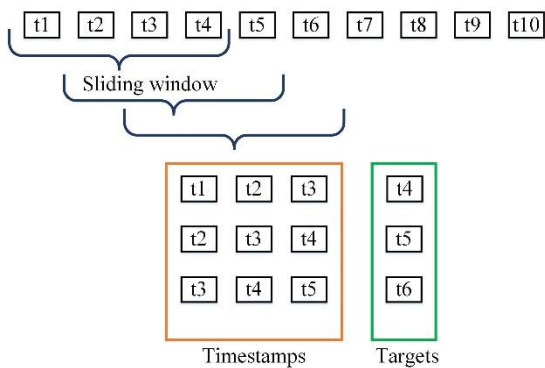


Figure 5. Sliding window method

After structuring the data as a supervised learning problem, the data were normalized in the 0-1 using MinMaxScaler. After the normalization, the data was split into training, test, and validation sets. These values were chosen for experimental studies, as the highest prediction accuracy was achieved in the combination of 80% training and 20% testing. In addition, 10% of the training data is split for validation. The validation data were used to optimize the parameters of the applied models. Model parameters were optimized using GridSearchCV from the Scikit Learn library so that the applied models could obtain the best prediction results.

### 2.4. Performance Evaluation Metrics

Mean Squared Error (MSE), Root Mean Squared Error (RMSE), MAE and R<sup>2</sup> metrics are mainly used to determine the error rate between the predictions and the actual values. MSE calculates the mean of the squares of the difference between actual and predicted values. MSE is calculated using Eq.1.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y - \hat{y})^2 \quad (1)$$

Here,  $y$  is the actual observation values,  $\hat{y}$  is the predicted values, and  $n$  is the total data.

RMSE calculates the square root of the mean of the squares of the difference between the actual observation values and the predicted values. RMSE is calculated as seen in Eq.2.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y - \hat{y})^2} \quad (2)$$

MAE refers to the mean of the absolute differences between the actual observation values and the predicted values. MAE is calculated using Eq.3.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y - \hat{y}| \quad (3)$$

R<sup>2</sup> indicates how well the dataset fits the regression line. It describes the goodness of fit of the data point on the regression line. It is the square value of the correlation coefficient. R<sup>2</sup> is calculated using Eq.4.

$$R^2 = \frac{\sum (y - \hat{y})^2}{(y - \bar{y})^2} \quad (4)$$

Here,  $\hat{y}$  is the predicted  $y$  values and  $\bar{y}$  is the mean of the  $y$  values.

### 2.5. Proposed Hybrid Prediction Model

In this study, a hybrid deep learning model was developed using CNN and RNN models. The architecture of the proposed model is shown in Figure 6.

CNN is an efficient model for automatically extracting features and learning from one-dimensional series data such as univariate time series. In this study, a hybrid model

was proposed using CNN to interpret sub-sequences that are input to RNN. CNN extracts features from the input data and transforms the univariate input data into multi-dimensional groups using convolution. The multi-dimensional datasets are then transmitted to the RNN for prediction.

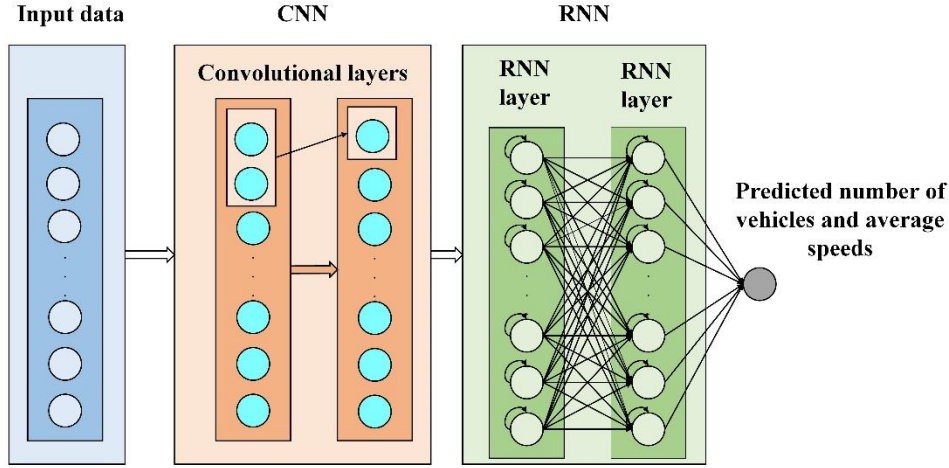


Figure 6. The architecture of the proposed model

CNN is an efficient model for automatically extracting features and learning from one-dimensional series data such as univariate time series. This study proposes a hybrid model using CNN to interpret sub-sequences that are input to RNN. CNN extracts features from the input data and transforms the univariate input data into multi-dimensional groups using convolution.

RNN contains feedback loops and encodes contextual information of a temporal sequence. For an input sequence given as  $\{F_1, F_2, \dots, F_T\}$ , the hidden states  $h_t$  and outputs  $y_t$  can be calculated as follows:

$$h_t = H(W_{ih}F_t + W_{hh}h_{t-1}) + b_h \quad (1)$$

$$y_t = W_{ho}h_t + b_o \quad (2)$$

Where,  $W_{ih}$ ,  $W_{hh}$ ,  $W_{ho}$  are weight matrices between input, hidden and output layers. Basically, for a  $x_t$  input received at time  $t$ , long-term memory  $C_{t-1}$  and transaction memory  $h_{t-1}$  are updated from the previous time step to time  $t$ . The developed model uses following equations for learning and prediction.

$$i_t = \sigma(w_{xi}x_t + w_{hi}h_{t-1} + w_{ci}c_{t-1} + b_i) \quad (3)$$

$$f_t = \sigma(w_{xf}x_t + w_{hf}h_{t-1} + w_{cf}c_{t-1} + b_f) \quad (4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(w_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (5)$$

$$o_t = \sigma(w_{xo}x_t + w_{ho}h_{t-1} + w_{co}c_{t-1} + b_o) \quad (6)$$

$$h_t = o_t \odot h(c_t) \quad (7)$$

where,  $i_t$  represents input gate,  $o_t$  represents output gate,  $f_t$  represents forgetting gate,  $c$  represents the cell activation vector,  $w$  weight matrix and  $b$  represents the bias vector.  $\odot$  represents the scalar product of two vectors.  $x_t$  and  $h_t$  represent input-output sequences.  $g_t$  refers to the input string.

The CNN-RNN architecture was developed to support the prediction model. The developed model contains CNN layers combined with RNN to extract properties from input data. First, using convolution, CNN extracts essential information from the input data and converts the univariate input data into multi-dimensional groups. Then, after the input data is sent to RNN, it is passed to the prediction stage.

In the developed model, CNN is used for feature extraction, while RNN is used to analyze and predict features extracted by CNN. The first step is to divide the input sequences into sub-sequences that can be processed by the CNN so that the developed hybrid model can be used in the weather prediction problem. Univariate time series data can be divided into 3 inputs and one output as input/output samples.

CNN interprets these sub-array samples and sends them to RNN for processing as input. Here, CNN has a one-dimensional convolutional layer with kernel size 1 and filter number 64 for reading sub-arrays. The number of filters refers to the number of times the input string is read.



Following the convolutional layer, a max pooling layer that is used to interpret the input feature and the dense layer that interprets the properties extracted by the convolution layer of the model are defined. The convolution and pooling layers are 3D layers. For this reason, flatten layer is used to reduce the feature maps to a one-dimensional vector to be used as an input to RNN.

Parameter analysis studies aim to reach the highest prediction accuracy with parameters such as the number of layers, neurons, epochs, and batch size. Adam was used as the optimizer. The GridSearchCV library was used to optimize the parameters of the developed model. The parameters of CNN and RNN were determined using GridSearchCV, and thus the model was created. The parameters of RF are max depth: 9, max features: log2,

max leaf nodes: 9, number of estimators: 25, respectively. SVM's parameters are C: 10, gamma: 1e-07, epsilon: 0.1 and kernel: linear, respectively.

### 3. EXPERIMENTAL RESULTS

In this study, the proposed hybrid model was extensively compared with LR, RF, SVM, MLP, CNN, RNN and LSTM. The results obtained according to MSE, RMSE, MAE and R<sup>2</sup> metrics were analysed comparatively. Table 2 and Figure 7 show comparative experimental results for predicting the number of vehicles passing through the junction.

Table 2. Experimental results for predicting the number of vehicles

Models	MSE	RMSE	MAE	R <sup>2</sup>
LR	18768.258	136.997	85.756	0.772
RF	12222.608	110.556	76.289	0.851
SVM	17585.555	132.610	86.439	0.786
MLP	11067.972	105.204	72.410	0.865
CNN	11555.746	107.498	72.582	0.859
RNN	10544.777	102.687	68.449	0.872
LSTM	9830.563	99.149	68.167	0.880
Proposed model	5302.627	72.819	54.036	0.929

Experimental results for the predicting the number of vehicles passing through the junction showed that the proposed model was more successful than the compared models.

After the proposed model, LSTM, RNN MLP, CNN, RF, SVM and LR were successful, respectively. Table 3 and Figure 8 show comparative experimental results for predicting the average speed of vehicles passing through the junction.

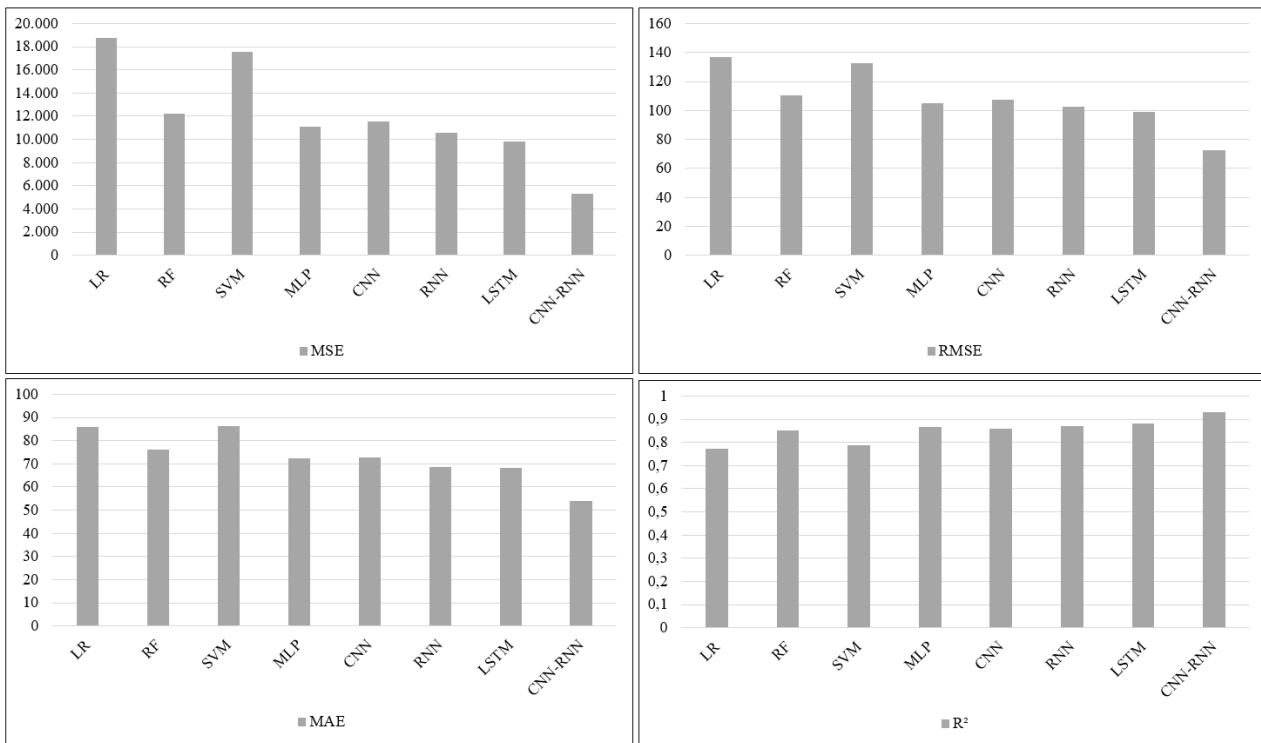


Figure 7. Experimental results for predicting the number of vehicles

Table 3. Experimental results for predicting average speeds of vehicles

Models	MSE	RMSE	MAE	R <sup>2</sup>
LR	71.702	8.467	5.334	0.805
RF	68.623	8.283	5.461	0.814
SVM	69.239	8.321	5.389	0.812
MLP	64.974	8.061	5.241	0.824
CNN	65.773	8.110	5.592	0.821
RNN	64.197	8.012	5.306	0.826
LSTM	62.038	7.876	5.319	0.832
Proposed model	30.186	5.494	4.180	0.934

Experimental results for the predicting the number of vehicles passing through the junction showed that the proposed model was more successful than the compared

models. After the proposed model, LSTM, RNN MLP, CNN, RF, SVM and LR were successful, respectively.

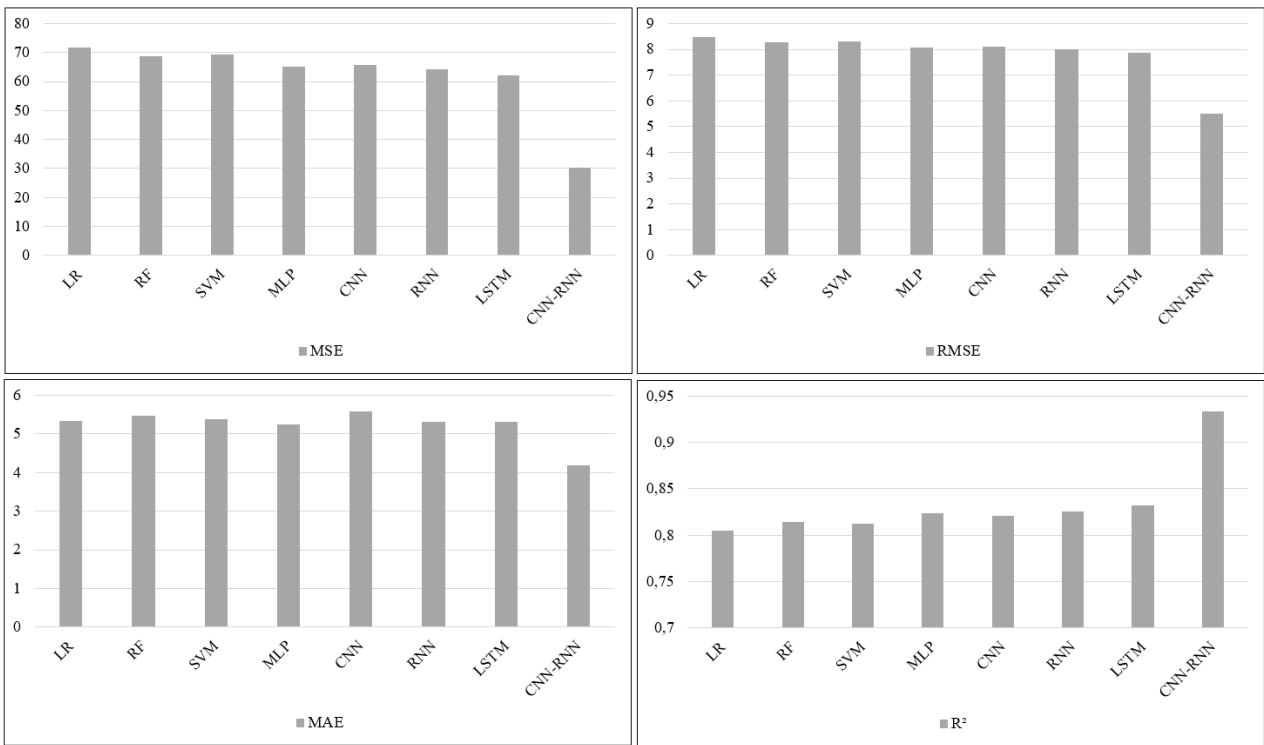


Figure 8. Experimental results for average speed prediction

Figure 9.a shows the graph of the predictions for the number of vehicles passing through the junction, and Figure 9.b shows the graph of the predictions of the average speeds of the vehicles passing the junction. As

seen in Figure 9, the proposed model successfully predicted the number of vehicles and average speed value fluctuations.

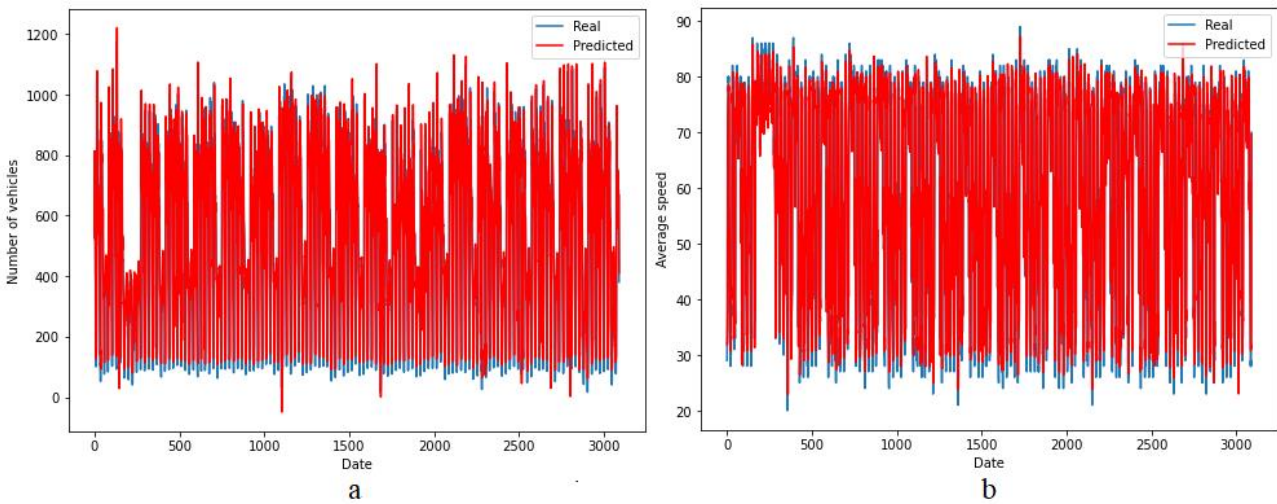


Figure 9. Prediction graphs of the proposed model

#### 4. CONCLUSIONS

Traffic congestion is a part of the daily life of those living in big cities such as the megacity of Istanbul. Although the metro and metrobus lines reduce this problem a little, the fact that these vehicles are located only between specific routes causes many people to be exposed to traffic jams. Therefore, traffic congestion has economic effects and is a waste of time. At the same time, it causes a decrease in work efficiency due to the loss of time caused by traffic

congestion, and the direct and indirect costs caused by traffic congestion cause an increase in the expenses of the enterprises. As a result, all these cause an increase in the price of the product or service offered.

In this study, a hybrid deep learning model was developed to predict the number and average speed of vehicles passing through the Bağcılar European Motorway. Bağcılar European Motorway is one of the most critical junction points of Istanbul. The developed model was

compared with LR, RF, SVM, MLP, CNN, RNN, and LSTM. For each model, the results obtained according to MSE, RMSE, MAE, and  $R^2$  metrics were analyzed comparatively. Experimental results showed that the proposed model has higher prediction accuracy than the compared models.

Experimental results showed that the developed hybrid model was more successful than the compared models. The developed model's success is that it can process very long input data that can be read in blocks by CNN and put together by RNN.

The success of LSTM compared to other models can be explained by the fact that LSTM has special units in addition to the standard units found in RNN. LSTM units contain a memory cell that can hold information in memory for long periods. In addition, a set of gates controls when information enters memory, when it leaves, and when it is forgotten. This structure allows the learning of longer-term dependencies.

The fact that RNN is more successful than CNN can be explained by the fact that CNN and RNN have different architectures. CNN is a feed-forward neural network using filters and pooling layers. RNN feeds the results back to the network. In CNN, the size of the input and the output are fixed.

The fact that MLP is more successful than CNN can be explained by the fact that MLP takes vectors as input and CNN takes tensors as input. The fact that RF is more successful than SVM can be explained by the fact that RF works with a mixture of numerical and categorical features. This allows RF to use the data as they are. SVM, on the other hand, maximizes the margin between different points and calculates the distance between points.

Experimental results for predicting the number of vehicles passing through the junction showed that the proposed hybrid model has 5302.627 MSE, 72.819 RMSE, 54.036 MAE, and 0.929  $R^2$ . Experimental results for predicting the average speed of vehicles passing through the junction showed that the proposed hybrid model has 30.186 MSE, 5.494 RMSE, 4.180 MAE, and 0.934  $R^2$  values.

The results will reveal that the importance of controlling vehicle traffic is increasing, especially in metropolitan cities such as Istanbul, with a large population. Therefore, determining the current traffic congestion level should be taken under control together with the main reasons causing the traffic congestion, and possible measures to prevent this congestion should be implemented. For this reason, various traffic data obtained instantly through sensors, cameras, and mobile devices are analyzed using artificial intelligence methods. For this purpose, the results obtained in this study can be adapted to real-world applications, and effective planning can be done for traffic management.

## REFERENCES

- [1] W. Broere, "Urban underground space: Solving the problems of today's cities". Tunnelling and Underground Space Technology, 55, 245-248, 2016.
- [2] G. Firdaus, & A. Ahmad, "Noise pollution and human health: a case study of municipal corporation of Delhi", Indoor and built environment, 19(6), 648-656, 2010.
- [3] D. Muley, M. Shahin, C. Dias, & M. Abdullah, "Role of transport during outbreak of infectious diseases: evidence from the past", Sustainability, 12(18), 7367, 2020.
- [4] Ş. İmre, & D. Çelebi, "Measuring comfort in public transport: a case study for İstanbul", Transportation Research Procedia, 25, 2441-2449, 2017.
- [5] J. Zhang, F.Y. Wang, K. Wang, W.H. Lin, X. Xu, & C. Chen, "Data-driven intelligent transportation systems: A survey", IEEE Transactions on Intelligent Transportation Systems, 12(4), 1624-1639, 2011.
- [6] S. Du, T. Li, X. Gong, Y. Yang, & S.J. Horng, "Traffic flow forecasting based on hybrid deep learning framework", In 2017 12th international conference on intelligent systems and knowledge engineering (ISKE), 1-6, 2017.
- [7] O. Mohammed, & J. Kianfar, "A machine learning approach to short-term traffic flow prediction: A case study of interstate 64 in Missouri," In 2018 IEEE International Smart Cities Conference (ISC2), Kansas City, Missouri, USA, 1-7, 2018.
- [8] W. Zhang, Y. Yu, Y. Qi, F. Shu, & Y. Wang, "Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning", Transportmetrica A: Transport Science, 15(2), 1688-1711, 2019.
- [9] Y. Gu, W. Lu, L. Qin, M. Li, & Z. Shao, "Short-term prediction of lane-level traffic speeds: A fusion deep learning model", Transportation research part C: emerging technologies, 106, 1-16, 2019.
- [10] J. Wang, R. Chen, & Z. He, "Traffic speed prediction for urban transportation network: A path based deep learning approach", Transportation Research Part C: Emerging Technologies, 100, 372-385, 2019.
- [11] İ.C. Taş, & A.A. Müngen, "Regional Traffic Density Estimation with Artificial Neural Networks and Support Vector Machines Methods", Adiyaman University Journal of Engineering Sciences, 8(15), 378-390, 2021.
- [12] İ. Takak, H. Görümez, H.İ. Türkmen, & M.A. Güvensan, "Short, Medium and Long Term Traffic Flow Rate Estimation and Visualization Tool", International Journal of Advances in Engineering and Pure Sciences, 33(4), 568-580, 2021.
- [13] A. Essien, I. Petrounias, P. Sampaio, & S. Sampaio, S. "A deep-learning model for urban traffic flow prediction with traffic events mined from twitter", World Wide Web, 24(4), 1345-1368, 2021.
- [14] K. Wang, C. Ma, Y. Qiao, X. Lu, W. Hao, & S. Dong, "A hybrid deep learning model with 1DCNN-LSTM-Attention networks for short-term traffic flow prediction", Physica A: Statistical Mechanics and its Applications, 583, 126293, 2021.

- [15] G. Zheng, W.K. Chai, J.L. Duanmu, & V. Katos, “**Hybrid deep learning models for traffic prediction in large-scale road networks**”, *Information Fusion*, 92, 93-114, 2023.
- [16] Internet: İstanbul Büyükşehir Belediyesi, Saatlik trafik yoğunluk veriseti, <https://data.ibb.gov.tr/dataset/hourly-traffic-density-dataset>
- [17] G.Y. Oukawa, P. Krecl, & A.C. Targino, “*Fine-scale modeling of the urban heat island: A comparison of multiple linear regression and random forest approaches*”, *Science of the total environment*, 815, 152836, 2022.
- [18] F. Rustam, A.A. Reshi, A. Mehmood, S. Ullah, B.W. On, W.Asalam, & G.S. Choi, “*COVID-19 future forecasting using supervised machine learning models*”, *IEEE access*, 8, 101489-101499, 2020.
- [19] M. Kayakuş, & M. Terzioğlu, “*Yapay sinir ağları ve çoklu doğrusal regresyon kullanarak emeklilik fonu net varlık değerlerinin tahmin edilmesi*”, *Bilişim Teknolojileri Dergisi*, 14(1), 95-103, 2021.
- [20] A.L. Balogun, & A. Tella, “*Modelling and investigating the impacts of climatic variables on ozone concentration in Malaysia using correlation analysis with random forest, decision tree regression, linear regression, and support vector regression*”, *Chemosphere*, 299, 134250, 2022.
- [21] Y. Yang, & W. Chen, “*Taiga: performance optimization of the C4.5 decision tree construction algorithm*”, *Tsinghua Science and Technology*, 21(4), 415-425, 2016.
- [22] M. Peker, O. Özkaraca, & B. Kesimal, “*Enerji tasarruflu bina tasarımı için ısıtma ve soğutma yüklerini regresyon tabanlı makine öğrenmesi algoritmaları ile modelleme*”, *Bilişim Teknolojileri Dergisi*, 10(4), 443-449, 2017.
- [23] V.A. Kumari, & R. Chitra, R. “**Classification of diabetes disease using support vector machine**”, *International Journal of Engineering Research and Applications*, 3(2), 1797-1801, 2013.
- [24] A.A. Heidari, H. Faris, S. Mirjalili, I. Aljarah, & M. Mafarja, “*Ant lion optimizer: theory, literature review, and application in multi-layer perceptron neural networks*”, *Nature-inspired optimizers*, 23-46, 2020.
- [25] I. Namatēvs, “*Deep convolutional neural networks: Structure, feature extraction and training*”, *Information Technology and Management Science*, 20(1), 40-47, 2017.
- [26] Z. Qiao, N. Sun, X. Li, E. Xia, S. Zhao, & Y. Qin, “*Using machine learning approaches for emergency room visit prediction based on electronic health record data*”, In *Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth*, IOS Press, 111-115, 2018.
- [27] A. Shewalkar, “*Performance evaluation of deep neural networks applied to speech recognition: RNN, LSTM and GRU*”, *Journal of Artificial Intelligence and Soft Computing Research*, 9(4), 235-245, 2019.
- [28] Y. Tian, R. Lai, X. Li, L. Xiang, & J. Tian, “*A combined method for state-of-charge estimation for lithium-ion batteries using a long short-term memory network and an adaptive cubature Kalman filter*”, *Applied Energy*, 265, 114789, 2020.
- [29] A. Farzad, H. Mashayekhi, & H. Hassanpour, “*A comparative performance analysis of different activation functions in LSTM networks for classification*”, *Neural Computing and Applications*, 31(7), 2507-2521, 2019.
- [30] X.H. Le, H.V. Ho, G. Lee, S. Jung, “*Application of long short-term memory (LSTM) neural network for flood forecasting*”, *Water*, 11(7), 1387, 2019.

# Twitter Veri Seti İçeriğinin Tanımlayıcı Analiz ile Keşfi: Çevrimiçi Yemek Siparişi Üzerine Bir Uygulama

Araştırma Makalesi/Research Article

 Yıldırım GÜNEŞ,  Murat ARIKAN

Endüstri Mühendisliği, Gazi Üniversitesi, Ankara, Türkiye

[yildirimgunes1973@gmail.com](mailto:yildirimgunes1973@gmail.com), [marikan@gazi.edu.tr](mailto:marikan@gazi.edu.tr)

(Geliş/Received:17.10.2022; Kabul/Accepted:16.03.2023)

DOI: 10.17671/gazibtd.1190184

**Özet**— Twitter, her konudaki kullanıcı görüşlerinin dolaşımında olduğu çevrimiçi paylaşım ağlarından biridir. Twitter'dan çekilen kullanıcı görüşlerinin analizi son dönemde yapılan akademik çalışmalarda dikkat çekici şekilde artış göstermektedir. Twitter yeni fikirlere kaynaklık yapan serbest formdaki metin içeriklerinin yanı sıra kullanıcı ve gönderilere ait meta veri ile de değerli bir bilgi kaynağıdır. Bu çalışmada çevrimiçi yemek siparişi konusunda toplanan veri setinin tanımlayıcı analiz araçları ile analizi yapılmış; buradan elde edilen sonuçlarla, işletmelerin iyileştirme/geliştirme maksadıyla kullanabilecekleri müşteri görüşlerinin veri seti içinden nasıl çıkarılabileceğini gösteren ve Twitter Veri Seti İçeriğinin Tanımlayıcı Analiz İle Keşfi olarak adlandırılan bir model sunulmuştur. Modelde analiz sonuçlarının reklam içerikli metinlerin etkisinden kurtarılması için analiz öncesi ve sonrasında uygulanabilecek yöntemler gösterilmiştir. Bu çalışmada 1 Ocak-31 Aralık 2020 tarih aralığında paylaşılmış toplam 35 428 adet tweet analiz edilmiştir. Çalışma sonunda, çevrimiçi yemek siparişine ilişkin Twitter'da yapılan paylaşımların ana gündem konularının; sipariş konusundaki kararsızlık, ürün ve hizmete duyulan güven, kadınların kişisel güvenlik endişesi, hastane yemeklerinin beğenilmemesi, sipariş sonrasında yemeklerin arkadaşlarla paylaşılmak zorunda kalınması, kuryelerin çalışma koşulları, ürün sunumunun beklenen şekilde gerçekleşmemesi, yemek siparişi üzerinden yardımlaşma, işletmelerin indirimli/joker uygulamaları, gece saatlerindeki yemek siparişi sonrasındaki pişmanlık, bir televizyon yemek programının siparişlere etkisi ve yemek siparişi ile karantina ilişkisi olduğu tespit edilmiştir.

**Anahtar Kelimeler**— twitter analizi, tanımlayıcı analiz, çevrimiçi yemek siparişi, çevrimiçi gıda siparişi, içerik analizi

## Exploring Twitter Data Set Content by Descriptive Analysis: An Application on Online Meal Order

**Abstract**— Twitter is one of the online social networking sites where users' opinions about any subject are circulating. The analysis of opinions drawn from Twitter has increased remarkably in recent academic studies. Twitter is a valuable source of information with metadata of users and posts, as well as free-form text content that provides a source for new ideas. In this study, the data set collected on online meal ordering was analyzed with descriptive analysis tools; with the results obtained from here, a model called Exploration of Twitter Data Set Content by Descriptive Analysis, on how to extract customer opinions that businesses can use for improvement/development purposes, is presented. In the model, methods that can be applied before and after the analysis are shown in order to save the analysis results from the effects of advertising texts. In this study, a total of 35 428 tweets posted between January 1, 2020 and December 31, 2020 were analyzed. At the end of the study, it has been determined that the main topics about online meal order on Twitter are; indecision about meal order, confidence in product and service, women's personal safety concerns, hospital meals are not liked, meals have to be shared with friends after ordering, working conditions of moto couriers, product presentation does not take place as expected, discount/wildcard applications of businesses, regret after ordering meal at night, the effect of a television cooking program on orders and the relationship between online meal ordering and quarantine.

**Keywords**— twitter analysis, descriptive analysis, online meal ordering, online food ordering, content analysis

## 1. GİRİŞ (INTRODUCTION)

Hayatın her alanına ilişkin ses, görüntü, serbest metin ya da doküman gibi içeriklerin çevrimiçi paylaşıldığı sosyal ağlar, söz konusu içerikleri geri besleme, iyileştirme ve geliştirme, inovasyon için kaynak olarak kullanan işletmeler açısından değerli birer maden ocağı haline gelmiştir. Facebook, Youtube, Twitter, Instagram gibi uygulamalardan çekilen ham veriler analizlerle işlenir ve daha iyi ürün, daha iyi hizmet olarak günlük hayatımıza girerler. Ancak birçok işletme bu ortamları sadece ucuz reklam aracı olarak kullanmaktadır.

Twitter son dönemlerde kullanıcı görüşlerinin elde edilebildiği mecralardan biri olarak veri madenciliği ve bunun bir alt uygulaması olan metin madenciliği çalışmalarına konu olmakla birlikte, Türkçe literatürde alana özel çalışmaların geliştirilmesine ihtiyaç vardır. Bu çalışmayla, son dönemde tüketiciler tarafından yoğun şekilde kullanılan ancak Türkçe literatürde Twitter analizine konu yapılmayan “çevrimiçi yemek siparişi” alanındaki literatüre katkı sağlanacağı değerlendirilmiştir. Çalışma kapsamında konuya ilişkin 1 Ocak -31 Aralık 2020 tarih aralığına ait 35 428 adet tweetin meta verileri üzerinde veri setinin tanımlanmasına yönelik analizler, bu analiz sonuçlarına dayanarak veri setinin içeriği ve ana gündem konularının keşfine yönelik sondajlamalar yapılmış ve *Twitter Veri Seti İçeriğinin Tanımlayıcı Analiz İle Keşfi* adı verilen bir model önerilmiştir.

Bir ürün/hizmet üreten işletmeler; iyileştirme ve geliştirme için telefonla, anketle, çevrimiçi uygulamalarla ürün/hizmet sundukları müşterilerden geri besleme alma çabası içindedirler. Bu çabalar için harcanan zaman ve emek ise müşterilerin isteksizliği, eksik bilgi ile geri besleme sağlamaları gibi nedenlerle istenen sonucu vermeyebilir. Öte yandan sosyal ağlarda kullanıcılar herhangi bir ürün/hizmete ait görüşlerini serbest ve istekli bir şekilde ifade etmekte, olumlu ya da olumsuz içerikli bu görüşlerini yeni paylaşımlarla çok geniş alana yayabilmektedirler [1]. Bu yayılma özelliği nedeniyle gündem belirleyebilme gücüne sahip Twitter ile; serbest metin formunda paylaşım yapılmakta, paylaşımlara görsel ya da metin eklentileri yapılabilmekte, bu paylaşımlar tweet, retweet, reply ile karşılıklı etkileşim özelliği kazanmakta ve geniş bir müşteri kesimin görüşlerinin elde edilebileceği kaynağa dönüşmektedir [2].

Söz konusu kaynaktan elde edilen veriler, analiz yöntemine uygun teknikler, algoritmalar kullanılarak işleminden geçirilir; kullanılan teknikler, yapılandırılmamış veriyi tanıma, işleme ve müşteri ihtiyaç ve tercihlerini anlama konusunda yardımcı olur [3]. Analitik çalışma seviyelerinden biri olarak tanımlayıcı analiz, geçmişte ne olduğunu ortaya çıkararak [4] müşteri ihtiyaç ve tercihlerinin önceliklendirilmesini sağlayabilir.

Finans, sağlık, enformasyon, gazetecilik, politika, pazarlama, iletişim ve psikoloji gibi birçok alanda etkin

olarak kullanılan Twitter’den belirli sınırlamalarla veri çekilmesine müsaade edilmektedir. Bu verilerin çeşitli veri madenciliği araç ve yöntemleri ile analiz edilmesi Twitter analizi olarak tanımlanır. Twitter analizinde; kullanıcılar, tweetler, tweet metin içerikleri, kullanıcılar arasındaki ilişkiler ayrı ayrı ele alınarak veri setine ilişkin farklı bilgiler ortaya çıkarılabilir. Bu bilgileri ortaya çıkaran tanımlayıcı, içerik, metin, duygu ve şebeke analizleri, kendine has uygulama ve detaylar içermektedir. İki ayrı veri seti üzerinde aynı analiz türü kullanılacak olsa bile veri setinin özellikleri ve içeriğine bağlı olarak farklı detay ve uygulamalara ihtiyaç duyulabilir. Analiz için kurulan modeller, verilerden elde edilecek bilgiyi, verinin bilgiye dönüştürülmesi sürecini etkiler, karar süreçlerini destekler [5]. Bu süreçler ise alana özel çalışmalar ile hızlandırılır.

Bu çalışma ile çevrimiçi yemek siparişi konusunda Twitter ortamında paylaşılan tweetlerin meta verileri üzerinde yapılan tanımlayıcı analiz, bu analiz sonuçları ile veri setinin içeriği ve gündem konularına ilişkin bilgilere nasıl ulaşılabileceğinin yöntemleri, önerilen model ile gösterilmiştir. Model, Chae [6]’nin çalışmasındaki Twitter analiz modelinin tanımlayıcı analiz safhası temel alınarak geliştirilmiştir. Model kullanılarak çevrimiçi yemek siparişi hizmeti kullanıcılarının ana sohbet konuları; hizmetin kullanılma zamanları, kullanıcı ve tweetlerin meta verilerine ilişkin istatistikî değerler elde edilmiş ve yorumlanmıştır. Çalışma kapsamındaki analizlerde pandemi öncesi ve pandemi dönemi ayrımı yapılarak pandemi etkisinin sonuçları da ortaya konulmuştur.

Çalışmanın amacı, Twitter kullanıcıları arasındaki “çevrimiçi yemek siparişi” alanına özel sohbetlerden, iyileştirme/geliştirme çalışmalarına kaynaklık edecek yeni fikirlerin tanımlayıcı analiz araçları ile nasıl elde edilebileceğini göstermek ve bunu diğer alanlara da uygulanabilir bir model ile sunmaktır. Bu maksatla yapılan çalışma; ikinci bölümde alana ilişkin *Literatür Araştırması*, üçüncü bölümde *Yöntem ve Model Önerisi*, dördüncü bölümde *Bulgular ve Tartışma*, beşinci bölümde *Sonuç ve Öneriler* başlıkları altında ele alınmıştır.

## 2. LİTERATÜR ARAŞTIRMASI (LITERATURE REVIEW)

Hizmet sektöründe müşteri odaklı işletmelerin, şirket dışı esas veri kaynağını sosyal ağlarda çeşitli şekillerde ortaya çıkan veriler oluşturmaktadır [7]. Otel ve restaurant gibi ağırlama sektörüne ilişkin konularda özellikle genç tüketicilerin ürün ve hizmet deneyimlerini internet üzerindeki ortamlarda hızlı bir şekilde paylaştıkları; bu paylaşımlardaki mesajlara, işletmeler tarafından oluşturulan viral pazarlama mesajlarından daha çok güvenildiği görülmektedir [8]. Lee ve Youn [9], sosyal ağlarda yapılan sohbetlerin müşteriler için önemli bir referans kaynağı olabildiğini; Allsop vd. [10] tüketicilerin, tavsiyeye başvurma ve tavsiye sağlama

konusunda en çok ağızdan ağıza iletişime başvurdukları ürün hizmet gurubunun yiyecek ve içecek işletmeleri olduğunu vurgulamaktadırlar.

Twitter analizlerinde veri seti büyüklükleri ve kapsanan zaman dilimleri çalışma alanlarına göre farklılık gösterebilmektedir. Örneğin kamuoyunu kısa süreli olarak etkileyen konularda, bu etkinin ortaya çıktığı birkaç gün, bir hafta gibi kısa bir zaman dilimi için veri toplanırken; bir ürün ve hizmet konusunda müşterinin görüşlerinin elde edilmesine yönelik çalışmalarda birkaç haftadan başlayıp daha uzun döneme uzanan zaman dilimi için veri toplama işlemi gerçekleştirilebilmektedir. Bu konuda çalışmanın konusu, amacı, ürün, hizmet çeşidine göre karar verilebilir.

Çetin ve Amasyalı [11] tweet metinleri üzerindeki duygu analiz çalışmasında 12 000 tweetten oluşan bir veri setini, Li vd. [12] Fukushima nükleer santral kazasına ait tweetlerden 1520 adetini, Meral ve Diri [13] farklı konulardaki algı analizine yönelik çalışmalarında Haziran-Temmuz 2013 dönemine ait 8 321 adet tweeti, Çevik vd. [14] Toyota'nın gaz pedalındaki bir sorun ile ilgili 2010 yılında paylaşılan tweetlerden 5000 tanesini, Akgül vd. [15] dört aylık bir süreci kapsayan toplam 6 800 tweeti, Koçak vd. [16] hayayolu pazarına yönelik Nisan-Mayıs 2016 dönemine ait 8 762 tweeti, Onan [17] bir aylık bir dönemde topladığı 10 600 tweetten oluşan veri setini, Albayrak vd. [18] sosyal medyaya yönelik veri analiz çalışmasında 10-12 Ekim 2017 dönemine ait toplam 12 739 tweeti, Çoban ve Tümüklü-Özyer [19] Twitter duygu analizi çalışmasında toplam 20 000 tweeti, Taşdelen [20] mültecileri konu alan çalışmasında 1545 tweeti, Akar vd. [21] İstanbul Sözleşmesi'ne yönelik tepkileri ölçmek için Nisan-Kasım 2020 dönemine ait 252 tweeti incelemiştir.

Türkçe literatürde doğrudan Twitter analizi kullanılarak çevrimiçi yemek siparişi konusunda ürün ve hizmete yönelik bir çalışma olmamakla birlikte, yiyecek sektörüne yönelik çalışmalar mevcuttur. Özasan ve Uygur [22] restaurant, fastfood, kafeterya, pastane ve diğer sınıf başlıkları altında faaliyet gösteren işletmelere ilişkin şikayet konularını incelemiş; Armağan ve Eskici [23] haz ve zevk alma, kolaylık ve rahatlık, yeni tatlar deneme arzusu gibi müşterileri çevrimiçi yemek sipariş verme konusunda teşvik eden unsurları incelemişlerdir.

İngilizce literatürde ise yiyecek, gıda ve restoran sektörüne ilişkin Twitter analizinin yapıldığı, müşteri memnuniyetsizlikleri, ilgi alanları ve algılarının ölçüldüğü çalışmalar yer almaktadır. Hingle vd. [24]'nin yemek yeme alışkanlıklarının tespitine yönelik çalışmasında, analitik yazılım araçları ile Twitter'ın gerçek zamanlı yiyecek tüketimi ve davranışsal faktörlere ilişkin veri toplanmasına uygun bir metot sağladığı belirtilmiştir. Chae [6], tedarik zinciri içinde yer alan çalışan ve

yöneticilerin Twitter'daki sohbet ve yorumlarını, Twitter analiz modeli ile incelemiştir. Chae [6]'nin Twitter analiz modeli; Tanımlayıcı Analitik, İçerik/Metin Analitiği ve Ağ Analitiği olmak üzere üç alt başlıktan oluşmaktadır.

Fan ve Gordon [25], sosyal medya analitik sürecini Veriyi Elde Etme, Anlama ve Sunma şeklinde üç ana başlık altında ele almıştır. Park vd. [26] Çin, Japon, Kore ve Tayland restoranlarına ilişkin lokanta algılarının ortaya çıkarılması amacıyla 29 Nisan-2 Temmuz 2014 tarih aralığına ait 86 015 tweeti incelemiş, çalışma sonucunda Çin restoranları hakkında ortalama duygu miktarının diğer Asya mutfaklarına göre daha düşük olduğunu, olumlu tweetlerin yiyecek kalitesi, olumsuz tweetlerin hizmet kalitesi, yiyecek kültürü hakkında olduğunu tespit etmiştir.

Mishra ve Singh [1], Ocak 2015-Ocak 2016 tarih aralığına ait tweetler üzerinde tanımlayıcı analiz, içerik analizi ve atık kategori analizi şeklinde üç analiz kullanmışlar, sığır etinde ortaya çıkan atığın tedarik zinciri süreçlerinde minimize edilmesine yönelik analiz modeli ve bu konuda önleyici tedbir önerileri sunmuşlardır. Bu çalışmada tanımlayıcı analiz kapsamında, tweetlere ait istatistikler, anahtar kelime ve diyez etiketi frekansları gibi ölçümler kullanılmıştır.

Singh vd. [27], Twitter'dan elde edilen verilerin incelenmesi için destekçi vektör makinesini kullanan metin analizi ve hiyerarşik kümelemenin uygulandığı bir Twitter analitik modeli önermişlerdir. Önerilen model ile sığır et ürünleri konusundaki memnuniyetsizlikler ve bunların giderilmesi için tedarik süreçlerinde kimler tarafından ne tür tedbirler alınması gerektiğine yönelik hususlar tespit edilmiştir. Bu yaklaşım ile tedarik zinciri karar vericileri için; müşteri geri bildirimini ve yiyecek ürün kalitesi ve akışı konularında olmak üzere iki konuda kelime grupları ortaya çıkarılmıştır. Bu sonuçlarla müşteri odaklı tedarik zinciri oluşturulması hedeflenmiştir.

El-Khchine vd. [28], tavuk tedarik zinciri yönetiminde Twitter verisinin ve K en yakın komşu noktalar, lojistik regresyon ve destekçi vektör makinesi algoritmalarının kullanımına dayanan bir analitik yaklaşım sunmuşlardır. Çalışma, tavuk ürünlerine ilişkin ana ilgi alanlarını tanımlayarak tüketici merkezli tedarik zincirinin gelişmesine öneriler getirmiştir.

Söz konusu çalışmalar içinde Chae [6]'nin çalışmasının farklı analiz yöntemlerinin bir araya getirilerek, Twitter verilerinden nasıl istifade edilebileceğini gösteren en geniş kapsamlı çalışmalardan biri olduğu görülmektedir. Twitter verisinin zenginleştirilmiş doğal yapısı, tweetler, kullanıcılar, diyez etiketleri, URL'lere ilişkin geniş çaptaki ölçüm setlerini kullanarak bilgi elde edilmesine imkan verir [6]. Çevrimiçi yemek siparişi konusuna yönelik bu çalışmada Chae [6]'nin modelindeki tanımlayıcı analiz safhasına, zamanla ilişkili aktivite istatistikleri ilave edilmiş, veri setinden reklam içerikli tweetlerin ayıklanması yöntemleri için öneriler getirilmiş,



tanımlayıcı analiz sonuçlarından ve diyez etiketi istatistiklerinden veri setinin gündem konularına ulaşma yolları gösterilmiştir.

### 3. YÖNTEM VE MODEL ÖNERİSİ (METHOD AND MODEL PROPOSAL)

#### 3.1. Analiz Araçları (Analysis Tools)

Veri setleri, bazı ön işlemlerden geçirildikten sonra analize tabii tutulurlar. Bu ön işlemler içerik/metin analizlerinde detaylı işlemler gerektirirken, tanımlayıcı analizlerde veri setlerinin basit ayıklama işlemlerinden geçirilmesi yeterli olabilmektedir. Bununla birlikte tanımlayıcı analizlerin detayı veri setinin içeriği, kapsamı, elde edilme yöntemi, amacı gibi hususlara göre değişiklikler gösterebilir. Bu çalışmada tanımlayıcı analiz sonuçları kullanılarak veri setinin içeriği ve ana gündem konuları belirlendiği için doğrudan içerik/metin analizi araçlarının kullanılmasına ihtiyaç duyulmamıştır. Bu nedenle çalışmada ön işlem olarak, reklam içerikli tweetlerin ve tekraren veri setine girmiş olan tweetlerin veri setinden ayıklanması yeterli olmuştur. Ayrıca veri çekme işlemleri ile veri setlerine çok sayıda haber içerikli tweetlerin girdiği tespit edilmiştir. Haber içerikli tweet frekansları da analiz sonuçlarını etkileyebilir. Yapılacak çalışmanın amacına göre haber içerikli tweetlerin ayıklanması konusunda karar verilmesi gerekir. Bu çalışma kapsamında, doğrudan kullanıcılara ait görüşleri içermese de haber içerikli tweetler, kullanıcı algısına olumlu ya da olumsuz direkt tesir edebileceği için veri setinden ayıklanmamıştır.

Metin analizlerinde kelime sayıları, cümle uzunlukları, kullanılan diyez etiketleri ve bunların frekansları; kullanıcıların eğilimleri, genel karakteristiği ve veri setinin içeriğine ilişkin ipuçları verir. İnternet ortamında birçok alanda olduğu gibi sosyal ağlarda da kullanılan diyez etiketleri ise kullanıcıların ilgi alanlarını ortaya koyar ve hedef kitleler ile iletişim kurulmasını sağlar [29]. Diyez etiketlerinin frekans analizi, diyez etiketlerinin popülaritesini gösterir, veri kümesinin ana gündem maddelerinin ve öncelikli olduğu konuların belirlenmesini sağlar[30].

Ana gündem konularının belirlenmesi için kullanılan yöntemlerden bir diğeri tweet metni içindeki kelimelerin frekans analizidir. Ancak kelime frekans analizi, diyez etiketlerinin frekans analizinden farklı olarak kelime kelimenin kök haline dönüştürülmesi, sözlük kullanımı gibi doğal dil işleme yöntemlerinin kullanılmasını gerektirir. Çalışma kapsamında çevrimiçi yemek sipariş alanına özel bir sözlük hazırlanmış, kelimeler en yaygın kullanılan yapım/çekim ekleri almış şekilleri ya da kök halleri ile sözlüğe dahil edilmiş ve veri setinin tweet metinlerindeki frekans tespiti bu sözlük ile yapılmıştır.

Veri seti içindeki tweetlerin kaçının orijinal (sohbet başlatan) tweet, kaç tanesinin retweet (RT) olduğu, veri setinde kullanıcıların gündem oluşturabilme etkilerini, veri setinin konu çeşitliliğini ve tweetlerin kullanıcılar

arasında yayılımını belirlerken; Twitter'da daha önce favoriye alma şeklinde adlandırılan ve Kasım 2015'den sonra beğenme (like) olarak kullanılan beğeni sayıları da konunun olumlu yönlerinin ortaya çıkarılmasını sağlayabilir. Tanımlayıcı analiz kapsamında ayrıca tweet miktarı, kelime sayısı, kullanıcıların paylaşımlarının ortalama değerleri kullanıcıların genel karakteristiğinin ortaya çıkarılmasını sağlar. Kim tweetledi, kim cevapladı ve kim retweetledi sorularının cevapları, iş alanına yönelik değer katmak isteyen uygulamacılar ve bir fenomeni inceleyen araştırmacılar açısından önemlidir [6].

Kullanıcı istatistikleri ile kullanıcı başına düşen ortalama tweet sayıları, en aktif, en etkili, en görünür (görünürlüğü en yüksek), en fazla beğeni alan kullanıcıların kimler olduğu işletmelerin hedef kitleleri arasında önceliklendirme yapabilmeye imkan verir. Aktif ve görünür kullanıcı/kullanıcı grupları hakkındaki bilgiler, yeni bilgilere ulaşılmasının da yolunu açar [6]. Veri seti içinde en aktif kullanıcı, yazıp paylaştığı orijinal tweet sayılarına göre belirlenir, bu ölçüm kullanıcının sosyal ağ ortamındaki aktivitelerinin sayılması işlemidir [31]. Kullanıcının etkinlik ölçümü ise takipçi sayıları ile yapılır [6], bununla birlikte bir kullanıcının yaptığı paylaşımının ne kadar beğeni aldığı da o paylaşımın ortamdaki etkisini gösteren bir değerdir.

Kullanıcıya ait ölçüm araçlarından bir diğeri de sosyal ağ ortamındaki görünürlüğün tespitidir. Görünürlük, alınan cevap tweeti (Reply) ile alınan RT'lerin toplamı (ya da bu her iki değer ayrı ayrı frekansları) şeklinde ifade edilebilir. Görünürlük ifadesinin bir başka kullanım şekli ise, diyez etiketinin kullanımı ile ilişkilidir. İnternet ortamında kullanıcıların yaptıkları sorgulama sonuçlarında görünebilmek için diyez etiketi kullanılır. Yazılan tweetin konusu ile ilgili kelimelerin tweete, diyez (#) sembolü eklenerek ayrıca yazılması, o kelime ile başka kullanıcıların yaptıkları aramalarda görünür olmayı sağlar. Bu şekildeki görünürlük uygulamasını özellikle reklam içerikli tweet paylaşımlarında görmek mümkündür ve diyez etiket frekanslarına ilişkin yorumlar bu durum dikkate alınarak yapılmalıdır.

Tweetlerin görsel, metin, belge, rapor, makale gibi eklentilerle genişletilmesini sağlayan tekdüzen kaynak bulucu/konumlayıcıların (Uniform Resource Locator/Identifier-URL) veri setindeki sayısı, içerikleri ve en çok kimler tarafından kullanıldığı veri setinin içeriğinin detaylandırılmasını sağlar. URL analizi için Karabulut ve Küçükşille [29] Python ile URL Lib kütüphanesi kullanarak bağlantı analiz algoritması geliştirmişlerdir. Bu çalışmada ise kullanıcıların tweetlere URL ekleme eğiliminin olup olmadığı ve URL'lerin veri setindeki etkisi incelenmiştir.

Yukarıda verilen istatistiklere göre yapılacak sıralamalar, yüzdelik dilim gruplarına ayrılarak karşılaştırma ve değerlendirmeler yapılabilir. Yüzdelik dilimlere göre ayırma işlemi frekanslara göre sıralama yapıldıktan sonra en üst sıradaki %1'lik, orta bölümdeki %90'lık ve son

bölümdeki %9'luk dilimler şeklinde inceleme yapılabilir. Ayrıca işletmeler, müşterilerdeki dönemsel değişikliklerin ne şekilde oluştuğunu görmek ve buna göre tedbir almak isterler [32]. Söz konusu dönemsel değişikliklerin ortaya çıkarılması amacıyla kullanıcı, tweet ve URL istatistiklerinin tümünün zaman dilimleri ile ilişkilendirilerek grafiğinin çıkarılması faydalı bilgilerin elde edilmesini sağlar [31]. Bu analizler çalışma amaçlarına uygun farklı kombinasyonlarla genişletilebilir.

Tweetin hangi kaynaktan üretildiği konusu tweetin platform bilgisinin tespiti ile elde edilir [33]. Bu bilgi elde edilen istatistikî değerlerin doğru yorumlanmasına, web ve uygulama geliştirme çalışmalarının önceliklendirilmesine ve otomatik üretilmiş olabilecek tweetlerin tespit edilerek ayıklanmasına katkı sağlayabilir.

### 3.2. Veri Setinin Analize Hazır Hale Getirilmesi ve Model Önerisi (Preparation of Dataset and Model Suggestion)

Çalışmada, Twitter API (Application Programming Interface-Uygulama Programlama Arayüzü)'sinin uyguladığı %1'lik, her seferinde son bir aya ait ve 3200 adet tweet çekebilme sınırları içinde kalınarak bir yıllık veri seti oluşturulmuştur. Tweet çekme işlemlerin tweetlerin %100'üne ulaşma imkanı da olmakla birlikte bu oldukça pahalı ve maliyetli olması nedeniyle büyük işletmeler tarafından tercih edilen bir yöntemdir. Veri çekme işlemlerinde, hazır bir uygulama olarak ücretli Twitonomy uygulaması kullanılmıştır. Standart veri çekme işlemlerinde, kullanıcılar tarafından korumalı/özel olarak işaretlenmiş tweetler, API ya da uygulamalar aracılığı ile veri toplama esnasında veri havuzuna çekilemezler [34]. Bu nedenle çalışma konusunun veri seti, yalnızca kamuya açık tweetlerden meydana gelmektedir. Bu durum, verilerin temsil yeteneği açısından bir kısıtlılık ortaya çıkarabilir. Aynı zamanda reklam ya da propaganda maksatlı yazılmış, tweet ve retweetlenmiş tweetlerin veri setinden tamamen ayıklanması mümkün olmayacağından, analiz kalitesinde düşüş olabilir [26] ve literatürdeki diğer çalışmalarda olduğu gibi reklam içerikli tweetleri ayıklamadan yapılacak çalışmalar *bu hususun analiz sonuçlarını olumsuz etkilemeyeceği* kabul edilerek sürdürülebilir. Ancak bu çalışmada, analiz kalitesini yükseltmek için, reklam içerikli tweetleri ayıklamada kullanmak üzere *alana özel reklam sözlüğü* hazırlanarak veri seti üzerinde kullanılmış, ayrıca reklam içerikli tweetlerin ayıklanmasına yönelik ilave öneriler getirilmiştir.

Veri setinin oluşturulmasında doğru verilere ulaşılabilmesi için veri çekmede kullanılacak anahtar kelimelerin amaca uygun olarak doğru belirlenmesi gerekir. Belirlenen anahtar kelimeler ile veri havuzu oluşturulurken ilgisiz verilerin de veri havuzuna girmesi mümkündür. Atan ve Çınar [35], arama kelimeleri ile arama yapılarak ilgili metinlere ulaşabilme konusunda, milyonlarca farklı insan tarafından kaleme alınan metinsel veriler içinden ilgili dokümanı tespit etmek için kurgulanmış her algoritmanın yanılmasının olası olduğunu, profesyonel arama motorlarının bu durumu

aşmak için çeşitli yaklaşımlar geliştirdiklerini, ancak yine de arama sonuçlarında ilgisiz verilerin varlığını da sürdürmeye devam edebileceklerini belirtmişlerdir. İlgisiz konudaki tweet gönderilerinin veri setine girmesini önlemenin/en aza indirmenin en etkili yollarından biri, doğru anahtar kelimelerin seçilmiş olmasıdır. O nedenle, veri çekme işlemi gerçekleştirilirken başlangıçtan itibaren veri seti kontrol edilmeli, eğer konuyla ilgisiz verilerin çekildiği tespit edilirse, anahtar kelimeler gözden geçirilmelidir.

Veri havuzunun oluşturulmasında kullanılacak anahtar kelime tespiti için sırasıyla; internet üzerinden Türkiye'de eve/işyerine hazır yemek servisi hizmeti veren şirketlerin kullandığı platformlar, Twitter kişisel kullanıcı hesabından Kasım 2009-Ağustos 2019 tarih aralığına ait 221 tweet metni, Google arama motoru üzerinden konuya ilişkin yorum, şikâyet ve öneriler [36-43] konuyla ilgili tartışma/karşılaştırma yazıları [44] incelenmiştir.

Yapılan bu sorgulama ve inceleme neticesinde çevrimiçi yemek siparişi konusunda; *yemek siparişi, yemeksepeti siparişi, döner siparişi, lahmacun siparişi, hamburger siparişi, pide siparişi* kelime çiftleri anahtar kelime olarak kullanılmış ve Twitter'dan 1 Ocak-31 Aralık 2020 tarihleri arasında veri çekme işlemi gerçekleştirilmiştir. Anahtar kelime tespiti için konuya ilişkin yapılan incelemelerde, alana özel bir sözlük çalışması yapılmıştır. Bu çalışmada tespit edilen tohum kelime listesi, eş anlamlı kelimeler sözlüğü kullanılarak genişletilmiş ve 2129 kelimelik bir sözlük elde edilmiştir. Tweetlerde argo, jargon, gündelik ve hatalı kelime kullanımının da yaygın olduğu dikkate alınarak, sözlüğe dahil edilmiştir. Bu sözlük kullanılarak veri setinin kelime frekans tespiti yapılmış ve 4'üncü bölümde verilmiştir.

Birden fazla anahtar kelime grubu ile veri çekilmiş olması ve bu kelime gruplarında ortak (*yemek ve siparişi*) kelimelerin yer alması nedeniyle, anahtar kelime grubunun tam ifade olarak sorgulatılmış olmasına rağmen bazı tweetlerin veri setine tekraren girdiği tespit edilmiştir, tekrar eden bu tweetlerin de veri setinden ayıklanması gerekir.

Genelde Twitter gibi sosyal ağlardan çekilerek oluşturulmuş veri setlerinin analizinde reklam içerikli paylaşımların, analiz sonuçlarını etkilemeyeceği ön kabulü ile reklam ayıklama işlemi yapılmamaktadır. Ancak sosyal ağların kullanımının yaygınlaşması, sosyal ağlarda reklam yapma maliyetinin çok düşük seviyede ya da maliyetsiz olması, bireysel reklamlardan büyük işletme reklamlarına kadar çok sayıda reklam içeriğinin sosyal ağlarda dolaşımına neden olmaktadır. Bu reklamları içine alan veri setleri üzerinde yapılacak analizlerin kalitesinde düşüş olabileceği ve yanıltıcı sonuçlara gidilebileceği öngörülmektedir. Bu çalışma için oluşturulan veri setindeki reklam tweet oranının aylık dilimler bazında incelendiğinde %12-16 arasında olduğu tespit edilmiştir. Bu hali ile yapılacak frekans bazlı analizlerin doğru sonuçlara ulaşması mümkün görülmemektedir. Çevrimiçi

etkileyicileri de (Influencer)<sup>1</sup> kullanarak reklamların yapıldığı sosyal ağlardan çekilmiş veri setlerinden reklam içerikli metinlerin %100 olarak ayıklanma imkanı olmasa da, çalışma kapsamında uygulanan reklam ayıklama yöntemlerinin, analizlerin farklı aşamalarında kullanımı ile reklam içerikli metinlerin analize etkilerini azaltmanın mümkün olduğu görülmüştür.

Çalışma kapsamında, reklam tweetlerinin ayıklanması için öznitelik seçim yöntemine benzer yolla 168 kelime/kelime öbeğinden oluşan bir sözlük hazırlanmış, sözlüğe kelimelerin hatalı yazım ve jargon kullanımları da dahil edilmiştir. Sözlüğün hazırlanması için; veri setinden rastgele 325 adet reklam içerikli tweet seçilmiş, bu tweetler tek tek incelenerek frekansı en yüksek olan ve reklama has nitelikleri ön plana çıkaran kelimeler seçilmiş ve *alana özel reklam sözlüğü* hazırlanmıştır. Sözlükteki tüm kelime/kelime öbekleri veri setindeki tüm tweetlerde arattırılmış ve bunlardan birinin tweet metninde yer alması durumunda söz konusu tweetin reklam tweeti olduğu kabul edilerek veri setinden çıkarılmıştır. Reklam sözlüğü kullanarak, veri setinden reklam içerikli tweetlerin ayıklanması için MS Excel 365'in veri analiz özellikleri kullanılmıştır.

Veri seti üzerinde yapılan incelemelerde, reklam içerikli tweetlerin veri setinden ayıklanmasında aşağıdaki maddelerde sıralanan beş yöntemin ayrı ayrı ya da birlikte kullanılabilmesi görülmüştür, bu çalışma kapsamında ilk ve son sıradaki yöntem birlikte kullanılmıştır:

(i) Sektöre ilişkin reklam içerikleri analiz edilerek, kullanılan kelimeler *alana özel reklam sözlüklerine* dönüştürülebilir ve sözlük tabanlı yaklaşım metodu ile veri setinden reklam içerikli metinler ayıklanabilir.

(ii) Veri setindeki tweetlerin kullanıcı bilgilerinde *kullanıcı (Handle)* ya da *kullanıcı adı (Name)* başlıkları altında yer alan bilgiler kontrol edilerek, sektörde faaliyet gösteren bir işletme adı olanlar tespit edilir ve veri setinden bu isimli kullanıcıların tweetleri ayıklanabilir. Reklam içerikli olanların yanı sıra, işletmelere ait bazı

<sup>1</sup> “Çevrimiçi etkileyiciler” ifadesi, sosyal medyada kullanılan İngilizce “Influencer” kelimesinin karşılığı olarak kullanılmıştır. Influencer, sosyal medyada aktif, hedef kitlesi geniş ve paylaşımları ile dijital tüketicileri etkileme gücü olan, paylaşım içeriklerinde markaların reklamını yaparak tüketicilerin tercihini etkileyebilen, paylaşımlara yorum yaparak hedef kitlelerinin dikkatini ürün ya da işletmelere çekebilen sosyal medyadaki kanaat önderleridir (Türkçe karşılığı olarak yaygın kullanıma sahip bir kelime olmamakla birlikte, ilham verici, yol gösterici, etkileyici, fikir önderi gibi kelimeler kullanılabilir). Çok sayıda takipçisi olan makro ve tematik paylaşımları ile daha sınırlı takipçi sayısına sahip mikro influencer olarak iki grup olarak isimlendirilirler [46].

tweetlerin kendi müşterilerinin sorularına cevapları olabileceği dikkate alınarak, bunlar çalışmanın amaçlarına bağlı olarak veri setinden ayıklanabilir ya da veri setinde tutulmaya devam edilir.

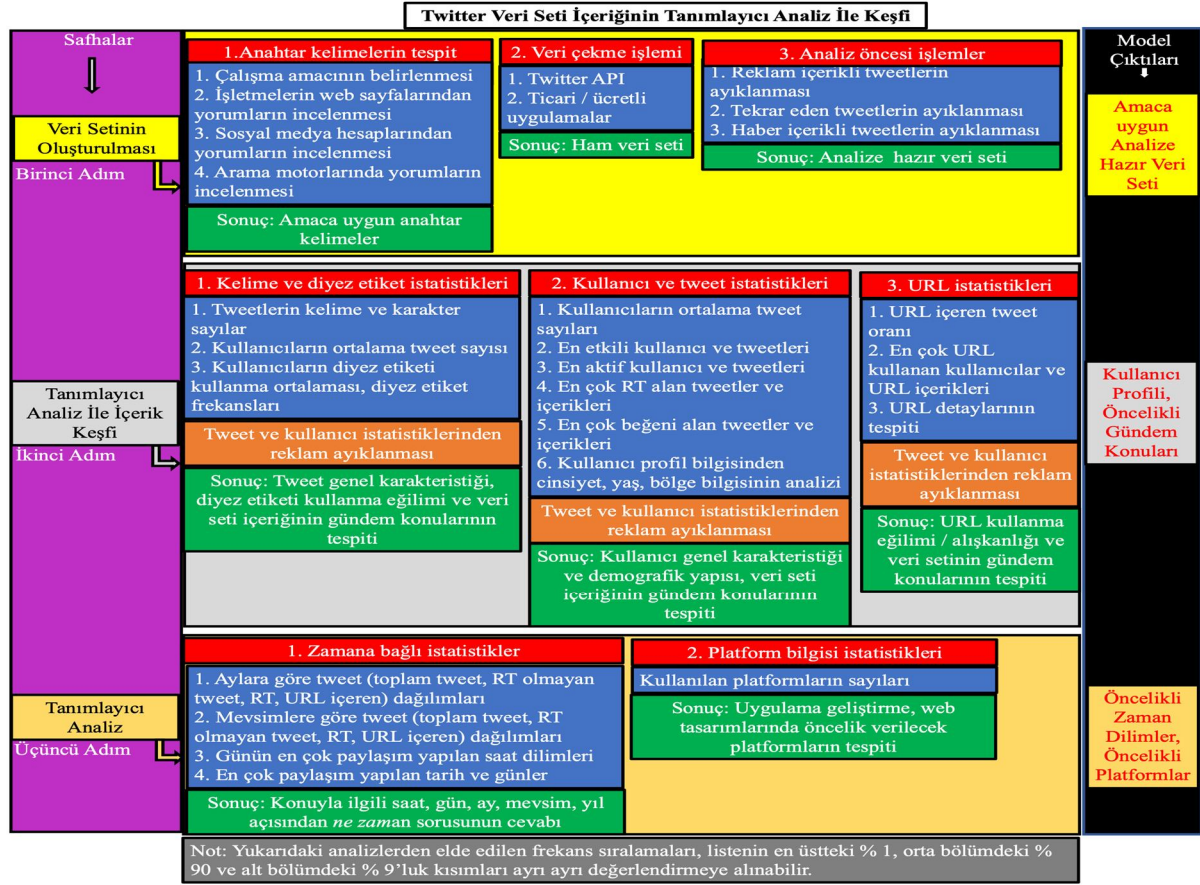
(iii) Yorum eklenmeden sektörde faaliyet gösteren bir işletmenin paylaşımını RT'leyen (metin kısmı “RT@alandaki işletmenin kullanıcı ismi” şeklinde başlayan) tweet gönderileri reklam tweeti kabul edilerek veri setinden ayıklanabilir.

(iv) Tweetin paylaşım yapılan platform bilgisi de reklam ayıklama işleminde bir araç olarak kullanılabilir. Kullanıcıların günlük tweet paylaşımları arasında reklam tweetleri yayımlamak, bu reklamların zaman içinde kaybolmadan belirli aralıklarla yenilenmesine devam etmek için sosyal ağların kontrolü, düzenlenmesi, takip edilmesi, bu işlemlere hız kazandırılması amacıyla çeşitli yönetsel araçlar, platformlar, web uygulamaları, ssd (solid state drive-katı hal sürücüsü)'ler gibi farklı araçların kullanımına ihtiyaç duyulur. Tweet yayımlama işini kolaylaştıran bu araçlar, tweetlerin meta verisi olarak *platform* başlığı altında görünmektedir. Bu çalışmada kullanılan veri setindeki tweetlerin platform bilgileri incelendiğinde; “Twitter for Android”, “Twitter for iPhone”, “Mobile Web (M2)”, “Twitter Web App”, “IFTTT”, “Clarabridge Engage” ve “Instagram” şeklinde bilgilerin yer aldığı görülmüştür. İlk iki sırada yer alan platform bilgisine sahip tweetler dışındakilerin reklam içerikli tweetler oldukları tespit edilmiştir.

(v) Veri seti üzerinde yukarıda sıralanan reklam ayıklama işlemleri yapılmaya da analiz sonundaki değerlerin reklam içerikli tweet, URL ya da kullanıcıya ait olup olmadığı kontrol edilmeli ve reklamlara ait olduğu tespit edilenler değerlendirme dışı bırakılmalıdır.

Yukarıda sıralanan ayıklama işlemleri sonucunda; tweetlerin %2'si (1070 tweet) bir takvim yılı dışında kalması ve takvim yılını esas alarak değerlendirme yapılabilmesi amacıyla, %7'si (2 848 tweet) reklam içerikli ve %7'si veri setine tekrardan girmiş (2864 tweet) tweetler olması nedeniyle ayıklanmıştır. %16'lık bu ayıklama sonucunda 35 473 tweet gönderisi kalmıştır. Twitter'den veri çekme işlemleri sırasında Excel tablosunda boş satırların ortaya çıktığı görülmüş ve analiz aşamasında hatalı sonuçlara neden olabilen bu şekildeki 45 boş satır veri setinden çıkarılmıştır. Kalan 35 428 tweet analize esas veri seti olarak kullanılmıştır.

Yukarıda belirtilen işlemler çerçevesinde önerilen model Şekil 1'de gösterilmiştir. Söz konusu model ile 35 428 tweetten oluşan veri seti, MS Excel'in veri analiz özellikleri ve Python programlama dili kullanılarak analiz edilmiş ve görselleştirilmiştir. Analiz sonuçları 4'üncü bölümde sunulmuştur.



Şekil 1: Twitter veri seti içeriğinin tanımlayıcı analiz ile keşfi için önerilen model  
(Proposed model for the discovery of Twitter dataset content by descriptive analysis)

#### 4. BULGULAR VE TARTIŞMA (RESULTS AND DISCUSSION)

Veri setini oluşturan 35 428 adet tweetin toplam kelime sayısı 698 051'dir. Bu sayıya işaret, sembol, resim, diyez etiketleri, kullanıcı sembolleri dahildir. Buna göre bir tweetin ortalama uzunluğu 19,7 kelimedir. Veri setindeki tüm tweetlerin boşluklu karakter sayısı 5 237 640, boşluksuz karakter sayısı 4 584 513'tür. Buna göre bir tweetin ortalama karakter sayısı, boşluklarla birlikte 147,8, boşluksuz 129'dur. Twitter uygulaması 280 karaktere kadar yazım için müsaade etmektedir. Yemek siparişi konusundaki sohbetlerin, 20 kelimelik ve 148 karakterden Twitter'ın genel kullanım eğilimine benzer uzunluktaki paylaşımlardan oluştuğu görülmektedir.

Veri setinin gündem konularının belirlenmesinde kullanılabilir ölçütlerden diyez etiket frekansları incelendiğinde, tweetlerin %9'unda (3 300 tweet) diyez etiketi bulunduğu, diyez etiketinin toplam kelime sayısı içinde %2 orana sahip olduğu ve yaklaşık 2,7 kullanıcıya bir diyez etiketi düştüğü görülmektedir. Bu değerlere göre diyez etiketi kullanım oranının düşük olduğu söylenebilir. Veri setindeki tüm tweetlerde birbirinden farklı olarak 10 690 diyez etiketi kullanılmıştır. Burada aynı köke sahip kelimelerin çekim ekleri ile türetilmiş halleri, hatalı ve eksik yazılmış şekilleri de ayrı kelimeler olarak sayılmıştır. Yabancı dilde yazılmış olan diyez etiketi dışında olan kelimeler için genel olarak mantıksal,

anlamsal ve aynı köke sahip olma gibi birbiri ile ilişkilendirilebilecek kelimeler aynı grup içine dahil edilerek, 10 690 diyez etiketi, 1537 gruba ayrılmıştır. Gruplandırma sonrası diyez etiketleri frekanslarına göre frekansı 50'den fazla olanlar Tablo 1'de frekansları ile birlikte gösterilmiştir.

Tablo 1. Tüm kullanıcıların tweetlerindeki diyez etiket frekansları (Hashtag frequencies used by all users)

Diyez Etiket Frekansları			
Diyez Etiketi	Frekansı	Diyez Etiketi	Frekansı
#gelal	800	#et/etdöner	77
#eveservis/evesi pariş	793	#restoran	73
#YemekSepeti	672	#burger	70
#siparis	547	#ankara	69
#yemeksiparisi/ yiyecek	408	#kurye	65
#paketservis	266	#lahmacun	62
#evdehayatvar/ EvdeKal	205	#getir	59
#corana	191	#online	59
#lezzet	172	#samsun	58
#tavuk	133	#food	56
#masterchef	109	#HayatEveSığar	55
#YökErkenMez uniyeteDurDe	107	#kebab	54
#istanbul	100	#dürüm	53
#Pizza	99	#canik	52
#pide	92	#hamburger	52
#saglik	84	#ilkadım	52

Tablo 1'in kelime bulutu gösterimi Şekil 2'de verilmiştir.

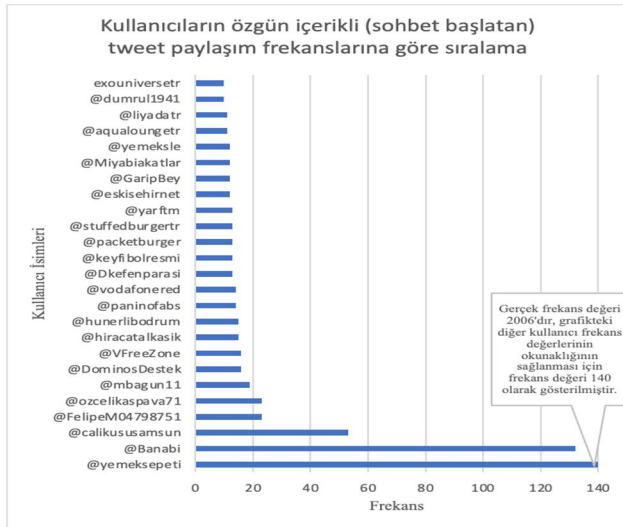




Şekil 7. Favoriye eklenme/beğeni alma frekans sıralaması  
(Frequency order of favorites/likes)

Potansiyel müşterilerin tespit edilmesine katkı sağlayan *en aktif* ve *en etkili* kullanıcı ölçütleri, bu kullanıcıların tweetleri üzerinden veri setinin gündem konularına ulaşılmasını sağlayabilir. Veri setindeki 35 428 tweet gönderisi, toplam 28 824 kullanıcıya aittir. Çevrimiçi yemek siparişi konusunda sohbet eden kullanıcıların bir yıl içinde ortalama 1,23 tweet gönderisi yaptıkları görülmektedir. Veri setinden RT'ler çıkarıldıktan sonra geriye 23 507 orijinal tweet kalmaktadır ki bu sayı kullanıcı sayısına bölüldüğünde bir kullanıcının özgün içerikli paylaşım sayısının bir yıl içinde 0,82 tweet olduğu görülmektedir. Buna göre kullanıcıların çevrimiçi yemek siparişi konusunda tweet paylaşımlarının yoğun olmadığı söylenebilir.

Özgün içerik paylaşımı yapan (sohbet başlatan/orijinal) tweet sayılarına göre belirlenen *en aktif kullanıcıların* ilk 25'inin kimler oldukları incelendiğinde; bu kullanıcıların altısı dışındakilerin ya işletmelere ait kullanıcı ya da alana özel reklam içerikli paylaşımında bulunan kullanıcılar oldukları Şekil 8'de görülmektedir. Söz konusu altı kullanıcının tweetlerinin içeriğinde yeni olarak; ürün teslim edilmediği halde kullanıcıya teslim edildi mesajı gönderilmesi, yemek siparişinin restorana ulaşmaması ve fiyat artışları konularının ortaya çıktığı belirlenmiştir.



Şekil 8. Özgün içerikli tweet frekans sıralaması  
(Frequency order of original tweets)

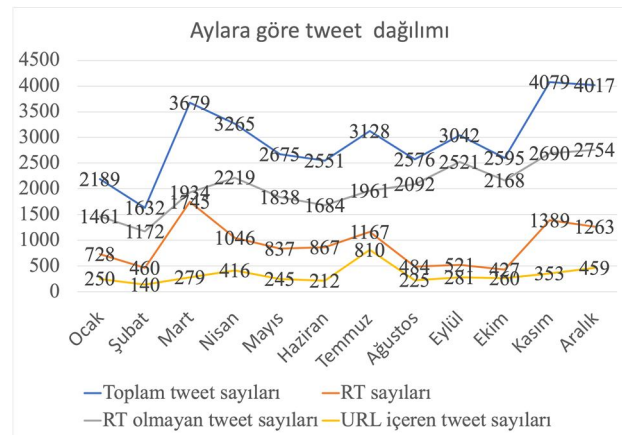
Bir başka ölçüm aracı olan takipçi sayıları ile tespit edilebilen *en etkili kullanıcı* sıralamasının ilk 25'i incelendiğinde, 15'inin haber içeriği üreten kullanıcılar oldukları görülmektedir. Bu kullanıcıların "çevrimiçi yemek siparişlerinin COVID-19 salgınının yayılmasına etkileri" ile "motosikletli kurye kazaları" konularında haber yaptıkları için veri seti içine girdikleri, takipçi sayılarının yüksek olması nedeniyle de en etkili kullanıcılar arasında yer aldıkları görülmüştür. Sıralamadaki kullanıcılardan dördü resmi kurum olması

nedeniyle değerlendirme dışında bırakıldığında altı potansiyel müşteri kalmaktadır.

En etkili kullanıcı olan altı potansiyel müşterinin gündem konularının; yemek sipariş hizmetinin hayati kolaylaştırıcı olması, siparişteki yoğunluk ve buna bağlı gecikmeler, bir yemek programının yemek siparişi vermeyi teşvik edici etkisi, sektörde faaliyet gösteren bir işletmenin müşterilerine kitap hediyesi kampanyasının olumlu etkisi, bazı müşterilerin koronavirüs salgını sonrası yemek siparişi verme alışkanlıklarını terk etmeleri olduğu tespit edilmiştir. Herhangi bir konunun, haber içeriği üreten Twitter kullanıcıları tarafından paylaşılması, müşteri görüşleri dışında kalıyor olmasına rağmen işletmeler tarafından dikkate alınması gerekir. Bu nedenle çevrimiçi yemek siparişlerinin COVID-19 salgınının yayılmasına etkileri ve motosikletli kurye kazaları konuları da kamuoyunu doğrudan etkileme gücü nedeniyle veri setinin esas gündem maddelerine dahil edilmelidir. Burada haber konusu olarak paylaşılan motosikletli kurye kazaları konusunun, diğer istatistiksel incelemelerde de başka kullanıcılar arasında kuryelerin çalışma koşulları şeklinde sohbet konusu olduğu görülmüştür. Bu nedenle kuryelerin çalışma koşullarına ilişkin tedbirlerin diğer ana gündem konularından daha önce ele alınarak (örneğin üç tekerli motosikletlerin kullanıma sokulması gibi) tedbirler üretilmesi ve alınacak tedbirlerin kamuoyu ile paylaşılması hususu işletmeler tarafından değerlendirilmelidir.

Gündem konularının detaylandırılmasına katkı sağlayan URL istatistikleri incelendiğinde; tweetlerin %9'unda URL bulunduğu, tweetine URL eklentisi yapan ilk 25 kullanıcıdan sadece üçünün potansiyel müşteri olduğu, diğerlerinin reklam içerikli ve işletmelere ait tweetler olduğu görülmüştür. Veri setindeki kullanıcılar arasında URL kullanımının yaygın olmadığı tespit edilmiştir.

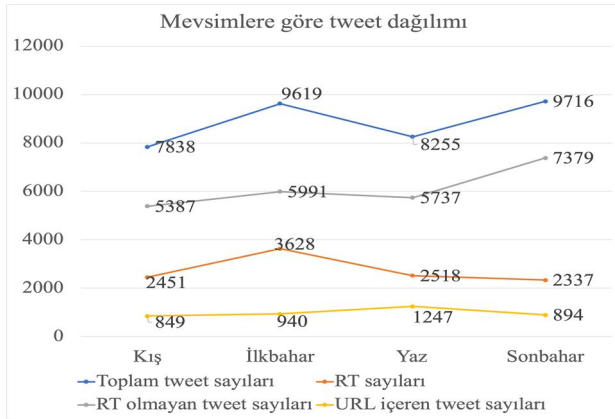
Yukarıda sıralanan ölçütlerin zaman birimleri ile ilişkilendirilmesi, işletmeler için kullanışlı ilave bilgilerin ortaya çıkarılmasını sağlayabilir. Müteakip maddelerde tweet, RT, URL frekanslarının farklı zaman dilimleri ile ilişkilendirilmesinden çıkarılan sonuçlar verilmiştir. Tweetlerin aylara göre dağılımı Şekil 9'da verilmiştir.



Şekil 9. Tweetlerin aylara göre akışı-zaman grafiği  
(Flow of tweets by month-time graph)

Şekil 9 incelendiğinde çevrimiçi yemek siparişi konusunda Twitter’da paylaşımların en düşük olarak Şubat ayında gerçekleştiği, Mart ayında yıl içinde ikinci yüksek seviyeye ulaştığı, Nisan ayında düşüşe geçerek Kasım ayı ortasına kadar birbirine yakın düzeylerde devam ettiği, Kasım ayı ortasından sonra Aralık ayına kadar yılın en yüksek seviyesine ulaştığı görülmektedir. Mart, Kasım, Aralık aylarında yemek siparişi konusunda paylaşımlardaki artış, bu dönemlerde işletmelerin satışlarında da artış olduğunu/olabileceğini göstermektedir.

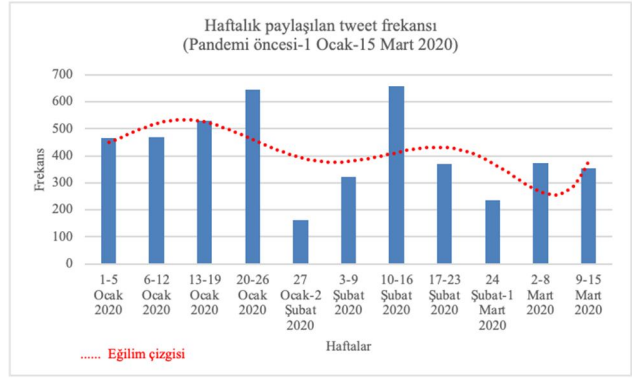
Şekil 10’da, Şekil 9’daki bilgilerin mevsimlere göre uyarlanmış hali verilmiştir. Mevsim olarak online yemek siparişi konusundaki sohbet ve paylaşımların İlkbahar ve Sonbahar dönemlerinde yüksek seviyede gerçekleştiği görülmektedir. Şubat ayında ve yaz döneminde okul tatili, Mayıs ayının Ramazan ayı olması nedeniyle paylaşımlarda düşüş olabileceği yönünde yorum yapılabilir ancak bunun kesin olarak ifade edilebilmesi için ayrıca veri setinin içerik analizi yapılması gerekir.



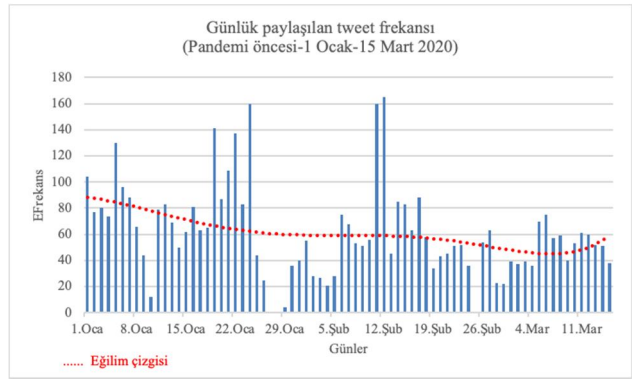
Şekil 10. Tweetlerin mevsimlere göre akışı-zaman grafiği  
(Flow of tweets by season graph)

Veri setini oluşturan zaman dilimi içinde pandemi döneminin de yaşandığı dikkate alınarak yukarıdaki aylık ve mevsimsel incelemeler, pandemi etkisini yansıtabilecek zaman dilimlerine ayrılarak müteakip şekillerde gösterilmiştir. Pandemi açısından tarih olarak, gündelik hayata yönelik tedbirlerin alınmaya başladığı ve okulların tatil edildiği 16 Mart 2020 tarihi esas alınmış; 1 Ocak-15 Mart 2020 tarih aralığı pandemi öncesi, 16 Mart-31 Aralık 2020 tarih aralığı pandemi dönemi olarak incelenmiştir.

Pandemi öncesi döneme ait haftalık ve günlük tweet paylaşımları Şekil 11a ve Şekil 11b’de gösterilmiştir. Her iki grafikten de paylaşımlarda düşüş eğilimi olduğu görülmektedir. Bu dönemde tweet paylaşımındaki haftalık ortalama 417, günlük ortalama 61’dir.

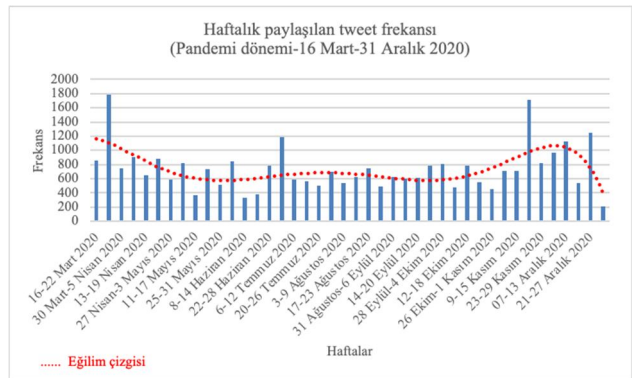


Şekil 11a. Haftalık tweet paylaşım frekansları-pandemi öncesi (Tweet frequencies on weekly basis- pre pandemic)



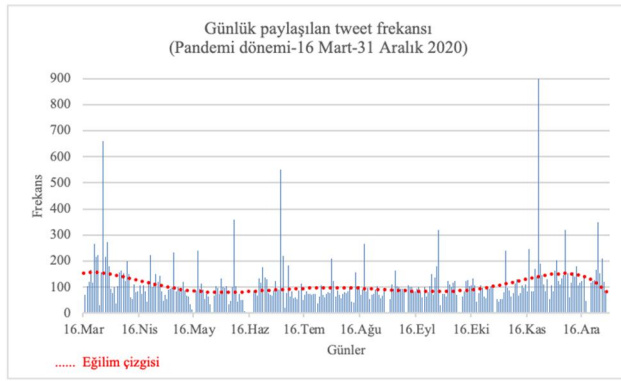
Şekil 11b. Günlük tweet paylaşım frekansları-pandemi öncesi (Tweet frequencies on weekly basis- pre pandemic)

Pandemi dönemine ait haftalık ve günlük tweet paylaşımları Şekil 12a ve Şekil 12b’de gösterilmiştir. Bu dönemde tweet paylaşımındaki haftalık ortalama 734, günlük ortalama 106’dır. Tweet paylaşımındaki ortalama değerler, pandemi öncesindeki ortalamaya göre % 74 arttığı, pandemi dönemindeki tweet paylaşımlarındaki değişkenliğin pandemi öncesindeki döneme göre daha kararlı bir yapıda olduğu görülmektedir.



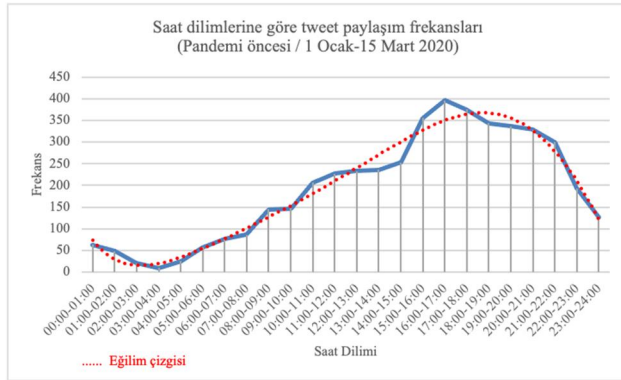
Şekil 12a. Haftalık tweet paylaşım frekansları-pandemi dönemi (Tweet frequencies on weekly basis-pandemic period)



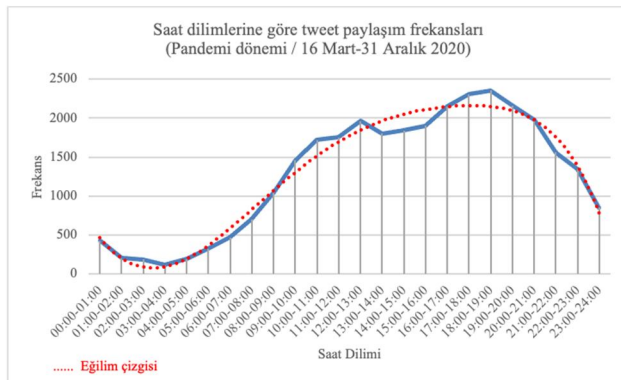


Şekil 12b. Günlük tweet paylaşım frekansları-pandemi dönemi (Tweet frequencies on weekly basis-pandemic period)

Şekil 13a ve Şekil 13b'de paylaşım saatleri esas alınarak tespit edilen tweet frekansları, pandemi öncesi ve pandemi dönemi şeklinde gösterilmiştir. Her iki şekilde de yemek siparişi verme konusunda paylaşımların saat 11:00-12:00'den itibaren yükselişe geçtiği, yüksek seviyenin 21:00-22:00'a kadar korunduğu görülmektedir. Gün içindeki en yüksek seviyenin; pandemi öncesi dönemde saat 16:00-17:00'da, pandemi döneminde saat 18:00-19:00'da ortaya çıktığı görülmektedir.



Şekil 13a. Saat dilimlerine göre tweet frekansları-pandemi öncesi (Tweet frequencies on hourly basis-pre pandemic)

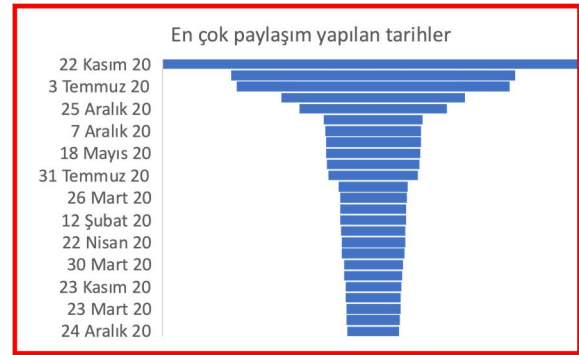


Şekil 13b. Saat dilimlerine göre tweet paylaşım frekansları-pandemi dönemi (Tweet frequencies on hourly basis-pandemic period)

Her iki dönemde ortaya çıkan tweet paylaşım saatlerindeki artış eğiliminin öğle ve akşam yemek

saatleri öncesi ve sonrasında gerçekleştiği görülmektedir. İşletmeler, paylaşımların arttığı bu saatleri, sipariş yoğunluğunun artış gösterebileceği/gösterdiği saatler kabul ederek satış politikalarında düzenlemeler yapabilirler. Örneğin müşteri tercihlerinin etkilenmesi için yapılacak reklamların, indirim ve joker gibi uygulamaların bu saatler dikkate alınarak yapılması daha etkili sonuçlar ortaya çıkarabilir.

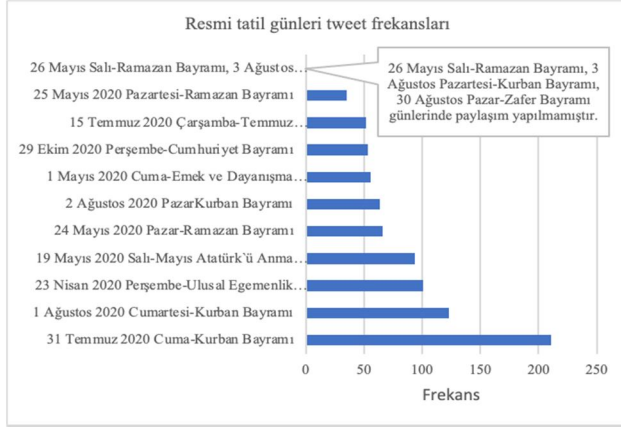
Yıl içinde en çok paylaşım yapılan tarih sıralaması Şekil 14'te gösterilmiştir. Sıralama incelendiğinde ilk üç sırada yer alan tarihlerin, işletmeler tarafından çok fazla sayıda reklam içerikli tweet paylaşımı yapılması nedeniyle öne çıktığı, bu üç tarihte potansiyel müşterilerin tweet paylaşımının çok az sayıda gerçekleştiği görülmektedir. Dördüncü sıradan itibaren gösterilen tarihlerde ise potansiyel müşteri paylaşımlarının yoğun olduğu görülmektedir.



Şekil 14. En çok tweet paylaşımı yapılan tarihler (The most tweeted dates)

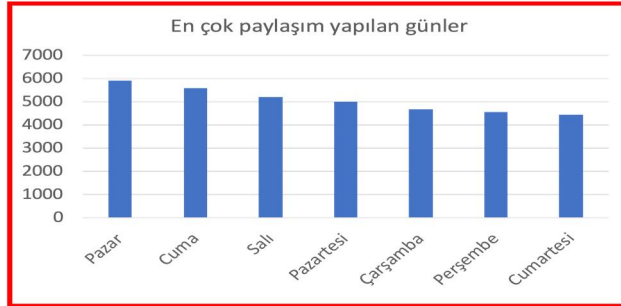
Şekil 14'te herhangi bir özelliği nedeniyle öne çıkan bir tarihin olmadığı görülmektedir. Dördüncü sıradaki 07 Aralık 2020 tarihindeki, kişisel güvenlik endişesi ile çevrimiçi yemek siparişi konusunun ilişkilendirildiği bir tweetin çok fazla sayıda RT yapılarak gün içinde yayıldığı görülmektedir. Gündem belirleyici nitelikteki bu paylaşım, ilk bakışta çevrimiçi yemek siparişi ile doğrudan ilişkili gibi gözükmeyebilir. Ancak tweet içeriği incelendiğinde, yemek siparişini eve teslim eden kuryeler nedeniyle böyle bir ilişkilendirmenin yapıldığı görülmektedir. "Kadınların kişisel güvenlik endişesi" ile ilgili konunun yukarıda yapılan diğer analizlerde de veri setinin ana gündem konuları arasında yer aldığı görülmektedir. Bu konuda işletmelerin doğrudan uygulamaya sokabileceği bir tedbir olmasa da çalışanların davranış şekilleri ve eğitimine yönelik uygulamalar yapılabilir. Örneğin pandemi döneminde ortaya çıkan temassız teslimat uygulaması, bu konuda alınmış bir tedbir olmamakla birlikte, tweetlerde ifade edilen kişisel güvenlik endişesinin azaltılmasına katkı sağlayacak bir uygulama olarak kabul edilebilir.

Resmi tatil günlerinde yapılan tweet paylaşım frekansları Şekil 15’de gösterilmiştir. Bu günlerden, 23 Nisan, 1,19, 24-26 Mayıs tarihlerinde korona virüsü nedeniyle sokağa çıkma yasağı uygulanmıştır. Veri setinin tümü için günlük ortalama tweet paylaşımının 97 iken, resmi tatil günlerinde bu ortalamanın 66 olduğu tespit edilmiştir. Resmi tatil günlerindeki gündem maddelerinin ise kuryelerin uygun olmayan davranışları, siparişin yerine ulaşmaması ve siparişte yaşanan gecikmeler, ürünlerde yabancı madde tespit edilmesi, yemek siparişine duyulan özlem, temassız teslimat şeklinde konuların olduğu görülmektedir.



Şekil 15. Resmi tatil günleri tweet frekansları (Tweet frequencies on public holidays)

Yıl içinde en çok paylaşım yapılan gün sıralaması Şekil 16’da gösterilmiştir.



Şekil 16. En çok tweet paylaşımı yapılan günler (The most tweeted days)

Şekil 16 incelendiğinde, günler arasında çok yüksek seviyelerde fark ortaya çıkmadığı, Cuma ve Pazar gününün en yüksek ilk iki sırada yer aldığı, Cumartesi gününün ise en az paylaşım yapılan gün olduğu görülmektedir. Şekil 16’da görüldüğü gibi tweet paylaşımlarındaki Cuma ve Pazar günlerinde ortaya çıkan artışın, Cumartesi günündeki azalışın çevrimiçi yemek siparişlerindeki artış ve azalış ile paralel olacağı/olduğu öngörülmektedir.

Twitter kullanıcı meta verilerinden biri olan platformlar incelendiğinde; % 91’inin “Twitter for Android”, “Twitter for iPhone” ve “Twitter for Web” uygulamalarından meydana geldiği, dördüncü sırada ise %6’lık dilimle “Clarabridge Engage” bilgilerinin yer aldığı, kalan %2’lik dilimde ise farklı 60 platformun kullanıldığı, bu 60 platform içinde en yüksek sayıda kullanılan platformun 180 kez kullanılmış olan “Twitter for iPad” şeklinde olduğu görülmüştür. Platformlara ilişkin bu bilgi, uygulamaların ve web sayfalarının geliştirilmesine ve bu maksatla hangi platformlara öncelik verilmesi gerektiğine yönelik fikirler verebilir. Dördüncü sıradaki platformun müşteri ilişkileri yönetimi amacıyla da kullanılan bir yazılım olduğu, bu tür bilgilerin veri seti içinde kurum ve sektörde faaliyet gösteren işletmeler tarafından otomatik olarak reklam maksatlı üretilmiş içeriklerin tespit edilmesini sağlayabileceği dikkate alınmalıdır.

Kullanılan model ile 2020 takvim yılının tamamını kapsayan veri setinden elde edilen sonuçlar incelendiğinde, sektörel olarak işletmelere yön verebilecek nitelikte fikirlere ulaşılabilirdiği görülmüş ve amaca uygun sonuçlar için yapılması gereken işlemler tespit edilmiştir. Bu çerçevede bu tür ticari ürün/hizmet konusu sektörlere ilişkin Twitter veri seti analizlerinde reklam içerikli tweetlerin ayıklanması işleminin gerekli bir adım olduğu öngörülmektedir. Metin ön işlemleri aşamasında yapılacak reklam ayıklama işleminin analiz sonuçlarının yanlıtlılığını azaltabileceği, bununla birlikte reklam içerikli/kaynaklı bilgilerin analiz sonuçlarında tekrar ortaya çıkabileceği için bu aşamada son bir ayıklama işlemine ihtiyaç duyulabileceği görülmüştür.

Bu alanda, potansiyel müşterilerin diyez etiketi ve URL kullanma eğiliminin düşük olduğu, diyez etiketi ve URL bulunan tweetlerin çoğunluğunun sektördeki işletmelere ait olduğu görülmüştür. Benzer durumun diğer ticari ürün/hizmeti konu alan tweet paylaşımlarında da ortaya çıkabileceği, bu nedenle Twitter’den bu tür konularda yapılacak veri çekmede, diyez etiketi yerine doğru seçilmiş anahtar kelimelerin daha yüksek sayıda potansiyel müşteri tweetlerine ulaşılmasını sağlayabileceği öngörülmektedir.

Sonuç olarak yapılan analizlerle çevrimiçi yemek siparişi konusunda faaliyet gösteren işletmelere, faaliyetlerini ne şekilde geliştirmeleri gerektiği konusunda yön verebilecek potansiyel müşterilerden elde edilen sohbet konuları olarak; (i) Gece geç saatlerde yemek siparişi verme, yeni başlanmış bir diyetin bozulması ya da evde yemek olduğu halde sipariş verme gibi nedenlerle yaşanan pişmanlık ve kararsızlık hissi, (ii) Özellikle koronavirüs salgını nedeniyle ürün ve hizmete güven konusunda yaşanan endişe, buna bağlı olarak yemek siparişi verme davranışını terk etme eğilimi, (iii) Evinde yalnız iken yemek siparişi veren kadınların yaşadıkları kişisel güvenlik endişesi (bu endişeyi gidermek için evde yalnız olmadıklarını kuryeye hissettirmek maksadıyla, ürün teslimatı esnasında evin içine doğru seslenme hareketi sergileme), (iv) Hastane çalışanları ve hastaların hastane yemeklerini beğenmemeleri nedeniyle yemek

sipariş hizmetine yönelmeleri, (v) Aynı ortamı paylaşan kişilerin başlangıçta sadece kendileri için sipariş ettikleri yiyecekleri arkadaşları ile paylaşmak zorunda kalmaları, (vi) Kuryelere emniyetli ve güvenli çalışma koşullarının sağlanması gerektiği, kuryelerin karıştıkları trafik kazaları, (vi) Sipariş edilen ürün sunumunun beklenen şekilde gerçekleşmemesi, (vii) Çevrimiçi yemek siparişi uygulamaları üzerinden ihtiyaç sahiplerine yemek yardımı (pandemi nedeniyle), (vii) Sektörde çalışan işletmelerin indirimli/joker uygulamaları, (viii) Televizyonda yayımlanan bir yemek programının, kişileri etkileyerek (programda kullanılan yemek görselinin de katkısıyla) yemek siparişi verme isteklerinin artırması, (ix) Yemek sipariş eden müşteriye ürünün ulaşmamış olmasına rağmen, cep telefonuna “ürün teslim edildi” mesajının gönderilmesi, (xi) Verilen yemek siparişinin ilgili restorana ulaşmaması, (xii) Restorana hizmet veren kuryenin, restorana gönderilmesinde gecikme yaşanması, (xiii) Yemek fiyatlarındaki artış, (xiv) Çevrimiçi yemek sipariş hizmetinin hayatı kolaylaştırması, (xv) Yemek siparişlerindeki gecikmeler, (xvi) Sektördeki bir işletmenin yemek siparişi yanında ücretsiz kitap hediyesi kampanyasının beğenilmesi tespit edilmiştir.

İşletmeler çevrimiçi yemek siparişi konusundaki paylaşımlarda; (i) Mart, Kasım, Aralık aylarında arttığını ve Şubat, Mayıs ayları ve yaz döneminde düştüğünü, (ii) Gün içinde 12:00-22:00 saat aralıklarında yükseliş olduğunu, 10:00-12:00 saat aralığı ile saat 22:00'dan sonraki zaman dilimlerinde düşüş olduğunu dikkate alarak pazarlama, reklam politikaları ve diğer faaliyetlerini yönlendirebilirler. Çalışma dönemi içindeki tweet paylaşımlarının günlük ortalamasında pandemi döneminde, pandemi öncesi döneme göre %74 oranında artış ve pandemi dönemi boyunca paylaşımlardaki günlük ortalamanın istikrarlı bir şekilde devam ettiği gözlenmiştir.

Twitter paylaşım platformu olarak en çok kullanılan platformların sırasıyla; Android, iPhone, Web, iPad olduğu tespit edilmiştir. Bu bilgi alanda faaliyet gösteren bir işletmenin uygulama/web geliştirme çalışmalarında önceliklendirme yapmasına imkan sağlayabilir.

## 5. SONUÇ VE ÖNERİLER (CONCLUSIONS)

Her konuda sohbetin gerçekleşebildiği Twitter, sektörel olarak yeni fikirlere kaynaklık etme potansiyeline sahiptir. Söz konusu potansiyelin, geliştirilen yeni modeller ve alana özel çalışmalarla işlenmesi ve sektörde faaliyet gösteren işletmelerin hizmetine sunulması gerekir. Özellikle alana özel yapılacak çalışmalar, analizlerde kullanılan algoritma ve modellerin etkinliğini, performansını artıracaktır. Bu tür çalışmalarda modellerin etkinliğini etkileyen önemli bir nokta ise, sosyal medya ortamlarında yoğunluğunu artıran reklamların veri setleri içinden ayıklanmasının gerekliliğidir. Bu çalışmada önerilen model ile tanımlayıcı analiz araçları kullanılarak, Twitter meta verilerinden ne tür bilgiler elde edilebileceği, bu bilgilerle ürün/hizmetlerin iyileştirilmesine kaynaklık edebilecek Twitter

kullanıcıları arasındaki esas sohbet, gündem konuları ve yeni fikirlere nasıl ulaşılabileceğinin adımları alana özel bir çalışma ile gösterilmiştir. Model ile detaylı metin ön işlemlerine, doğal dil işleme metodlarına ihtiyaç duymaksızın veri setinin içeriğinin ve ürün/hizmetin olumlu ve olumsuz yönlerinden en etkili ve gündemde olanların tespit edilmesi sağlanmıştır. Modelde veri setleri içinden reklam içerikli tweetlerin, çalışma kapsamında oluşturulan alana özel reklam sözlüğü kullanılarak ayıklanması yapılmıştır. Çalışma kapsamında yemek alanına özel bir sözlük hazırlanmıştır.

Ticari ürün/hizmet konulu bu çalışmanın bir benzeri Facebook, Instagram, Youtube gibi diğer sosyal ağlarda yapılan paylaşımlardan oluşan veri setleri üzerinde yapılabilir. Bu çalışma Twitter kullanıcı profillerinde yer alan diğer meta veriler kullanılarak demografik analizlerle (cinsiyet analizi, yaş grupları, coğrafi konum bilgileri) genişletilerek Twitter'den işletmelerin kullanabileceği ne tür bilgiler elde edilebileceğine ışık tutacak çalışmalar yapılabilir. Ayrıca sosyal medyadan elde edilecek veri setleri ile işletmelerin kendi havuzlarındaki müşterilerine ilişkin veri setlerinde karşılaştırmalı analiz çalışmaları yapılarak, analiz sonuçlarında paralellik olup olmadığı test edilebilir. Söz konusu bu çalışmalara, alana özel sözlükler dahil edilerek performans ölçümleri gerçekleştirilebilir.

## KAYNAKLAR

- [1] N. Mishra, ve A.A. Singh, “Use of Twitter Data for Waste Minimisation in Beef Supply Chain”, *Ann Oper Res*, 270 (2016), 337-359, 2016.
- [2] Z. Shi, H. Rui, A. B. Whinston, “Content Sharing in a Social Broadcasting Environment: Evidence From Twitter”, *MIS Quarterly* 38(1), 123-142, 2014.
- [3] K. H. Tan, Y. Z. Zhan, G. Ji, F. Ye, C. Chang, “Harvesting big data to enhance supply chain innovation capabilities: An analytic infrastructure based on deduction graph”, *Int. J. Production Economics* 165, 223-233, 2015.
- [4] T. Nguyen, L. Zhou, V. Spiegler, P. Ieromonachou ve Y. Lin, “Big data analytics in supply chain management: A state-of-the-art literature review”, *Computers and Operations Research*, 98 (2018), 254-264, 2018.
- [5] C. G. Ralha ve C. V. Silva, “A multi-agent data mining system for cartel detection in Brazilian government procurement”, *Expert Systems with Applications*, 39(14), 11642-11656, 2012.
- [6] B. K. Chae, “Insights from hashtag #supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research”, *Int. J. Production Economics* 165 (2015), 247-259, 2015.
- [7] Y. Çakırel, “İşletmelerde Büyük Veri”, *Kırklareli Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 5(1), 52-62, 2016.
- [8] İ. S. Pantelidis, “Electronic Meal Experience: A Content Analysis of Online Restaurant Comments”, *Food Service Management, Cornell Hospitality Quarterly* 51(4), 483-491, 2010.
- [9] M. Lee ve S. Youn, “Electronic Word of Mouth (eWOM): How

- eWOM Platforms Influence Consumer Product Judgment*”, International Journal of Advertising, 28 (2009), 473-499, 2009.
- [10] D. T. Allsop, B. R. Bassett ve J. A. Hoskins, “*Word-of-Mouth Research: Principles and Applications*”, Journal of Advertising, 47(4), 398-411, 2007.
- [11] M. Çetin ve M. F. Amasyalı, “*Eğitici ve Geleneksel Terim Ağırlıklandırma Yöntemleriyle Duygu Analizi*”, In Signal Processing and Communications Applications Conference (SIU), Girne, KKTC, 2013.
- [12] J. Li., A. Vishwanath ve H. R. Rao, “*Retweeting the Fukushima nuclear radiation disaster*”, Communications of ACM, 57(1), 78-85, 2014.
- [13] M. Meral ve B. Diri, “*Twitter Üzerinde Duygu Analizi*”, **IEEE 22nd Signal Processing and Communications Applications Conference**, 978(1/14), 4799-4874, 2014.
- [14] M. Çevik, S. Öztürkcan ve N. Kasap, “*Sosyal Medya Analitiği: Twitter İçin Büyük Veri Yaklaşımı*”, **20’nci Ulusal Pazarlama Kongresi**, Eskişehir, Türkiye, 10-13 Haziran 2015.
- [15] A. S. Akgül, C. Ertano ve B. Diri, “*Twitter Verileri İle Duygu Analizi*”, Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 22(2), 106-110, 2016.
- [16] B. Kocak, İ. Polat ve C. B. Kocak, “*Twitter Kullanıcılarının Havayolu Pazarına Yönelik Duygu Kutuplarının Belirlenmesi: Bir Fikir Madenciliği Örneği*”, **Global Business Research Congress (GBRC)**, İstanbul, 26-27 Mayıs 2016,
- [17] A. Onan, “*Twitter Mesajları Üzerinde Makina Öğrenmesi Yöntemlerine Dayalı Duygu Analizi*”, Yönetim Bilişim Sistemleri Dergisi, 3(2), 1-14, 2017.
- [18] M. Albayrak, K. Topal ve V. Altıntaş, “*Sosyal Medya Üzerinde Veri Analizi: Twitter*”, Süleyman Demirel Üniversitesi İİBF Dergisi, 22(Kayfor15 Özel Sayısı), 1991-1998, 2017.
- [19] Ö. Çoban ve G. Tümöklü-Özyer, “*Twitter Duygu Analizinde Terim Ağırlıklandırma Yönteminin Etkisi*”, Pamukkale Üniversitesi Müh. Bilim Dergisi, 24(2), 283-291, 2018.
- [20] B. Taşdelen, “*Twitter’da Suriyeli Mültecilere Karşı Çevrimiçi Nefret Söylemi*”, Gümüşhane Üniversitesi Sosyal Bilimler Enstitüsü Elektronik Dergisi, 11(2), 562-575, 2020.
- [21] E. Akar, Ö. Bayar ve N. Koçtürk, “*İstanbul Sözleşmesi: Twitter Üzerinden Sözleşmeye Karşı Çıkanların Görüşlerine Yönelik Bir Söylem Analizi*”, Uluslararası İnsan Çalışmaları Dergisi, 4(7), 2021.
- [22] Y. Özasan ve S. M. Uygur, “*Negatif Ağızdan Ağıza İletişim ve Elektronik Ağızdan Ağıza İletişim (E-WOM): Yiyecek-İçecek İşletmelerine Yönelik Bir Araştırma*”, Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi, 28(3), 2014.
- [23] E. Armağan ve Y. Eskici, “*Tüketicilerin Online Yemek Servislerine Karşı Tutum, Davranış ve Satın Alma Niyetleri*”, EKEV Akademi Dergisi, ICOAEF Özel Sayısı, 2019.
- [24] M. Hingle, D. Yoon, J. Fowler, S. Kobourov, M. L. Schneider, D. Falk, R. Burd, “*Collection and Visualization of Dietary Behavior and Reasons for Eating Using Twitter*”, Journal of Medical Internet Research, (J Med Internet Res,15(6), e125, 2013.
- [25] W. Fan, M.D. Gordon, “*Unveiling the Power of Social Media Analytics*”, Communications of the ACM, 57(6), 74-81, 2014.
- [26] B. S. Park, J. Jang, C. M. Chihyung Ok, “*Analyzing Twitter to Explore Perception of Asian Restaurants*”, Journal of Hospitality and Tourism Technology, 7(4), 2016.
- [27] A. Singh, N. Shukla ve N. Mishra, “*Social Media Data Analytics to Improve Supply Chain Management in Food Industry*”, Transportation Research Part E 114, 398-415, 2018.
- [28] R. El-Khchine, A. Amar, Z.E. Guennoun, C. Bensouda ve Y. Laaroussi, “*Machine Learning for Supply Chain’s Big Data: State of the Art and Application to Social Network’s Data*”, **MATEC Web of Conferences 200**, 00015, 2018.
- [29] Y. E. Karabulut ve E. U. Küçükşille, “*Twitter Profesyonel İzleme ve Analiz Aracı*”, Teknik Bilimleri Dergisi 8(2), 17-24, 2018.
- [30] M. Baykara, U. Göktürk, “*Sosyal Medya Paylaşımlarının Duygu Analizi Yöntemiyle Sınıflandırılması*”, **UBMK’17, 2nd International Conference on Computer Science and Engineering**, Antalya, 2017.
- [31] A. Bruns ve S. Stieglitz, “*Towards More Systematic Twitter Analysis: Metrics for Tweeting Activities*”, International Journal of Social Research Methodology, 16(2), 91-108, 2013.
- [32] E. Akçetin ve H. Turgut, “*Eksik Verili İşletme Veri Tabanlarında Veri Analizi ve Veri Madenciliği Uygulamaları*”, Route Educational and Social Science Journal, 22(2), 181-199, 2015.
- [33] İnternet: Yardım Merkezi, Tweetleme <https://help.twitter.com/tr/using-twitter/how-to-tweet>, 20.07.2022.
- [34] D. Boyd ve K. Crawford, **Critical Questions For Big Data: Provocations for a Cultural, Technological and Scholarly Phenomeno**, Information, Communication, Society 15(5), 662-679, 2012.
- [35] S. Atan ve Y. Çınar (2019), **Borsa İstanbul’da Finansal Haberler ile Piyasa Değeri İlişkinin Metin Madenciliği ve Duygu (Sentiment) Analizi ile İncelenmesi**, AÜ SBF Dergisi, 74(1), 1-34, 2019.
- [36] İnternet: Şikayetler, <https://www.sikayetvar.com/yemek-sepeti?page=5>, 10.08.2019.
- [37] İnternet: Şikayetler, <https://www.sikayetvar.com/dominos-pizza/memnuniyet/siparis>, 10.08.2019.
- [38] İnternet: Şikayetler, <https://www.sikayetvar.com/hot-doner/siparis/yemek>, 10.08.2019.
- [39] İnternet: Şikayetler, <https://www.sikayetvar.com/bereket-doner>, 10.08.2019.
- [40] İnternet: Şikayetler, <https://www.sikayetvar.com/yemek-sepeti/lahmacun>, 10.08.2019.
- [41] İnternet: Şikayetler, <https://www.sikayetvar.com/hacioglu-restaurant>, 10.08.2019.
- [42] İnternet: Şikayetler, <https://www.sikayetvar.com/yemek-sepeti/firma/hamburger>, 10.08.2019.
- [43] İnternet: Şikayetler, <https://www.sikayetvar.com/yemek-sepeti/pide>, 10.08.2019.
- [44] İnternet: A. Papuççuyan, **Yemek Siparişinde Piyasaya Yeni Giren Glovo’yu Yemeksepeti ile Kıyasladık**, <https://webrazzi.com/2018/12/21/yemek-siparisinde-piyasaya-yeni-giren-glovoyu-yemeksepeti-ile-kiyasladi/>, 23 Ekim 2019.

- [45] S. Tetik, Twitter Verisi ile R'de Metin Analizi Uygulaması, Lisans Bitirme Tezi, Yıldız Teknik Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, 2017.
- [46] İnternet: <https://www.baylanajans.com.tr/sosyal-medya-pazarlamasinda-influencer-etkisi/>, 1.08.2022.

# Fikir Geliştirme Laboratuvarı: Şeffaf, Güvenilir ve Disiplinlerarası Bir Model

*Araştırma Makalesi/Research Article*

 Savaş TAKAN<sup>1\*</sup>,  Duygu ERGÜN<sup>2\*</sup>

<sup>1</sup>Yapay Zeka ve Veri Mühendisliği Bölümü, Ankara Üniversitesi, Ankara, Türkiye

<sup>2</sup>Güzel Sanatlar Ortak Dersler Bölümü, Atılım Üniversitesi, Ankara, Türkiye

[stakan@ankara.edu.tr](mailto:stakan@ankara.edu.tr), [duygu.ergun@atilim.edu.tr](mailto:duygu.ergun@atilim.edu.tr)

(Geliş/Received:05.10.2022; Kabul/Accepted:29.03.2023)

DOI: 10.17671/gazibtd.1184672

**Özet**— Disiplinlerarası çalışmalarda öğrenme ve uygulama döngüsü oldukça önemlidir. Teknoloji transfer ofisleri ve teknokentler, tam da bu öğrenme ve uygulama döngüsünü sağlamak için hayata geçirilmiş organizasyonel yapılardır. Ancak teknoloji transfer ofisleri ve teknokentler, fikir üretiminden ziyade, üretilmiş fikirlerin uygulama alanları olarak hayata geçirilmiştir. Dolayısıyla teknoloji transfer ofisleri ve teknokentlerin bir ön basamağı olarak düşünülebilecek, fikir üretimine ve geliştirilmesine odaklanan bir mekanizmaya ihtiyaç bulunmaktadır. Toplumsal ihtiyaçlar ve sorunların çözümüne yönelik yenilikçi fikirlerin üretilmesinde akademisyenler, öğrenciler, sektörel aktörler ve toplum birbirlerinden kopuk olduğundan, sistematik bir zeminde bir araya gelememektedir. Bu durum, yenilikçi fikirlerin rassal, bireysel ve yetersiz bir şekilde geliştirilmesine, fikirlerin korunamamasına ve dolayısıyla zaman ve maliyet kayıplarına neden olmaktadır. Çalışmamızda, söz konusu problemlerin çözümüne yönelik, yenilikçi fikirlerin üretilmesi, takibi ve korunmasına odaklanarak akademisyenleri, öğrencileri ve çeşitli sektörel, toplumsal aktörleri bir araya getiren, fikrin üretildiği andan projeye dönüştürülmesi sürecinde tüm bu unsurların disiplinlerarası şekilde çalışabileceği yeni bir fikir laboratuvarı modeli önerilmiştir. Önerdiğimiz model, fikirlerin korunumu, tartışılması ve takibine yönelik şeffaflığın sağlanabilmesi amacıyla, blok zincirinin veri yapısından esinlenilerek geliştirdiğimiz ve Etiket adını verdiğimiz yeni bir veri yapısı üzerine inşa edilmiştir. Geliştirdiğimiz veri yapısı sayesinde, fikirlerin korunması, geliştirilmesi ve takibinin yanı sıra mantıksal ilişkilerin yaratılabilirdiği güvenilir bir tartışma ortamı elde edilmiştir.

**Anahtar Kelimeler**— disiplinlerarası çalışmalar, bilişim iletişimi, blok zinciri, üniversite eğitim modeli, fikir laboratuvarı

## A Transparent, Reliable, and Interdisciplinary Model for the Idea Development Laboratory

**Abstract**— The learning and application cycle is essential in interdisciplinary studies. Technology transfer offices and Technopolises are organizational structures implemented precisely to ensure this learning and application cycle. However, Technology Transfer Offices and Technopolises have been implemented as application areas for ideas rather than idea generation. Therefore, there is a need for a mechanism that can be considered a preliminary step of technology transfer offices and technopolises, focusing on idea generation and development. Since academics, students, sectoral actors, and society are disconnected from each other in generating innovative ideas to solve social needs and problems, they need to unite systematically. This leads to stochastic, individual, inadequate development of innovative ideas, lack of protection, and thus loss of time and cost. In our study, we propose a new idea laboratory model for solving these problems, focusing on the generation, follow-up, and protection of innovative ideas, bringing together academics, students, and various sectoral and social actors, where all these elements can work in an interdisciplinary manner from the moment the idea is generated to its transformation into a project. The model we propose is built on a new data structure that we call Tag, inspired by the data structure of the blockchain, to ensure transparency for the protection, discussion, and follow-up of ideas. Thanks to the data structure we have developed, we have achieved a reliable discussion environment where logical relationships can be created as well as the protection, development, and tracking of ideas.

**Keywords**— interdisciplinary studies, informatics communication, blockchain, university education model, idea laboratory

## 1. GİRİŞ (INTRODUCTION)

1972'deki OECD konferansının disiplinlerarasılık için beş itici güç belirlemesinden bu yana [1], yükseköğretimdeki kurumsal yapıların disiplinlerarası araştırma ve eğitimin gelişmesinde çok önemli olduğu kabul edilmiştir. Disiplinlerarası çalışmalar, belirli disiplinler bağlamında eğitilmiş araştırmacılar için yalnızca bilişsel veya epistemolojik bir zorluk değil, aynı zamanda disiplinler arası faaliyetlerin geliştirildiği kurumlar için örgütsel bir zorluktur [2]. Üniversiteler, bu tür gelişmeler için özellikle ilginç bir ortamı temsil etmektedir; çünkü bu kurumlar aynı anda hem geleneğin koruyucuları hem de deney alanlarıdır [3]. Üniversite, disiplinler bilginin üretildiği ve yeniden üretildiği geleneksel yer olduğundan, disiplinlerarasılık tartışılırken bu gerilim belirginleşir [4]. Öte yandan, yeni bilgi üretim yollarına izin verebilecek kavramsal ve organizasyonel yenilikler üretmek için yeterli özgürlüğe ve özerkliğe sahip bir yerdir [5]. Üniversitelerin dönüşümleri hem organizasyon düzeyinde hem de araştırma düzeyinde disiplinlerarasılık gibi işbirlikçi yaklaşımları gerektirmektedir. Ancak, bu değişikliklere rehberlik edecek evrensel bir genel veya hazır model bulunmamaktadır [4]. Yine de bu işbirlikçi araştırma yaklaşımlarına içgörü sağlayan, ancak aynı zamanda hem epistemik hem de kurumsal düzeylerdeki zorlukları ve engelleri vurgulayan disiplinlerarasılık uygulamaları hakkında büyüyen bir literatür bulunmaktadır [6].

Disiplinlerarası çalışmalarda öğrenme ve uygulama döngüsü oldukça önemlidir [7]. Bu döngü, yetenekli insan kaynakları yetiştirmek, yeniliği beslemek ve rekabet gücünü artırmak adına öğrenen bir organizasyon olmayı gerektirir. Öğrenen organizasyon, yeni teknolojiyi edinir ve öğrenerek daha kullanışlı teknolojiler yaratır. Dinamik öğrenen organizasyon, hızla değişen ortamlara kolayca uyum sağlar; bir sürdürülebilirlik aracı olarak kendi geleceği için çözümler üretir [8].

Sürdürülebilir bir rekabet avantajı elde etmek için organizasyonlar, öğrenme yoluyla yenilikçi fikirler yaratarak sürekli değişiklikler ve organizasyonel stratejiler geliştirir [9], [10]. Bu gelişim vasıtasıyla, sektörel ortamlara başa çıkma becerisine dönüştürebilecek insan kaynakları sağlanabilir. Bu nedenle yenilik için mevcut temel stratejilerden biri, fikir odaklı öğrenen ve öğrenen organizasyonlardır [11-13]. Yeni fikirler, öğrenmenin sonuçlarından biri olabileceği gibi, öğrenme sürecinde kazanılan bilgi ve becerilerin işyerine aktarılmasıyla daha büyük fikirlere de aracı olabilir [14], [15]. Performansı artırmak için öğrenme sürecinde edinilen bilgi, beceri ve deneyimlerin işletmelere aktarılması gerekir [14]. Bu öğrenme transferi, öğrenen organizasyon faaliyetlerinden elde edilen deneyim veya bilgi ile örgütsel performans arasındaki ilişkiye aracılık etmede çok önemli rol oynar.

Teknoloji transfer ofisleri ve teknokentler, tam da bu öğrenme döngüsünü sağlamak için hayata geçirilmiş organizasyonel yapılardır [16]. Teknoparkların en belirgin motivasyonu, üniversite ve şirketler arasındaki iş birliğinin

geliştirilmesini teşvik etmesidir [17]. Bu yönüyle oldukça yenilikçi nitelikteki bu organizasyonlar, sektör ve akademi iş birliğinin güçlenmesi başta olmak üzere pek çok açıdan öğrenme transfer döngüsüne katkı sunar. Ancak teknoloji transfer ofisleri ve teknokentler, birer teknoloji uygulama alanı olarak kurgulanmıştır. Yani doğrudan yaratıcı fikir üretimine odaklanmaz. Daha ziyade, fikirlerin uygulama alanları olarak hayata geçirilmiştir [18]. Dolayısıyla teknoloji transfer ofisleri ve teknokentlerin bir ön basamağı olarak düşünülebilecek, yaratıcı fikir üretimine ve bu fikirlerin geliştirilmesine odaklanan bir mekanizmaya ihtiyaç bulunmaktadır. Winks'in de ifade ettiği üzere yükseköğretimde yaratıcılığı teşvik eden ve ortaya çıkaran süreçler, kültürler ve öğrencilerin, öğretim üyelerinin disiplinler arasında ve kendi bölümleri içinde fikir geliştirme yolları hakkında yeterli öneri geliştirilmemiştir [19]. Buna ek olarak teknoloji transfer ofisleri ve teknokentler, fikirlerin uygulama alanları oldukları için, öğrenen organizasyon olma sürecinde fikir üretimine yönelik bizce en fazla yatırım yapılması gereken temel insan kaynağını, yani eğitim gören kitleyi ve akademisyenleri büyük oranda devre dışı bırakır [20]. Diğer bir ifadeyle, yapısal düzeneğini bu insan kaynağı odağında inşa etmez. Teknoloji transfer ofisleri ve teknokentlerin varoluş stratejileri açısından bu kaçınılmaz ve olağandır. Dolayısıyla bu çalışmada, üniversite eğitimi gören kitlenin ve akademisyenlerin, sözü edilen teknokent ve teknoloji ofisi gibi yapılara transferinden bir adım öncesinin tasarımına odaklanılmıştır.

Mevcut durumda, oluşan ihtiyaçlar ve sorunların çözümü için yenilikçi fikir üretimi konusunda akademisyenler, öğrenciler, sektörel aktörler ve toplum birbirlerinden kopuk olduğundan [21], sistematik bir zeminde bir araya gelememektedir. Bu durum, yenilikçi fikirlerin rassal, bireysel ve yetersiz bir şekilde geliştirilmesine, fikirlerin korunamamasına ve dolayısıyla zaman ve maliyet kayıplarına yol açabilmektedir. Buna ek olarak yenilikçi fikirlerin üretilebilmesi ve geliştirilebilmesi için disiplinlerarası bir anlayış oldukça elzemdir. Toplumsal ihtiyaçlara ve problemlere yönelik fikir üretip bu fikirleri projeye dönüştürme konusunda öğrencilerin ve akademisyenlerin teşvik edilmesine ve eğitilmesine ihtiyaç bulunmaktadır. Son olarak öğrencilerin, sektörün problemlerini tanıyıp öğrenmesine ve böylece sektöre entegrasyon süresinin azaltılmasına ihtiyaç vardır.

Çalışmamızda, yukarıda değinilen problemlerin çözümüne yönelik, yenilikçi fikirlerin üretilmesi ve korunmasına odaklanarak bilim insanlarını, öğrencileri ve çeşitli sektörel, toplumsal aktörleri bir araya getiren, fikrin üretildiği andan projeye dönüştürülmesi sürecinde tüm bu unsurların disiplinlerarası şekilde çalışabileceği bir fikir laboratuvarı modeli önerilmiştir.

## 2. İLGİLİ ÇALIŞMALAR (RELATED WORK)

Çalışmamızda fikirlerin üretimi ve korunması odaklı üniversite eğitiminin bir parçası olarak hayata geçirilen ve akademi ile özel sektör iş birliğinde disiplinlerarası çalışmalarla ilerleyen bir fikir laboratuvarı modeli

önerilmiştir. Bu bağlamda ilgili çalışmalar, üniversite eğitiminde disiplinlerarası ve çok aktörlü fikir laboratuvarları çerçevesinde incelenmiştir.

Baptista ve arkadaşları, üç Latin Amerika üniversitesinde disiplinlerarası araştırma için kurumsal alanların yaratılmasına yönelik girişimlere odaklanmaktadır [4]. Makale, disiplinlerarası merkezlerin kurumsallaşma süreçlerini analiz etmektedir. Analiz, kurumsallaşma süreçleri arasında bir çeşitlilik bulunduğunu göstermektedir ve bölgede disiplinlerarasılığın kurumsallaşması için tek bir model olmadığı fikrini desteklemektedir. Lindvig çalışmasında, disiplinlerarası eğitim unsurlarının önceden var olan monodisipliner yapılar tarafından nasıl oluşturulduğu, desteklendiği ve hatta güçlendirildiğine ilişkin yolların haritasını çıkarmak için geleneksel fakülte yapılı bir üniversitede çeşitli disiplinlerarası eğitim çabalarının gelişimini incelemiştir [22]. Akabinde, tek disiplinli ortamlarda disiplinlerarası eğitim yaratmanın, tek disiplinliliği disiplinlerarasılıkla karşı karşıya getirme durumu olmadığı; üniversite yönetimi ve idaresinin yanı sıra araştırmacılar ve öğrenciler için toplam eğitim çıktısını artırmak üzere peyzajdan, mekana ve yönetime kadar tüm yapılardan en iyi şekilde yararlanmak ile ilgili olduğu sonucuna varılmıştır.

Birleşik Krallık Üniversitesi'ndeki eğitim uygulamalarında yenilik ve yaratıcılığın önündeki engelleri ve bu konudaki olanakları inceleyen bir çalışmada [19], üniversite öğretim kadrosuna sunulan disiplinlerarası ortamların daha fazla kullanılmasının, yaratıcı uygulamayı ve akran öğrenimini artırmaya katkıda bulunabileceği ortaya konmuştur. Buna ek olarak, dört tasarım ilkesi önerilmiştir: Yenilik günlük alanlarda gerçekleşir, ortak alanlara ve sosyal işlevlere ihtiyaç duyar, farklı ölçekleri bulunur ve ortak çalışma alanları kolektif bir kültür gerektirir. Araştırmacılara göre bir yenilikçilik kültürünün normalleştirilmesi, alan kullanımında hem fiziksel hem de davranışsal ayarlamalar gerektirir; bu da fakültenin akran öğrenimi için alanları yeniden tasarlamak üzere hem fakülte hem de kurumların birlikte çalışması gerektiğini düşündürür. Fakülteler ile sosyal ya da sektörel kurumların bir arada çalışmasına yönelik Groulx ve arkadaşları tarafından gerçekleştirilen farklı bir çalışmada, sürdürülebilirlik girişimleri oluşturmak için geçerli olan kapsamlı üniversite-toplum iş birlikleri kavramı incelenmiştir [23]. Sonuçlar, toplumsal/bölgesel kalkınma bağlamında, topluluk deneyiminin, eşitlik ve katılımın, esnek program tasarımının ve üniversite-topluluk ortaklıklarının, başarının anahtarı olarak öğretme ve öğrenmede risk almayı destekleyen bir kurumsal kültür için önemini vurgulamaktadır.

Barringer ve arkadaşları, bilim politikalarının onlarca yıldır disiplinlerarası araştırmayı teşvik etmesine rağmen üniversitelerin buna aynı şekilde karşılık verememesi konusundaki nedenleri açıklamak için, ülke çapındaki en iyi 156 araştırma üniversitesi için kapsamlı bir veri toplayarak, incelemelerde bulunmuştur [24]. Çalışmada, büyük miktarda metinsel veri toplandığı için (stratejik planlar ve üniversite rektörlerinin konuşmaları gibi),

disiplinlerarası etkileşimin tanınmasına yardımcı olması adına doğal dil işlemeye yönelinmiştir. Sonuçta, bu üniversitelerde idari destek ve fakülte katılımı yüksekse, disiplinlerarası çalışmalara yapısal bağlılık düzeylerinin de arttığı gözlenmiştir. Ek olarak, Tıp Fakültesi bulunan üniversitelerin, bu konuda daha başarılı olabileceği gözlenmiştir. Ancak idari destek ve fakülte katılımının koşullu doğası gereği üniversitelerin sadece akranlarını taklit etmemesi gerektiği, bunun yerine hangi stratejilerin uygulanacağını belirlemek için araştırma kapasitelerini, statülerini ve mevcut fakülte desteği ile disiplinlerarası çalışmalara katılım düzeylerinin belirlenmesi gerektiği önerilmiştir.

Cavallone ve arkadaşları çalışmalarında, özel ve kamu sektörü kuruluşlarının değer beklentilerini ve bu kuruluşların ortak değer yaratımındaki potansiyel rollerini araştırmıştır [25]. Araştırmacılar değer kavramını, akademisyenler ve uygulayıcıların, yükseköğretimde eğitim hizmetlerinin kalitesinin, etkinliğinin ve evrenselliğinin artırılması olarak ele almıştır. Araştırmada, yükseköğretim süreçlerinde öğrencilerin performansını artırmak için hem ulusal hem de uluslararası düzeyde iş paydaşları ile yaratıcı ortaklık kurma eğiliminin teşvik edilmesi gerektiği sonucuna varılmıştır. Buna ek olarak, üniversiteler ile dış ortamda faaliyet gösteren şirketler arasındaki alışverişin, değer yaratma açısından hem öğrencileri hem de iş paydaşlarını içeren bir öğrenme ekosistemi oluşturacağı belirtilmiştir. Son olarak yönetim açısından bakıldığında, üniversitelerde kalite iyileştirmesinin, yöneticilerin ortak değer yaratımına verdiği destek ile artırılabilirliği savunulmuştur.

Miller ve arkadaşlarının araştırması, disiplinlerarası bir araştırma ekibinin parçası olarak çalışan öğrenci deneyimlerine ilişkin bir vaka çalışması sunmaktadır [26]. Araştırmada, sosyal hizmet, inşaat mühendisliği ve bilgisayar bilimlerinden yüksek lisans düzeyindeki öğrencilerden oluşan bir ekip, ulaşım dezavantajının toplulukta yaşayan yaşlı yetişkinlerin ve bağımlı çocuklarla evsizlik yaşayan bekar ebeveynlerin deneyimlerini nasıl etkilediğine ilişkin verileri toplayan bir mobil cihaz uygulamasının tasarımında iş birliği yapmıştır. Sonuçlar, öğrencileri disiplinlerarası uygulamalı araştırmalara dahil etmenin, bilgiyi disiplinlerin geleneksel sınırlarının ötesine genişletmeye zorlamak, eleştirel ve yaratıcı düşünme becerilerini teşvik etmek gibi faydaları olduğunu ortaya koymuştur. Jæger, çalışmasında yükseköğretimde disiplin, disiplinlerarasılık ve disiplinlerarasılığın çerçevelemesi ve bu çerçevelemenin Avrupa yükseköğretim müfredatı üzerindeki etkilerini araştırmıştır [27]. Araştırmacı bu tür disiplinlerarası katkıların, Avrupa toplumlarının karşı karşıya olduğu karmaşık siyasi ve sosyal sorunlara gerekli yanıtlar geliştirebileceğini savunur.

Belwal ve arkadaşları çalışmalarında proje tabanlı öğrenmenin, üniversite eğitimi ile gerçek yaşamı nasıl ilişkilendirebileceğini araştırmıştır [28]. Proje tabanlı öğrenmede öğrencilerin, iletişim, özgüven, İngilizce dil becerileri, planlama, düşünme ve zaman yönetimi becerileri gibi davranışsal yetkinliklerden büyük ölçüde



yararlandıkları belirlenmiştir. Ayrıca proje tabanlı öğrenmenin, öğrencilere geleneksel öğretim ve öğrenme ortamında edindiklerinin ötesinde farklı bir beceri seti sunduğu tespit edilmiştir. Björklund ve arkadaşları çalışmalarında, yükseköğretimde işbirlikçi çabaları ilerletmenin önündeki temel kolaylaştırıcılara ve engellere ışık tutmak amacıyla Aalto Üniversitesi'nin yenilikçi platformlardan biri olan Tasarım Fabrikası Küresel Ağ'ını incelemiştir [29]. Öğrencilerden, personelden ve eğitimcilerden gelen bakış açıları da dahil olmak üzere 17 Tasarım Fabrikası ile yapılan 25 derinlemesine görüşmeye dayanan bulgular, esnekliği teşvik eden kurumsal politikaların, fiziksel bir disiplinlerarası alanı, topluluk oluşturmayı ve bu konudaki üst düzey desteği güvence altına aldığını göstermektedir. Disiplinlerarası eğitim çalışmalarının sınırlarını genişletmek için disiplinlerötesilik (transdisciplinary) geleneğinden yararlanan farklı bir çalışma, gelecekte, üniversite eğitiminin duvarların çok ötesinde gerçekleşebileceği için, yeni bir bütünlük öğrenme anlayışının önemine işaret eder [30].

Brandenburg ve arkadaşları, Avustralya'daki Federasyon Üniversitesi'nde disiplinler arası bir arası araştırma grubunun (üç ayrı okul) doğuşunu, gelişimini ve uygulamasını incelemiştir [32]. Yazarlar, araştırmacıların okullar arası ve disiplinler arası bir ortamda yaratıcı bir şekilde iş birliği yapma fırsatlarına çok değer verdiklerini, diğer yandan hibe almanın ve araştırma yayınlamanın dolaylı ve daha az değer gördüğünü bulmuştur. Bertel ve arkadaşları, Aalborg Üniversitesi (AAU) Mega projelerini, sürdürülebilir kalkınma ilkelerinin ve sistematik disiplinlerarasılığın "probleme dayalı öğrenmeye" entegre edilmesi için bir vaka çalışması gerçekleştirmiştir [30]. Bulgular, disiplinler arası mega projelerin öğrencileri sürdürülebilir kalkınmaya katılmaya motive etme potansiyelinin bulunduğunu göstermektedir. Diğer yandan araştırmacılar, öğrencilerin sahiplenmeleri, iş birliğine öncelik vermeleri ve disiplinlere olan katkıyı ve disiplinler arasındaki bağlantıyı görebilmeleri için özellikle erken aşamalarda sistematik çerçeveleme ve rehberli kolaylaştırmanın gerekliliğini vurgular.

Sürekli değişen çalışma ortamı nedeniyle, üniversitelerin "değişken, belirsiz, karmaşık ve muğlak (variable, uncertain, complex, ambiguous - VUCA)" bir çalışma ortamına uyum sağlayabilen öğrenciler yetiştirmesine duyulan ihtiyaçtan hareket eden bir çalışmada, UNIS-X deneysel öğrenme pedagojisinin etkililiği incelenmiştir [33]. UNIS-X yaklaşımı, tek bir kursta dört ilkeyi (proje tabanlı öğrenme; disiplinlerarasılık; fakülte ve dış ortaklar arasında yakın iş birliği ve aktif mentorluk) kapsamaktadır. Araştırmada, öğrencilerini bir VUCA çalışma ortamıyla başa çıkmak için beceri edindirmek amacıyla UNIS-X deneysel öğrenme pedagojisini kullanan bir üniversite incelenmiştir. Sonuçta, UNIS-X grubunun, UNIS-X olmayan gruba kıyasla problem çözme becerilerinde daha yüksek bir gelişmeye görüldüğü belirlenmiştir. Hannon ve arkadaşları, disiplinler arası müfredat ve öğretimin,

disiplin temelli bir organizasyon yapısını simgeleyen bir Avustralya üniversitesinde nasıl anlaşıldığı, uygulandığı ve desteklendiğinin yanı sıra disiplin sınırlarını aşan öğretim uygulamalarını araştırmıştır [34]. Sonuçta, disiplinlerarası bilgiyle etkileşimin, katılan öğrenciler ve öğretim personeli arasındaki akademik kültür ve kimlikler üzerinde derin etkileri olduğu ortaya konmuştur. Buna ek olarak, disiplinlerarası eğitimin koordinasyonunda ve yönetiminde önemli zorluklar ortaya çıktığı ve bu zorlukların giderilmesinde, kurumsal yapıların katkısının önemli bir faktör olduğu vurgulanmıştır.

Literatürde öne çıkan disiplinlerarası eğitim araştırmalarında genel olarak disiplinlerin birlikte çalışabilecekleri ortamların, üniversiteler ve bu kurumlardaki aktörler için değer, deneyim, yaratıcılık, iş birliği gibi konularda performansı artıracakları ortaya koyulmuştur. Buna ek olarak, sektörel aktörlerle iş birliklerinin ve proje tabanlı öğrenmenin faydalarına odaklanılmıştır. Literatürde yer alan araştırmalardan farklı olarak bizim çalışmamızda, disiplinlerarası çalışma ortamında fikirlerin üretilmesi, paylaşılması, geliştirilmesi, takibi ve korunmasına odaklanılmıştır. İlâveten, mevcut modellerden farklı olarak fikirlerin korunması, takibi ve yönetilmesi, geliştirdiğimiz yapının merkezini oluşturan bir yazılım ile güvence altına alınmıştır. Diğer bir farklılık ise geliştirdiğimiz modelde eğitim süreçleri ile disiplinlerarası çalışmaların eş zamanlı gerçekleştirilmesidir. Buna ek olarak, geliştirdiğimiz disiplinlerarası çalışma yapısı, yaşayan bir organizma gibi, öğreten, öğrenen, keşfeden, geliştiren ve dönüştüren bir mekanizma olarak tasarlanmıştır. Son olarak, diğer tüm bilimlerle her geçen gün güçlenen ilişkisinden dolayı, geliştirdiğimiz fikir laboratuvarının merkezine, bilgisayar bilimleri yerleştirilmiştir.

### 3. İLGİLİ MODELLER (RELATED MODELS)

Araştırma alanlarının çeşitliliği, disiplinlerarası uygulama yapısı ve sektörle iş birliği bakımından en dikkat çekici uygulama, MIT Medya Laboratuvarı'dır (MIT Media Lab)<sup>1</sup>. 1985 yılında MIT (Massachusetts Institute of Technology) bünyesinde kurulan Media Lab, dünyanın önde gelen araştırma ve akademik kuruluşlarından biridir. Media Lab tasarımcıları, mühendisleri, sanatçıları ve bilim insanları, geleneksel disiplinler tarafından kısıtlanmadan, insanların hayatlarını, topluluklarını ve çevrelerini anlamalarını ve dönüştürmelerini sağlayan teknolojiler ve deneyimler yaratmaya çalışır. MIT Medya Laboratuvarı, farklı ilgi ve araştırma alanlarını bir araya getiren disiplinler arası bir araştırma kültürünü desteklemektedir. Burada öğrenciler ve araştırmacılar, sosyal robotik, fiziksel ve bilişsel protezler, öğrenme için yeni modeller ve araçlar, topluluk biyomühendisliği ve sürdürülebilir şehirler için modeller gibi çeşitli disiplinlerde yüzlerce proje üzerinde birlikte çalışmaktadır. Medya Laboratuvarı'nda geliştirilen araştırma ve projeler, sürekli bir araştırma ve keşif alanı olarak yan şirketler, sergiler, performanslar ya da üye

<sup>1</sup> <https://www.media.mit.edu/about/overview/>

şirketlere teknoloji transferi şeklinde sıklıkla Laboratuvar dışında da büyük ve gelişir.

Cambridge Üniversitesi Disiplinlerarası Araştırma Merkezleri (Interdisciplinary Research Centres)<sup>2</sup>, üniversitenin araştırma politikası komitesinin açık onayına sahip, bölümler arası girişimleri desteklemek üzere kurulmuş bir merkezdir. Merkez, disiplinlerarası zorluklarla mücadele ederek ve temel tematik alanlarda ortak bir bölümler arası vizyon yaratarak, bireysel bölümlerin araştırma önceliklerini ve stratejilerini genişletir. Disiplinlerarası Araştırma Merkezi, genellikle, bir Stratejik Araştırma Girişimi tarafından altı yıllık bir süre boyunca geliştirilen hedefler ve başarılar üzerine kuruludur. Merkezde veriye dayalı keşif, kanser, kalp damar hastalıkları, biyoçeşitliliği koruma, enerji, biyoloji mühendisliği, küresel gıda güvenliği, enfeksiyon hastalıkları, dilbilim, sinirbilim, halk sağlığı ve kök hücreler gibi çalışma alanları yer almaktadır.

Oxford Üniversitesi E-Araştırma Merkezi (Oxford e-Research Centre)<sup>3</sup> Mühendislik Bilimi Bölümü'nün bir parçası olan çok disiplinli bir veri bilimi araştırma ve eğitim enstitüsüdür. Burada, akademik araştırma ve endüstriyel uygulamalar için yenilikçi dijital metodolojiler, bilgiler ve hesaplama çözümlerini araştırılmakta ve uygulanmaktadır. Merkez, akademisyenler, öğrenciler, veri bilimcileri, doktora sonrası araştırmacıları, araştırma ve toplum yararına disiplinlerarası projeler üzerinde birlikte çalışan profesyonel araştırmacılar, yazılımcılar ve bilgi mühendisleri de dahil olmak üzere 50'den fazla personel için ortam sağlar. Merkezde, iklim hizmetleri, siber güvenlik, veri yönetimi ve tekrarlanabilirlik, dijital beşerî bilimler, enerji, dijitalleştirme, nesnelerin interneti, dil modelleme, bilimsel yayıncılık ve iletişim, bilimsel hesaplama ve görsel analitik gibi çalışma alanları mevcuttur.

Max Planck Multidisipliner Bilimler Enstitüsü (Max Planck Institute for Multidisciplinary Sciences)<sup>4</sup> geniş bir araştırma yelpazesini kapsayan, böylece disiplin çeşitliliğini ve iş birliğini teşvik eden, toplamda 16 bölüm ve 25'ten fazla araştırma grubu bulunan multidisipliner bir merkezdir. Enstitüdeki konuların yelpazesi fizik ve kimyadan yapısal hücre biyolojisine, nörobilim ve biyomedikal araştırmalara kadar uzanmaktadır. Bu geniş yelpaze ile enstitü, kendisini, uluslararası araştırma ortamında eskisinden daha iyi konumlandığını belirtir. Araştırma grupları, merkezi bilimsel ve genel hizmet tesisleri aracılığıyla desteklenir. Yarı bilim insanı olmak üzere 1000'den fazla kişiyle enstitü, multidisipliner yaklaşımıyla Max Planck Topluluğu'nun en büyük organizasyonudur. Yüksek bir bilimsel standardı korumak için, uluslararası bilim insanlarından oluşan bir danışma kurulu, merkezde gerçekleştirilen araştırmaları düzenli olarak değerlendirir. Bilim insanlarının yanı sıra endüstri

ve siyaset temsilcilerinden oluşan bir mütevelli heyeti, enstitünün çalışmalarının daha geniş sosyal bağlama entegrasyonunu destekler.

2006 yılında Pekin Üniversitesi'nde (PKU) kurulan İleri Disiplinlerarası Çalışmalar Akademisi (Academy for Advanced Interdisciplinary Studies - AAIS)<sup>5</sup>, birden fazla alanın iş birliğiyle temel bilim hakkında bilinenleri genişletmeyi amaçlayan öncü bir enstitüdür. AAIS son birkaç yıldır kendini yeni disiplinlerin inşasına, öncü araştırmaların geliştirilmesine ve yetenekli genç beyinlerin bütünleşmiş bir bakış açısıyla eğitilmesine adanmıştır. AAIS'de, modern bilim ve teknolojik ilerlemelerin disiplinler arası etkileşimleri teşvik ettiği görüşünden yola çıkmıştır. PKU'da disiplinler arası araştırmalar ve eğitim için ana platform olan AAIS, üniversitenin çeşitli bölümlerinden ve fakültelerinden gelen akademik kaynaklar tarafından desteklenmektedir. AAIS, BioMed-X Araştırma Merkezi, Yaşam Bilimleri Merkezi, Kantitatif Biyoloji Merkezi, Nano Ölçekli Bilim ve Teknoloji Merkezi, Fonksiyonel Görüntüleme Merkezi, Çevre ve Sağlık Merkezi, Bilim Felsefesi ve Tarihi Merkezi dahil olmak üzere 10'dan fazla araştırma çeşitliliğine sahiptir.

Disiplinlerarası Araştırma Konsorsiyumu (The Consortium for Interdisciplinary Research - CIR)<sup>6</sup>, Berkeley Üniversitesi çatısı altında, kampüs ve ötesinde disiplinler arası araştırma ve programlamayı desteklemek ve geliştirmek üzere kurulmuştur. 2013 yılında UC Berkeley'in Edebiyat ve Bilim Fakültesi: Sanat ve Beşerî Bilimler Bölümü tarafından kurulan CIR, kampüs içine ve küresel bir izleyici kitlesine değişim ve programlama fırsatlarından yararlanarak disiplinler ve bölümler arası araştırma ve akademik faaliyet imkânı sunar. CIR şu anda Berkeley Din Araştırmaları Merkezi (BCSR), Eleştirel Teori, Beşerî Bilimler için Yeni Stratejiler, Berkeley Sanatlar ve Fikirler gibi programları desteklemektedir. BCSR yılda 30-40 etkinlik üretmekte, milyonlarca hibe fonunu yönetmekte ve yaklaşık 30 kampüs bölümü ve 200 fakülteden oluşan danışman, akademisyen ve lisansüstü öğrenciden oluşan bir topluluğu bir araya getiren akademik, sosyal yardım ve yayıncılık faaliyetlerini koordine etmektedir.

Harvard Üniversitesi bünyesinde, lisans öğrencilerini araştırmacı olmaya teşvik etmek adına çok çeşitli uygulama merkezleri bulunmaktadır<sup>7</sup>. Ancak bu merkezler multidisipliner yapıda olmadığı gibi doğrudan sektör ile entegre de değildir. Diğer yandan öğrencileri odağına alan bilebildiğimiz en temel örnek Stanford Disiplinlerarası<sup>8</sup> adıyla Stanford Üniversitesi'nde bulunmaktadır. Burada lisans düzeyinde disiplinlerarası akademik çalışma laboratuvarları bulunmaktadır. Stanford Disiplinlerarası'nda Stanford'un iş birliği kültürü, dünya, sağlık ve entelektüel yaşam için hayati önem taşıyan

<sup>2</sup> <https://www.cam.ac.uk/research/research-at-cambridge/interdisciplinary-research-centres>

<sup>3</sup> <https://oerc.ox.ac.uk/>

<sup>4</sup> <https://www.mpinat.mpg.de/643867/organization>

<sup>5</sup> <http://www.aais.pku.edu.cn/en/enabout/>

<sup>6</sup> <https://cir.berkeley.edu/>

<sup>7</sup> <https://uraf.harvard.edu/find-opportunities/browse-uraf-oppo/research>

<sup>8</sup> <https://interdisciplinary.stanford.edu/>

alanlarda yenilikçi keşifleri teşvik etmektedir. Ancak bu örnekte, sektörel aktörler geri planda bırakılmıştır. Diğer yandan, Goethe Üniversitesi Disiplinlerarası Araştırma Merkezi<sup>9</sup>, Tokyo Üniversitesi Multidisipliner Bilimler Bölümü<sup>10</sup> gibi merkezler de bulunmaktadır. Genel olarak bu multidisipliner yapılardan bazıları konsorsiyum, bazıları enstitü, bölüm ya da lisansüstü eğitim programı, diğer bazıları da araştırma merkezi olarak konumlandırılmıştır. Bölüm ya da eğitim programı olarak düşünülen disiplinlerarası yapılanmalar günümüzde her ne kadar yaygınlaşıyorsa da Rienties ve Heiot'un araştırmasına göre öğrencileri disiplinler arası gruplara ve sınıflara yerleştirmenin 'otomatik olarak' istenen disiplinler arası fikir alışverişine, disiplinler arası sosyal ağların gelişimine veya gelişmiş akademik performansa yol açmadığı ampirik olarak test edilmiş ve doğrulanmıştır [35]. Bu açıdan disiplinlerarası çalışmaların bir araştırma merkezi ya da bir laboratuvar ortamında ve bir fikir odağında geliştirilmesinin daha uygun olabileceği düşünülmektedir.

Ionian Üniversitesi İnsan ve Sosyal Bilişim Laboratuvarı (Humanistic & Social Informatics Laboratory) da multidisipliner uygulama laboratuvarı modellerine örnek verilebilir. Burada bir yanda Psikoloji, Dilbilim, Tarih gibi ilgili bilimsel ve sanatsal alanlardan türetilen hümanist ve sosyal verilerin çıkarılması, temsili ve işlenmesi için Bilgi ve İletişim Teknolojilerinin (BİT) uygulanmasına odaklanılır. Öte yandan, BİT ürün ve hizmetlerini optimum kalitede ve insan merkezli bir bakış açısıyla sunmak için bu bilimsel ve sanatsal alanlardan uyarlanmış metodolojileri kullanırlar. İnsan ve Sosyal Bilişim Laboratuvarı çalışma alanları içerisinde Kültürel Bilişim, Müzik Bilişimi, Tarih Bilişimi, Hesaplamalı Dilbilim, Oyun Yazılımı Tasarım ve Geliştirme, Multimedya, Sanal Gerçeklik, Görüntü İşleme, Bilgisayar Destekli Sistemler, Dijital Medya, Semantik Uzay Modelleme, Web Duygu Analizi ve Semantik Web, Eğitimde Bilgi işlem teknolojileri gibi araştırma ve geliştirme faaliyetleri yer almaktadır.

Ülkemizde ise bir araştırma ve uygulama laboratuvarı şeklinde olmasa da İstanbul Teknik Üniversitesi (İTÜ) çatısı altında, tüm proje önerilerinin tek bir birim tarafından değerlendirilerek geliştirildiği bir proje başvuru birimi bulunmaktadır. Söz konusu birim araştırma dekanlığınca yönetilmekte olup, her kategoride proje başvurularına açık ve bu başvuruların geliştirilmesi için hizmet vermektedir. Proje odaklı bu ortak çatının dışında İTÜ, Ortadoğu Teknik Üniversitesi (ODTÜ) ya da Boğaziçi Üniversitesi'nde, geliştirdiğimiz türden öğrenci odaklı, akademik proje üretimi ve sektörel iş birliği sağlayan bir organizasyon yapısı, bilebildiğimiz kadarıyla bulunmamaktadır. Bundan farklı olarak ülkemiz üniversitelerinde çoğunlukla her birinin kendi altyapısına uygun çeşitlilikte araştırma ve uygulama merkezleri bulunmakta, bu merkezlerde sınırlandırılmış araştırma

alanları çerçevesinde ve genellikle sektör-akademi iş birliği dışında araştırmalar yürütülmektedir.

Çalışmamızda tasarladığımız fikir laboratuvarı, değinilen örneklerden farklı olarak öğrencilere ve akademisyenlere, bağımsız çalışma (freelance) ortamı yaratarak sektörel ve kamusal aktörlerle ortaklaşa fikir ve proje geliştirme imkânı sunar. Bunun yanı sıra hem öğrencileri, akademisyenleri ve sektörel, kamusal aktörleri odağına alması hem de bu odağı araştırma ve uygulamanın yanı sıra multidisipliner iş birliklerine dahil etmesi bakımından, yukarıdaki örneklerden ayrışmaktadır. Önerimizin, yukarıda değinilen modellerden ayrıştığı diğer bir nokta, merkezine bilgisayar bilimlerini koymasındır. Bilgisayar bilimleri günümüzde sayısız pek çok disiplinin bir parçası olarak konumlanmaktadır. Bu açıdan bilgisayar bilimleri, çok çeşitli disiplinleri ilgilendiren ve gelişimine katkı sunan kritik bir alandır. Bilgisayar bilimlerinin merkezde konumlandırılması, tasarlanan araştırma ve uygulama laboratuvarının, akademik anlamda tüm disiplinlerle iş birliği içerisinde olması anlamına gelir ve bu da laboratuvarının iş birliği halinde olacağı multidisipliner yapıyı ve sektörel yelpazeyi son derece genişletir.

Özetlemek gerekirse, üniversiteler bünyesindeki uygulama laboratuvarları genellikle lisansüstü eğitime ve araştırmacı yetiştirmeye odaklıdır. Sektörle iş birliği içerisinde bulunan yapılar ise genellikle mezuniyet sonrası iş kurma ve yürütme odağında ilerler. Mevcut durumdan hareketle fikirlerin üretimini, korunmasını, yönetilmesini ve uygulama pratiğini geliştirebilecek öğrenci, akademisyen ve sektörel aktörler odaklı disiplinlerarası yaratıcı bir fikir laboratuvarı işlevi görebilecek bir yapının eksikliği dikkat çekmektedir. Geliştirdiğimiz yapı, söz konusu ihtiyacı gidermek amacıyla geliştirilmiş ve tüm bu süreçlerin rahatlıkla yönetilmesini sağlayan bir yazılım etrafında şekillendirilmiştir. Geliştirdiğimiz model ile her yıl, belirlenen stratejik bir limitte patent ve proje üretimi gerçekleştirilmesi hedeflenmektedir.

#### 4. YÖNTEM (METHOD)

Çalışmamızda, yenilikçi fikirlerin üretilmesi ve korunması amacıyla akademisyenleri, öğrencileri ve çeşitli sektörel, toplumsal aktörleri bir araya getiren, fikrin üretildiği andan projeye dönüştürülmesi sürecinde tüm bu unsurların disiplinlerarası şekilde çalışabileceği bir fikir laboratuvarı modeli önerilmiştir.

Geliştirdiğimiz modelde tüm unsurlar fikir ağını oluşturmaktadır. Üretilen ve geliştirilmesi planlanan fikirler, fikir ağında paylaşılır. Öğrenciler, akademisyenler ve sektörel, kamusal aktörlerden oluşan misafir kullanıcılar, fikir ağında hesap açabilir ve fikirlerini paylaşabilir. Paylaşılan fikirlerin projelendirilmesi ve yürütülmesi sürecinde, üst yönetimce belirlenen bir oranda öğrenci, akademisyen ve sektörel katılımcı barındırması gerekmektedir. Dolayısıyla fikir ağında paylaşım açılan

<sup>9</sup> [https://www.goethe-university-frankfurt.de/43272306/Interdisciplinary\\_Research\\_Centres](https://www.goethe-university-frankfurt.de/43272306/Interdisciplinary_Research_Centres)

<sup>10</sup> [https://www.c.u-tokyo.ac.jp/eng\\_site/info/academics/fas/dids/](https://www.c.u-tokyo.ac.jp/eng_site/info/academics/fas/dids/)

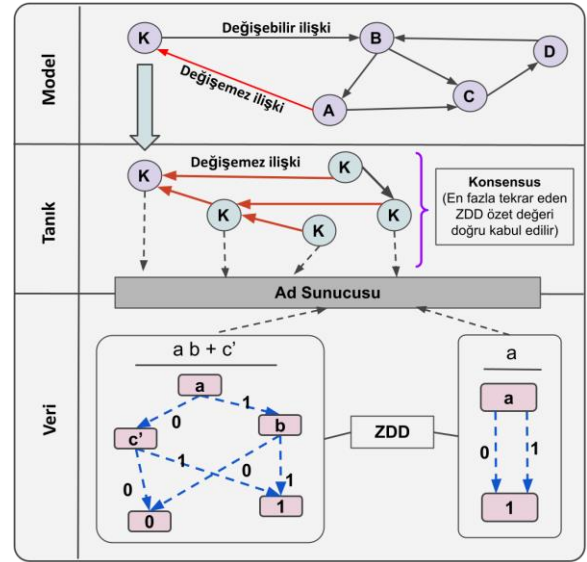
proje önerileri, bu belirli oranda ve çeşitlilikte katılımcının beğenisine ve katılımına sunulmaktadır.

Tasarladığımız model, fikirlerin korunumu, tartışılması ve takibine yönelik şeffaflığın sağlanabilmesi amacıyla, blok zincirinin veri yapısından esinlenilerek geliştirdiğimiz ve Etiket adını verdiğimiz yeni bir veri yapısı üzerine inşa edilmiştir. Geliştirdiğimiz veri yapısı sayesinde, fikirlerin korunumu ve takibinin yanı sıra mantıksal ilişkilerin yaratılabildiği güvenilir bir tartışma ortamı elde edilmiştir. Çalışmanın devamında, ilk olarak Etiket Veri Yapısı ve Etiket Ağı'na, ardından Etiket Ağı ile oluşturulan Fikir Geliştirme Modeli'ne ayrıntılı şekilde yer verilmiştir.

#### 4.1. Etiket Ağı (Tag Web)

Geliştirdiğimiz veri yapısında tüm unsurlar "Etiket" olarak kabul edilmektedir. Etiketlere ait veriler, adres olarak tutulmaktadır. Burada verinin kendisi yerine adresin tutulmasının nedeni, adresin veriden daha az yer kaplaması, tekrarı engellemesi, esnek tasarım ve kolay bakım imkânı sunmasıdır. Etiket, verinin ilişkilerini tutan değişebilir ve değişemez adreslerle birlikte değişemez verilerin adresleri bulunmaktadır. Değişebilir ilişkiye sahip olduğundan dolayı etiket ağında ekleme, yaratma, değiştirme ve silme işlemleri rahatlıkla gerçekleştirilebilmektedir. İlişkiler, her türlü ilişki yapısını modelleyebilmek amacıyla, etiketler vasıtasıyla adlandırılabilir. Bir etiketin silinmesi, değiştirilmesi, yaratılması ve eklenmesi durumunda, ona bağlı olan etiketlerin özet değerleri Algoritma 1 ile hesaplanarak güncellenir.

Geliştirdiğimiz Etiket veri yapısı, Şekil 1'de görülebileceği üzere üç katmanlıdır. Bu üç katmanın tamamına Etiket Ağı adı verilmiştir. 1. katman, değişebilir ve değişemez ilişkilerden oluşan model katmanıdır. Burada ilişkiler değişebilir ve değişemez olduğu için her çeşit yapı modellenilebilir. 2. Katman ise model katmanındaki etiketlerin kopyalarının tutulduğu ve yine hem değişemez hem de değişebilir ilişkilerin bulunduğu tanık katmanıdır. Örneğin Şekil 1'de, model katmanındaki K etiketinin, tanık katmanında dört adet kopyası bulunmaktadır. Tanık katmanında etiketlerin doğruluğu, fikir birliği algoritması ile tespit edilir. Burada kullanılan fikir birliği algoritması, çoğunlukta olan ZDD kök özet değerlerini doğru olarak kabul eder. 3. Katman ise veri katmanıdır. Veri katmanı ile etiketler arasında bir ad sunucusu yer alır. Bu ad sunucusu, ZDD kök özet değerine bakarak verinin adresini belirler.



Şekil 1. Etiket ağının genel yapısı. Burada kırmızı oklar değişemez ilişkileri, siyah oklar ise değişebilir ilişkileri ifade etmektedir. (General structure of the tag network. Here, red arrows represent immutable relationships, and black arrows represent mutable relationships.)

Şekil 1'de görülebileceği gibi, geliştirdiğimiz yapıda model ile veri birbirinden ayrılmıştır. Verinin indekslenmesi, Zero-Suppressed Decision Diagrams (ZDD) [36] kullanılarak gerçekleştirilmiştir. ZDD yapısında veriler, kanonik, kompakt ve çarpımların toplamı şeklinde ifade edilmektedir. Kanoniklik sayesinde, aynı mantıksal ifadenin özet değerleri eşit olacaktır. Şekil 1'de, K etiketinin verisi olarak  $a b + c'$  ifadesi gösterilmiştir. Burada gösterilen ifadeler, çarpımların toplamı olduğundan dolayı, diğer etiketlerle manipülasyonlar oldukça kolaydır. Örneğin aşağıdaki denklemde, A etiketi ile K etiketinin toplanması gösterilmiştir. Bu işlem ZDD'de  $O(n)$  karmaşıklığındadır.

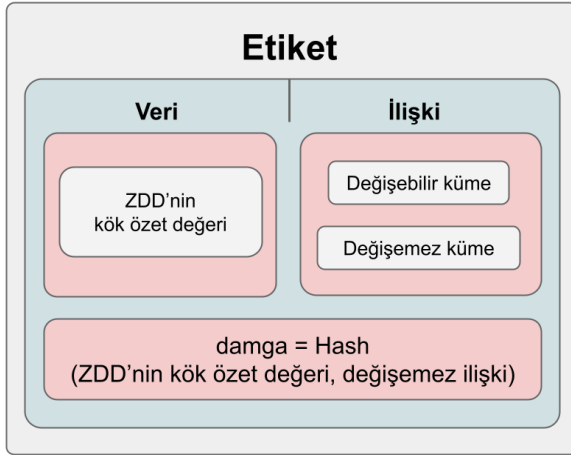
$$K = a b + c'$$

$$A = a b + d$$

$$A + K = a b + c' + d$$

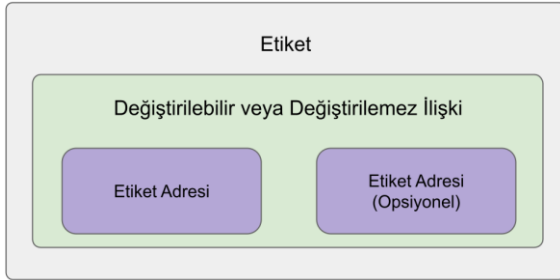
Geliştirilen modelde etiketler arasındaki ilişkiler liste olarak tutulabileceğinden, geleneksel blok zincirinin blok yapısını kapsayan bir yapı oluşur. Ayrıca ZDD'nin kök özet değeri, en uzun kök değeri karmaşıklığında hesaplanır. Hesaplama yapılırken, Algoritma 1'de gösterildiği gibi topolojik sıralama, derin öncelikli arama ve dinamik programlama kullanılır. Bunun yanı sıra Algoritma 1'de, modelin döngüsel çok geçişli çizgeyi desteklemesi için geçişler ile düğümler yer değiştirmiştir. Etiketlerin, değişebilir ve değişemez ilişkilerle organize edilmesinin en temel nedeni, tasarım kolaylığı sunmak ve döngüsel çizge gibi pek çok modelin gerçekleştirilmesini sağlamaktır. Diğer yandan blok zincirinin özetleme yapısında değiştirilebilir ilişkiler bulunmadığı için tasarımsal ve yukarıda sözü edilen verisel manipülasyonlar

oldukça zordur. Şekil 2’de, geliştirdiğimiz etiket veri yapısı genel özellikleriyle gösterilmiştir.



Şekil 2. Etiket veri yapısının genel gösterimi (General representation of the tag data structure)

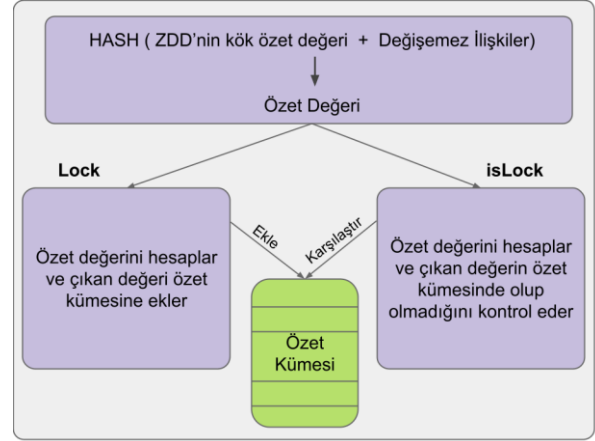
Etiketler arası ilişkiler, Şekil 3’te görülebileceği gibi, iki adet etiket adresinden oluşur. Bunlardan ilki zorunlu, ikincisi ise opsiyoneldir. Tanık katmanında ilişkiler, opsiyonel etiket almaz. Diğer katmanlarda ilişkiler en az iki etiket almak zorundadır. Bunun nedeni, model katmanındaki değişmez ilişki ile tanık katmanındaki değişmez ilişkiyi birbirinden ayırmaktır. Etiket ile veri arasındaki ilişki, Şekil 1’de görülebileceği gibi ad sunucuları tarafından organize edilir. Herhangi bir etiket, verisine ulaşmak istediğinde ZDD’nin kök özet değerini ad sunucusuna bildirir. Ad sunucusu, uygun olan veri ile etiket arasında köprü işlevi görür.



Şekil 3. İlişkilerin genel yapısı (General structure of relationships)

Modelde değişmezlik kontrolü, özet mekanizması ile sağlanmıştır. Burada özet, değişmez ilişkilerin ve ZDD kök özet değerinin özetlenmesini ifade eder. İstenildiği zaman özet hesaplanır ve özet kümesine eklenir. Geliştirdiğimiz veri yapısında bu işlem, lock() fonksiyonu ile gerçekleştirilmektedir. Daha sonra, etikette herhangi bir değişiklik olup olmadığını kontrol etmek için yeni bir özet değeri hesaplanır ve özet kümesinin içerisindeki eski özet değerleri ile karşılaştırılır. Bu andan sonra, etiketin bağlantılarında veya değişmez veride bir değişiklik gerçekleşirse, farklı özet çıkacağından, verinin değişip değişmediği ve değişmişse bulunduğu pozisyon otomatik olarak tespit edilecektir. Çıkan sonuçlar eşitse yapının değiştirilmediği; sonuçlar eşit değilse yapının değiştirildiği anlamına gelir. Bu işlem, isLock() fonksiyonu ile

gerçekleştirilmektedir. Etiketinin Lock ve islock durumunun genel yapısı Şekil 4’te gösterilmiştir.



Şekil 4. Lock ve Unlock durumu (Lock and Unlock status)

Etiket özet değerinin hesaplanması için kullanılan örnek özet bulma formülü aşağıdaki gibidir:

$$\text{hash}(\text{ZDD'nin kök özet değeri} + \sum_{i=0}^n \text{değişmez ilişki}_i)$$

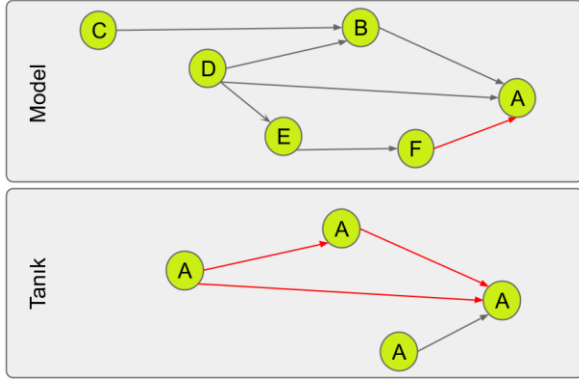
Şekil 5’te yer alan algoritma sayesinde, düşük karmaşıklık ile değişikliklerin sistemdeki tüm noktalara taşınması sağlanmaktadır. Algoritmada güncellemeler, model ve tanık katmanlarındaki çizgenin değiştirilemez ilişkileri üzerinde gerçekleşir. Bu algoritma, derinlik ilk arama, dinamik programlama ve topolojik sıralama kullanılarak geliştirilmiştir. Sistem çok geçişli bir çizge olduğundan, bütün geçişleri gezmesi için düğümlerle kenarlar yer değiştirilmiştir. Böylece tüm kenarların gezilebilmesi sağlanmıştır. Bu sayede tüm sistem O(E) karmaşıklığı ile gezilmiş olur.

```
def update(graph, parent, weight_cost, start_edge):
    stack = [start_edge]
    visited = []
    while stack:
        edge = stack.pop()
        visited.append(edge)
        for neighbor_edge in graph.neighbors(edge):
            if neighbor_edge not in visited:
                visited.append(neighbor_edge)
            if edge not in parent:
                parent[neighbor_edge] = edge
                weight_cost[neighbor_edge] = weight_cost[edge] + graph[neighbor_edge]['weight']
                stack.append(neighbor_edge)
def run(graph, start_edge):
    weight_cost = defaultdict(int, {edge: 0 if start_edge else -sys.maxsize for edge in graph.edges})
    weight_cost[start_edge] = 0
    while len(required) > 0:
        update(graph, parent, weight_cost, start_edge)
```

Şekil 5. Değişikliklerin çizge yapısında güncellenmesi için kullanılan pseudo kod (Pseudo code used to update changes in graph structure)

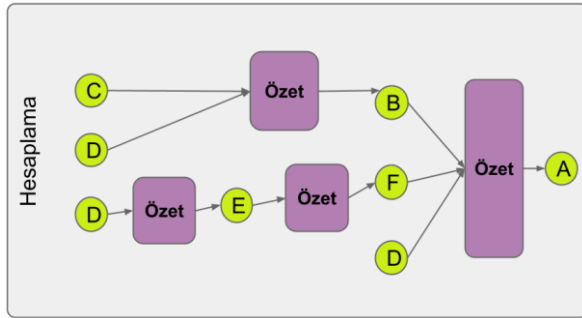
Geliştirdiğimiz modelde, güven mekanizması bulunmaktadır. Güven mekanizması, değişmez ilişkilerin toplamından meydana gelir. Güven mekanizması bu yönüyle, etiketlerin sıralanmasını sağlar. Yani, güven değeri yüksek olan etiket, daha güvenlidir ve üst sıralarda

yer alır. Diğer yandan hiçbir değişmez ilişki bulunmayan etiketler, şüpheli etiket olarak değerlendirilir. Örneğin Şekil 6'da, model katmanındaki A verisinin güven değeri, onu destekleyen değişmez ilişkilerin sayısı kadardır, yani 4'tür.



Şekil 6. Algoritma 1 ile hesaplanan A değeri (A value calculated by algorithm 1)

Model katmanında, ilişkilerin özetlenmesi ile oluşan değere ise *bağlam* adı verilmiştir. Şekil 6'daki özetleme hesaplamalarının sonucunda A'ya ulaşan değer, A'nın bağlamını oluşturmaktadır. Anlaşılacağı üzere, herhangi bir verinin yerinin değişmesi, A'nın bağlamını da değiştirecektir. Şekil 6'daki modelin bağlam değerinin hesaplanması Şekil 7'de gösterilmiştir.



Şekil 7. Şekil 6'daki modelin bağlam değerinin hesaplanması (Calculating the context value of the model in Figure 6)

Geliştirdiğimiz modelde etiketler oylanabilmektedir. Bir etiketin oyu, etiketin kendi oyu ile barındırdığı alt etiketlerin oyunun toplamına eşittir. İstenirse, etiketler toplanırken bir etki faktörü kullanılabilir. Herhangi bir etiketin oyu çok yüksek olsa dahi, değişmez ilişkilerinde başka hiçbir etiket bulunmuyorsa, güvenilirliği düşük olacaktır. Yani sistemde, güvenilirlik ve oy olmak üzere iki farklı sıralama mekanizması bulunmaktadır. Oy mekanizmasının sisteme dahil edilmesinin nedeni, tanık katmanında hiçbir etiketi bulunmayan, yani şüpheli olarak addedilen etiketlerin, kullanıcıların oyları ile sisteme entegre edilebilmesini sağlamaktır. Diğer bir nokta, sistemde kullanıcıların da bir etiket olarak kabul edilmesidir. Böylece, etiketleri sabote eden ya da bir şekilde etiket paylaşım sürecine zarar veren kullanıcıların oyları, otomatik olarak silinir ve böylece modelin kendini koruması sağlanır.

Yöntem 1:

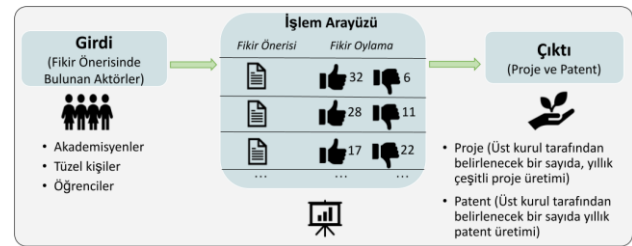
$$\square\square = \square\square\square\square\square\square\square\square + \sum_{\square=0}^{\square} \square\square\square\square\square\square\square\square\square\square\square\square * \text{Etki Faktörü}$$

Yöntem 2:

$$\square\square = \square\square\square\square\square\square\square\square + \sum_{\square=0}^{\square} \square\square\square\square\square\square\square\square\square\square\square\square$$

#### 4.2. Fikir Geliştirme Laboratuvarı (Idea Development Laboratory)

Bu çalışmada, yenilikçi fikirlerin üretilmesi ve korunması amacıyla akademisyenleri, öğrencileri ve çeşitli sektörel, toplumsal aktörleri bir araya getiren, fikrin üretildiği andan projeye dönüştürülmesi sürecinde tüm bu unsurların disiplinlerarası şekilde çalışabileceği bir fikir laboratuvarı modeli önerilmiştir. Model, şeffaflık, değiştirilemezlik ve takip edilebilirliği sağlaması dolayısıyla blok zincirinden esinlenerek geliştirdiğimiz ve Etiket adını verdiğimiz yeni bir veri yapısı üzerine inşa edilmiştir. Geliştirdiğimiz bu yapının üzerine, kullanıcı arayüzü tasarlanmıştır. Bu sayede, yaratıcı fikirlerin takibi ve korunmasına yönelik katılımın, kullanım kolaylığının, başarımın ve sürdürülebilirliğin artırılmasına çalışılmıştır. Geliştirdiğimiz modeli oluşturan tasarımlar bu bölümde ele alınmıştır. Modelin genel tasarımı Şekil 8'de gösterilmektedir.



Şekil 8. Modelin genel tasarımı (General design of the model)

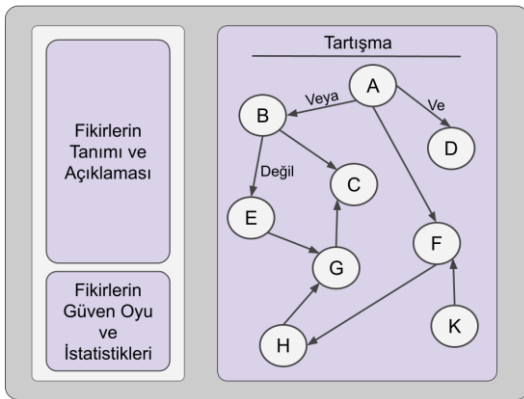
Her bir proje önerisinde, üst kurul tarafından belirlenen bir oranda öğrenci ve akademisyen bulunmak zorundadır. Etki faktörü ve beğeni oranına göre sıralanan projeler, en yüksek puanı alan ilk sırada olacak şekilde işlem arayüzünde listelenir. Öğrenciler, akademisyenler ve tüzel kişiler, bu yapı aracılığıyla hem proje önerisinde bulunabilir hem önerilen projeleri oylayabilir hem de öneriye sunulmuş projeler arasında kendilerine uygun olanları beğenip, bu projelerde yer almayı talep edebilirler. İşleme konulacak projelerde, söz konusu projeyi oylayan kişiler arasından, proje çalışanları belirlenir. Proje çalışanları, projede bulunması gereken oranlarda öğrenci, akademisyen ve tüzel kişi olacak şekilde projeyi öneren aktör tarafından belirlenir. Projelerin sorumluluğu, proje önerisinde bulunan aktör ya da aktörlere aittir. Mezuniyet koşulu olarak her bir lisans öğrencisinin, üst kurul

tarafından belirlenen sayıda projede görev almış olması gerekir. Tamamlanan her bir projede görev alanlar, projede üstlendikleri sorumluluk düzeyinde etki faktörü kazanırlar.



Şekil 9. İşlem arayüzü (Operation interface)

Önerdiğimiz modele, geliştirilebilir arayüzü sayesinde tüm katılımcılar kişisel hesapları ile giriş yapabilir ve sistemde her bir katılımcının kendine ait fikir paylaşım profili bulunur. Bu kişisel profil sayesinde katılımcılar, kendi etiketleri (fikirlere) ile ilgili her gelişmeyi, kendi hesaplarından takip edebilir. Kullanıcı sisteme giriş yaptığında, ürettiği, paylaştığı ya da takip ettiği tüm fikirleri (etiketleri) ve bu fikirlerle ilişkili diğer fikir paylaşımlarını kendi ekranında özetlenmiş bir biçimde görebilir. Her bir kullanıcı, profil sayfasında kendi fikirleriyle ya da ilgi ve yetkinlik alanlarıyla alakalı fikirleri (etiketleri) yönetebilir. Kullanıcı, kendi ilgi ve yetkinlik alanlarıyla ilgili fikir önerilerinde bulunabilir, önerilen fikirleri destekleyebilir, önerilen fikirlere karşı çıkabilir ya da alternatif fikir üretebilir. Diğer yandan, paylaşılan fikirleri oylayarak fikirlerin geliştirilmesine, projelendirilmesine katkı sunabilir ve bu projelerde yer almak için başvuruda bulunabilir. Kullanıcılar, sisteme giriş yaptıklarında genel olarak hangi fikirlerin popüler olduğunu ya da kendi fikirlerinin nasıl geri dönüşler aldığı takip edebilir.

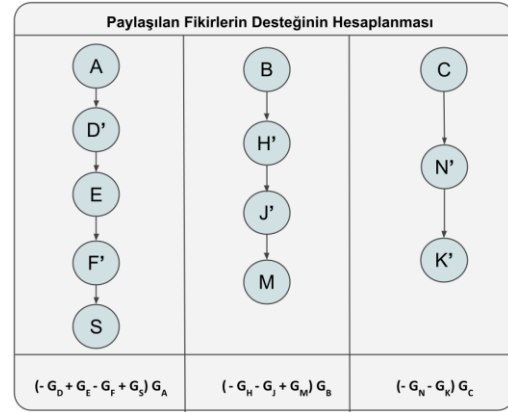


Şekil 10. Fikir laboratuvarının arayüzü (Interface of the idea lab)

Şekil 10'da görülebileceği gibi modelin arayüz ekranında, paylaşılan ve takip edilen fikirlerin gösterimi, söz konusu fikirlerin güven oyu ve istatistikleriyle birlikte tartışma bölümü yer almaktadır. Fikirlerin tanımı ve açıklaması,

proje değeri taşıdığı düşünülen ve paylaşılmak istenen fikirlerin yer aldığı kısmı ifade eder. Paylaşılan fikirlerin güven oyu, fikir paylaşım platformunda o fikrin ne kadar destek gördüğü ile ilintili olarak artar ya da azalır. Herhangi bir fikri paylaşan kişinin (etiket) bireysel güven oyu, paylaştığı fikrin güven oyuna belirli bir oranda etki eder. Bu da güven oyu yüksek olan kişilerin paylaştığı fikirlerin daha fazla etkiye sahip olması anlamına gelir. Fikirlerin, diğer fikirlerle ya da aynı fikrin alternatif versiyonları ile olan ilişkisi tartışma bölümünde yer alır. Burada fikirler arasında üç çeşit ilişki bulunmaktadır. "Ve" ilişkisi, paylaşılan fikirlerin desteklendiğini ifade eder. "Veya" ilişkisi, paylaşılan fikirlere bir alternatif getirildiğini ifade eder. "Değil" ilişkisi ise fikirlerin desteklenmediğini ifade eder. Modeldeki bu ilişkilerle kurulan fikir yapısı, ZDD'de tutulmaktadır.

Geliştirdiğimiz model bir uygulama olarak öğrenci, akademisyen ve sektörel-toplumsal aktörlerce kullanılabilir. Uygulamayı kullanan kişi ya da kurumların, kendi profil sayfaları olacak ve bu sayfada kendi uzmanlık alanları, ilgi alanları, yetkinlikleri, yetkinlik kazanmak istedikleri alanlar görülebilecektir. Üretilen fikirlerin sisteme girişi, korunması ve takibi bu profil sayfalarından yapılabilecektir. Her bir kullanıcı, sisteme giriş yaptığında, sistemde yer alan profil sayfasında kendisiyle ilgili önceki fikir ve proje geçmişini, kendi oluşturduğu ya da kendisini ilgilendiren konulardaki fikir önerilerini görebilecek ve bu fikirlerle ya da farklı fikirlerin mimarlarıyla etkileşim sağlayabilecektir. Böylece fikirler tartışılabilir, kullanıcılar arasındaki etkileşim artarak ve geliştirilerek, fikir paylaşım platformunda sosyal medya etkisi oluşacaktır.



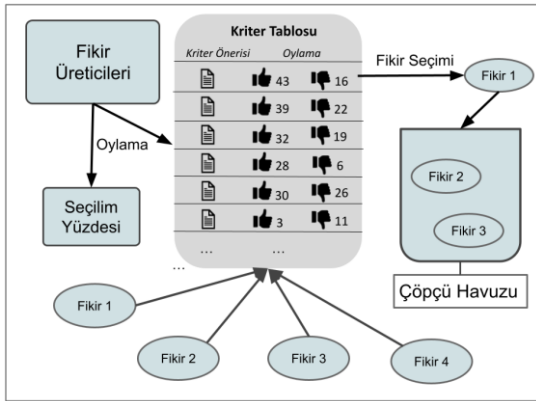
Şekil 11. Paylaşılan fikirlerin desteğinin hesaplanması (Calculation of support for shared ideas)

Şekil 11'de görülebileceği gibi ZDD'de tutulan fikirler, çarpımların toplamı şeklinde olduğu için, fikir etkileşimleri kolaylıkla belirlenebilir. ZDD yapısında fikirlerin sırası önemli değildir. Burada önemli olan, fikrin desteklenmesi, desteklenmemesi ya da bir alternatifinin üretilmesidir. Örneğin Şekil 11'de A fikri, E ve G fikirleri tarafından desteklenmiş; D ve F fikirleri tarafından yalanlanmıştır.

Durumu formalize etmek istersek, X ve K bir fikir olarak kabul edildiğinde, X fikrini geliştiren kişinin güven değerine  $G_x$  denir. Ayrıca, K fikrini destekleyen fikirler 1 değeri ile desteklemeyen fikirler ise -1 değeri ile ifade edilir. Örneğin,  $G_A (D' G_D + E G_E + F' G_F + S G_S) = G_A (- G_D + G_E - G_F + G_S)$  işleminin sonucu A fikrinin güven değerini ifade eder.

Bir fikir desteklendiğinde, o fikir tanık katmanında kopyalanır. Bu durum o fikrin güven değerini artırır. Fikri destekleyen kişi, o fikre senkron olduğu için, söz konusu fikirdeki değişikliklerin her birini takip edebilir.

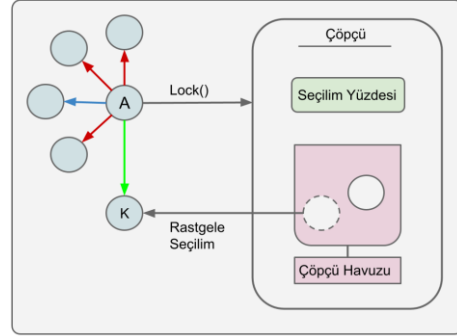
Tasarımızda, fikir çöpçüsü algoritmasına yer verilmesi düşünülmüştür. Fikir çöpçüsü algoritmasının tasarlanmasında, rastgele seçim algoritması kullanılmıştır. Fikir çöpçüsü sayesinde farklı bölgelerden, farklı alanlardan, güven oyu düşük olan fikirler de içine alacak şekilde belirli kriterlerle oluşturulmuş rastgele fikir seçimleri devreye girer. Örneğin bu sayede, platforma yeni katılan ve henüz güven oyu oluşmamış kişi ya da kurumların ürettiği fikirler, rassal olarak platformdaki diğer fikir takipçilerinin karşısına çıkarılır. Çöpçü algoritmasında, her fikir seçiminden sonra, başarı kriterlerine göre seçim fonksiyonu kendini günceller. Çöpçü algoritmasının genel tasarımı Şekil 12'deki gibidir.



Şekil 12. Çöpçü algoritması ile fikir seçiminin ve seçim yüzdesinin nasıl bulunduğu dair genel gösterimi (General representation of how the selection of ideas and the percentage of selection is found with the scavenger algorithm)

Çöpçü algoritması sayesinde, sınırlı güven oyu nedeniyle fikirlerini geniş ilişki ağına ulaştıramayan fikir üreticilerinin, rassal seçimlerle daha geniş kitlelerin karşısına çıkabilecek şekilde hareket ettirilmesi ve böylece fikir geliştirme süreçlerine dahil olabilmeleri sağlanır. Bunun için fikir üreticileri tarafından kararlaştırılacak kriterlere göre azımlık olarak (düşük güven oyuna sahip) belirlenen fikirler, kendileri ya da bir başkası tarafından çöpçü algoritmasının havuzuna kaydedilir. Her bir fikrin, fikir üreticilerinin belirlediği oranda, rassal fikirlerle ilişkilendirilme zorunluluğu bulunması önerilir. Herhangi bir fikir, eğer yeterince rassal fikir barındırmıyorsa, Lock() fonksiyonu çalıştırıldığında otomatik olarak çöpçü algoritması ile iletişime geçerek, bu algoritmanın sağladığı düşük etkili fikirleri bünyesine katar. Bu da hangi sebepten

ve hangi konuda olursa olsun etki oranı düşük kimselerin kayda değer olma ihtimali bulunan fikirlerinin belirli bir oranda görünür olmasını sağlar. Böylece, pek çok fikir ile diğeri arasında normal şartlarda bir mesafe (çizge) olmasına rağmen, sistem bazı önemli ve özel durumlarda uzak mesafedeki fikirleri, diğer kullanıcılara sunmuş olur. Çöpçü algoritmasının genel mekanizması Şekil 13'te gösterilmiştir.



Şekil 13. Çöpçü mekanizmasının genel gösterimi (General representation of the scavenger mechanism)

Geliştirilen modelde, paylaşılan fikirler düzenli olarak analiz edilebilir, teknolojik eğilimler gözlenebilir ve kapsamlı öngörüler edinilebilir. Bu öngörülerle ilgili bilgilendirmeler hızlı şekilde ilgili kurumlara ulaşabilir. Böylece, konu başlıkları ve yetkinlikler özelinde fikir üretimlerinin yönelimleri izlenebilir.

Sosyal medya etkisini de kullanan bu modelde, bilgi kirliliği ve manipülasyonu önlemeye yönelik "üretilen fikirlerin korunmasına, paylaşımına, değişimine ve takip edilebilirliğine odaklanılmıştır. Takip edilebilir ve korunabilir fikir paylaşımı sunan modelin temelinde, Blok zincirinin özetleme mekanizmasından ilham alarak geliştirdiğimiz ve Etiket adını verdiğimiz yapı kullanılmıştır. Bu sayede, fikirlerin gelişim, paylaşım ve değişim süreçleri mümkün hale gelirken; bu süreçlerin takibi ve korunması güvence altına alınmıştır. Model, fikir üretiminin yalnızca akademisyenler değil, öğrenciler ve tüzel kişilerin de katılımıyla geliştirilebileceği bir ortam sunar. Model, çok çeşitli ölçeklere uyarlanabilir ve ek olarak workshop, yarışma, kurs, kongre / konferans, dergi / yayın ve çeşitli disiplinlerarası etkinliklerle entegre edilebilir.

Verilerin korunması ve takibine yönelik her ne kadar yaygın ve güncel bir teknoloji olsa da geliştirdiğimiz modelde blok zinciri teknolojisi doğrudan kullanılmamıştır. Bunun nedeni, blok zincirinin döngüsel çizge yapısını desteklememesi, modelleme açısından bakım kolaylığı ve esneklik sunmaması ve değişebilir ilişkilere izin vermemesidir. Ancak yine de fikirlerin korunması ve takibine yönelik geliştirdiğimiz yapıya en yakın teknoloji olması dolayısıyla Etiket yapısı sonraki bölümde Blok zinciri ile karşılaştırılarak tartışılmıştır.



## 5. GEÇERLİLİĞE YÖNELİK TEHDİTLER (THREATS TO VALIDITY)

Blok zinciri pek çok katmandan meydana geldiği için, geliştirdiğimiz model, blok zincirinin tamamı yerine yalnızca blok zinciri veri yapısı ile karşılaştırılmıştır. İleride, farklı problemleri çözmek için diğer katmanlar projeye entegre edilerek, daha performanslı yaklaşımlar geliştirilmesi planlanmaktadır.

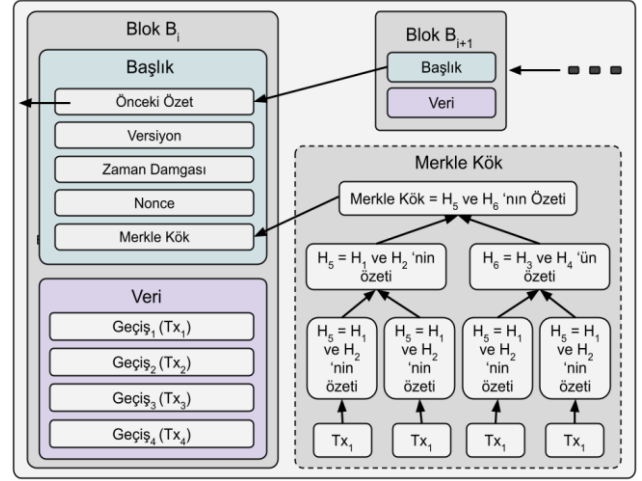
Geliştirdiğimiz modelin fiziksel olarak test edilmesi için, dört adet STM32 ve Lora Modülü kullanılmış ve bu doğrultuda verinin okunabilirliği, depolanabilirliği ve manipülasyonuna yönelik testler gerçekleştirilmiştir. Bunun sonucunda sistemin fiziksel olarak problemsiz çalışabilirliği tespit edilmiştir.

Geliştirdiğimiz çöpçü mekanizmasının testleri gerçekleştirilmemiştir. Bunun nedeni, çöpçü mekanizmasının rastgele seçim algoritması üzerine inşa edilmiş olmasıdır. Bu yapı, geçerliliği ve güvenilirliğinin bilinir olması ve teknolojik bir yenilik getirmemesi dolayısıyla bu çalışmada test edilmemiştir. Çalışmamızda, zaman ve alan açısından yenilik getiren kısımların testlerine odaklanılmıştır.

Son olarak çalışmamızda, finansal yetersizlikler nedeniyle geliştirdiğimiz modelin büyük kalabalıklar üzerinde denenmesi mümkün olmamış ve fikir birliği algoritmaları sınamamıştır. Bu nedenle, çalışmamızla ilgili olarak ileride bu tarz bir sınama gerçekleştirilmesi gerekli ve faydalı görülmektedir.

## 6. DEĞERLENDİRME (EVALUATION)

Geliştirdiğimiz model, özünde, blok zincirine bir alternatif olarak geliştirilmiştir. Çünkü senaryo bölümünde görülebileceği üzere, bazı yapıların blok zinciri altında gerçekleştirilmesi çok zordur. Aynı zamanda blok zinciri pek çok katmandan oluşan bir yapıya sahiptir. Şüphesiz her katman, bu proje ile uyumlu olmayacaktır. Bu nedenle projede, yalnızca gerekli katmanın kullanılmasına yönelinmiştir. Bu doğrultuda, blok zinciri veri yapısı ile bu çalışma kapsamında geliştirdiğimiz veri yapısı karşılaştırılarak, avantaj ve dezavantajlar ortaya konulmuştur.



Şekil 14. Standart blok zinciri veri yapısı (Standard blockchain data structure)

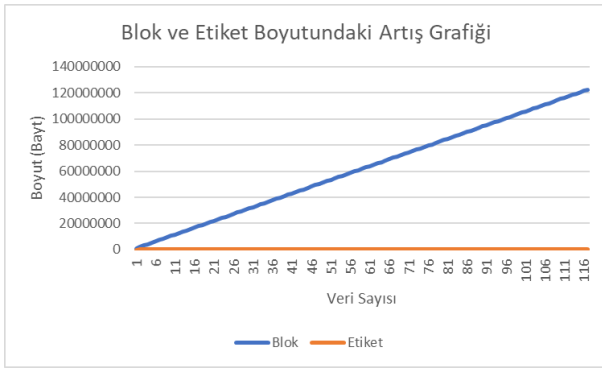
Standart bir blok zinciri yapısında, bağlı liste ve bağlı listenin içindeki her düğüm bir ağaçtan oluşur. Bu yapıda, Merkle kök değeri sayesinde verilerin değiştirilip değiştirilmediği  $O(\log n)$  karmaşıklığı ile görülebilmektedir. Bunun yanında, herhangi bir veriye ulaşmak istendiğinde ilk önce  $O(n)$  karmaşıklığı ile blok bulunur, daha sonra yine  $O(n)$  karmaşıklığı ile veriye ulaşılır. Blok yapısının veri miktarı aşağıdaki gibi formüle edilebilir:

$$\begin{aligned} \text{Blok büyüklüğü} &= \text{Önceki özet (256 Bayt)} + \text{Versiyon (256 Bayt)} \\ &+ \text{Zaman damgası (256 Bayt)} + \text{Nonce (256 Bayt)} + \text{Merkle kök (256 Bayt)} \\ &+ N * \text{Geçiş Büyüklüğü (10MB)} = 256 \text{ bayt} * 5 + 10 \text{ MB} = 10.00128 \text{ MB} \end{aligned}$$

Etiket yapısının veri miktarı ise aşağıdaki gibi formüle edilebilir:

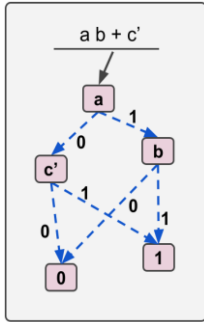
$$\text{Etiket büyüklüğü} = N * \text{Veri adresi (256 Bayt)} + \text{İlişki sayısı} * (256 \text{ Bayt}) = 512 \text{ bayt}$$

Yukarıdaki işlemlerden görülebileceği üzere verinin, modelin dışında tutulmasından dolayı modelde herhangi bir veri artışı gerçekleşmemektedir. Bu sayede yazılım mühendisliği açısından, veri manipülasyonu, tekrar problemi ve veri bakımı gibi avantajlar elde edilebilmektedir. Aynı zamanda veri ile model birbirinden ayrıldığı için modelleme daha kolay gerçekleştirilebilmektedir. Veri sayısı artarken blok ve etiketlerin boyutundaki değişim, Şekil 15'te gösterilmiştir.



Şekil 15. Veri sayısı artarken blok ve etiketlerin boyutundaki değişim (Change in size of blocks and tags as the number of data increases)

Grafik 1, önerilen veri yapısı için Etiket mekanizmasının nasıl avantaj sağladığını göstermektedir. Her bloğun veya etiketin boyutu 10 MB olarak ve her adresin boyutu 256 bayt olarak ayarlanmıştır. Bu deney düzeneğinden görüldüğü gibi, işlenecek veriler blokların toplam miktarı ile doğrusal olarak artar. Ayrıca Etiket mekanizması tarafından işlenecek veriler, geleneksel blok zincirine kıyasla neredeyse yok denecek kadar azdır.



Şekil 16. ZDD gösterimi (Zero-Suppressed Decision Diagram representation)

ZDD sayesinde herhangi bir mantıksal ilişki, düşük zaman ve alan karmaşıklığına sahiptir.  $O(n)$  karmaşıklığında iki tane ZDD birleşebilmektedir. Aynı zamanda çarpımların toplamı biçiminde yazılmış herhangi bir mantıksal ilişki, lineer karmaşıklığa sahiptir. ZDD yapısı kanonik olduğu için, kök özet değeri mantıksal ifadeye özgüdür ve farklı biçimdeki aynı mantıksal ifadeler aynı özet değerine sahiptir. Örneğin  $ab + ac$  ve  $a(b + c)$  eşitliği aynı kök özet değerine sahiptir. Bu yapı, bilebildiğimiz kadarıyla herhangi bir blok zincirinde bulunmamaktadır. Ayrıca, verileri indekslemek için B-Tree kullanılarak yaklaşık  $O(\log n)$  karmaşıklığında arama sağlanabilir. Blok zincirinde böyle bir yapı, bilebildiğimiz kadarıyla bulunmamaktadır.

Etiket mekanizmasının avantajları Çizelge 1'de gösterilmiştir. Burada veri ile model ayrı tutularak, verilere rahatlıkla erişebilmek mümkün hale gelmiştir. Ayrıca Etiket yapısı, Blok zincirinden farklı olarak, değiştirilebilir ilişki yapısını da içerisinde barındırmaktadır.

Geliştirdiğimiz modelde üç önemli unsur bulunmaktadır. Bunlardan ilki, değişebilir ilişkilerdir. İkincisi, verilerin modelden ayrı tutulmasıdır, üçüncüsü ise oluşturulan çizgede yalnızca etiketlerin bulunmasıdır.

Tablo 1. Blok ve Etiket mekanizması / Veri odaklı karşılaştırma (Block and Tag mechanism / Data driven comparison)

Tür	Veri	yedekleme	Değişme	Silme	Ekleme	Değişme
Blok	İçeride	Model ve veri	Yok	Yok	Veri + 5* adres	Yok
Etiket	Dışarıda	Model veya veri	Var	Var	2*adres	Var

Çizelge 2'de, Blok zinciri veri yapısı ile geliştirdiğimiz Etiket veri yapısı karşılaştırılmıştır. Blok zincirinde her bir blok, bağlı liste yapısında saklanır. Blok zinciri veri yapısı, bunun dışında farklı bir saklama türüne izin vermemektedir. Bağlı listedeki operasyonlar, esnek değildir ve değiştirilmezdir. Buna karşın önerdiğimiz model, çeşitli yapıları barındırabildiği için Blok zincirindeki bloklara göre esneklik ve değişebilirlik avantajlarına sahiptir. Bloklar arasındaki bağlantılar, Blok zincirinde blokların farklı konumlara taşınmasını engellerken; önerdiğimiz Etiket yapısı, çizge tabanlı olması ve esnekliği dolayısıyla, bir etiketin farklı bir konuma taşınmasında herhangi bir probleme yol açmaz. Veri taşınabilirliği aynı zamanda indeksleme ile ilintilidir. Blok zincirindeki standart bloklarda kullanılan bağlı liste yapısında bu indekslemeyi oluşturmak oldukça zor iken; önerdiğimiz Etiket yapısı, her türlü indeksleme mekanizmasını kolaylıkla içinde barındırabilecek özelliklere sahiptir.

Tablo 2. Blok ve Etiket mekanizması / Veri yapısı odaklı karşılaştırma (Block and Label mechanism / Data structure driven comparison)

Tür	Veri Yapısı	Esneklik	Taşınabilirlik	İndeks	Arama
Blok	Bağlı liste	Yok	Yok	Yok	$O(n)$
Etiket	Döngülü Çizge + ZDD	Var	Var	Var	$O(\log n)$

Çizelge 3, iki yaklaşımı güven odaklı bir perspektiften ele almaktadır. Buna göre her bloğun güveni, bir önceki ekli bloğa bağlıdır ve blokların güveni, derece veya hiyerarşik bir şekilde gerçekleştirilemeyebilir. Öte yandan, Etiketler

birbirleri arasında farklı türde ilişkiler kurabilir ve farklı Etiketler için farklı güven dereceleri bulunabilir.

Tablo 3. Blok ve Etiket mekanizması / Güven odaklı karşılaştırma (Block and Label mechanism / Reliability-oriented comparison)

Tür	Güven Bağımlılığı (blok sayısı)	Güven Bağımlılığı (Yöntem)	Bakım onarım
Block	1	Sabit	Blok düzeyinde bütünlük yok. Önceki bloğa bağlı.
Tag	n	Derece tabanlı	Gerekirse farklı Etiketler işlenebilir.

Geliştirdiğimiz modelin fiziksel olarak test edilmesi için, dört adet STM32 ve Lora Modülü kullanılmış ve testler gerçekleştirilmiştir. Konsensüs algoritmaları bu aşamada entegre edilmemiş ancak bilginin saklanması, aranması ve transferi ile ilgili istemci-sunucu yapısında testler uygulanmıştır. Sonucunda, modelin fiziksel işlerliği tespit edilmiştir. Kodlar, GitHub linkinde<sup>11</sup> mevcuttur.

## 7. SONUÇ (CONCLUSION)

Fikir üretimi, toplumsal gelişim ve küresel ihtiyaçlar konusunda kritik rol oynar ancak yaratıcı fikirlerin korunması, takibi, paylaşılması ve ilgili noktalara ulaştırılması konusunda yaygın bazı problemler bulunmaktadır. Örneğin üniversiteler bünyesindeki uygulama laboratuvarları genellikle lisansüstü eğitime ve araştırmacı yetiştirmeye odaklıdır. Sektörle iş birliği içerisinde bulunan yapılar ise genellikle mezuniyet sonrası iş kurma ve yürütme odağında ilerler. Mevcut durumda, fikirlerin üretimini, korunmasını, yönetilmesini ve uygulama pratiğini geliştirebilecek öğrenci, akademisyen ve sektörel aktörler odaklı disiplinlerarası yaratıcı bir fikir laboratuvarı işlevi görebilecek bir yapının eksikliği dikkat çekmektedir. Çalışmamızda söz konusu ihtiyaçtan hareketle, yenilikçi fikirlerin üretilmesi ve korunması amacıyla akademisyenleri, öğrencileri ve çeşitli sektörel, toplumsal aktörleri bir araya getiren, fikrin üretildiği andan projeye dönüştürülmesi sürecinde tüm bu unsurların disiplinlerarası şekilde çalışabileceği bir fikir laboratuvarı modeli önerilmiştir.

Geliştirdiğimiz yapı ile fikir paylaşımına yönelik güvenilirlik problemleri ortadan kaldırılarak, yenilikçi fikir üretimi konusunda birbirinden kopuk olabilen akademisyenler, öğrenciler, sektörel aktörler ve toplumun ortak hareket edebileceği bir zemin oluşturulmuştur. Böylelikle fikrinsel çözümlerin rassal bir şekilde

gelişmesinin, bireysel ve yetersiz kalmasının, dolayısıyla zaman ve maliyet kayıplarının ortadan kaldırılması hedeflenmiştir.

Geliştirdiğimiz model, fikir üretimine odaklanması dolayısıyla, teknoloji transfer ofisleri ve teknokentlerin ön basamağı olarak düşünülmüş ve tüm bu süreçlerin rahatlıkla yönetilmesini sağlayan bir yazılım etrafında şekillendirilmiştir. Geliştirilen modelde fikirlerin takibi ve korunması amaçlandığı için, gözlemlenebilirlik ve izlenebilirlik açısından en çok kullanılan teknoloji olan blok zincirinden esinlenilmiştir. Ancak blok zinciri teknolojisi, mantıksal ilişki kurulamaması, döngülü çizge oluşturulamaması, değişebilir ilişkileri tanımlayamaması gibi çeşitli probleme sahiptir. Blok zincirinin, geliştirdiğimiz fikir ağına entegre edilmesindeki problemleri giderebilmek için, Etiket adını verdiğimiz üç katmanlı bir veri yapısı geliştirilmiştir. Böylece, önerdiğimiz fikir ağının döngüsel çizgeyi desteklemesi, model ile verinin birbirinden ayrılması, değişebilir ilişkilerin tanımlanması, indeksleme mekanizmalarının desteklenmesi ve mantıksal ilişkilerin modele entegrasyonu sağlanmıştır. Önerdiğimiz veri yapısının işlevselliği, blok zincirin blok yapısı ile karşılaştırılarak test edilmiştir. Sonuçta, geliştirdiğimiz yapının, zaman ve alan karmaşıklığının yanı sıra sürdürülebilirlik açısından daha yüksek başarımla ortaya koyduğu tespit edilmiştir.

Geliştirdiğimiz yapının, yenilikçi fikirlerin üretilmesi ve geliştirilebilmesi için disiplinlerarası bir anlayışa ve çalışma ortamına duyulan ihtiyacı gidermesi beklenmektedir. Ayrıca toplumsal ihtiyaçlara ve problemlere yönelik fikir üretip bu fikirleri projeye dönüştürme konusunda öğrencilerin ve akademisyenlerin teşvik edilmesi ve eğitilmesine katkı sunması beklenmektedir.

Çalışmamızda önerdiğimiz yapının işlevselliğine dair geliştirdiğimiz testlere ilaveten, ileriki çalışmalarda fikir laboratuvarı modelinin bir bütün olarak pilot uygulamaların gerçekleştirilmesinin faydalı olacağı düşünülmektedir.

## KAYNAKLAR (REFERENCES)

- [1] L. Apostel, "Interdisciplinarity Problems of Teaching and Research in Universities, Organisation for Economic Cooperation and Development", Paris Centre for Educational Research and Innovation, France, 1972.
- [2] S. Veine vd., "Reflection as a core student learning activity in higher education - Insights from nearly two decades of academic development," International Journal for Academic Development, vol. 25, no. 2, pp. 147-161, 2020.
- [3] B. Stuart, "After the darkest hour... Integrity and engagement in the development of university research." **The University Research System. The Public Policies of the Home of Scientists**, Editor: Witrock, B. & Elzinga, A., Escocolmo, Almqvist & Wiksell International, 1985.

<sup>11</sup> <https://github.com/kursuApp/tag/>

- [4] B. Vienni Baptista, F. Vasen, and J. C. Villa Soto, "Interdisciplinary Centers in Latin American Universities: The Challenges of Institutionalization," *Higher Education Policy*, vol. 32, no. 3, pp. 461–483, 2019.
- [5] P. Weingart and B. Padberg, **University Experiments in Interdisciplinarity: Obstacles and Opportunities**. Transcript Verlag, 2014.
- [6] R. J. Lawrence, "Interdisciplinary science: A coming of age," *NYAS Sciences Magazine*, 18, 2016.
- [7] M. M. Hynes and W. J. Hynes, "If you build it, will they come? Student preferences for Makerspace environments in higher education," *Int. J. Technol. Des. Educ.*, vol. 28, no. 3, pp. 867–883, 2018.
- [8] D. A. Garvin, A. C. Edmondson, and F. Gino, "Is yours a learning organization?" *Harv. Bus. Rev.*, vol. 86, no. 3, pp. 109–116, 2008.
- [9] O. de Pablos Patricia and M. D. Lytras, "Competencies and human resource management: implications for organizational competitive advantage," *Journal of Knowledge Management*, vol. 12, no. 6, pp. 48–55, 2008.
- [10] H. M. Chen and W. Y. Chang, "The essence of the competence concept: Adopting an organization's sustained competitive advantage viewpoint," *Journal of Management & Organization*, vol. 16, no. 5, pp. 677–699, 2010.
- [11] K. G. Lewis, "Pathways toward improving teaching and learning in higher education: International context and background," *New Dir. Teach. Learn.*, vol. 2010, no. 122, pp. 13–23, 2010.
- [12] A. Sohel-Uz-Zaman and U. Anjalin, "Knowledge innovative organization: The effect of constant organization renewal," *Journal of Service Science and Management*, 2(04), 384, 2009.
- [13] T. Lockwood and T. Walton, **Corporate Creativity: Developing an Innovative Organization**. Simon and Schuster, 2010.
- [14] R. Wang, "Evolutionary game of knowledge sharing in master-apprentice pattern of innovative organization," *International Journal of Innovation Science*, vol. 11, no. 3, pp. 436–453, 2019.
- [15] D. Cruz-Amarán, M. Guerrero, and A. D. Hernández-Ruiz, "Changing Times at Cuban Universities: Looking into the Transition towards a Social, Entrepreneurial and Innovative Organization," *Sustain. Sci. Pract. Policy*, vol. 12, no. 6, p. 2536, 2020.
- [16] G. A. Olcay and M. Bulu, "Technoparks and technology transfer offices as drivers of an innovation economy: Lessons from İstanbul's innovation spaces", *Journal of Urban Technology*, 2016.
- [17] İ. Durak, H. M. Arslan, and Y. Özdemir, "Application of AHP–TOPSIS methods in technopark selection of technology companies: Turkish case," *Technology Analysis & Strategic Management*, 34(10), 1109–1123, 2022.
- [18] A. M. Soares, J. L. Kovalski, S. Gaia, and D. M. de G. Chiroli, "Building Sustainable Development through Technology Transfer Offices: An Approach Based on Levels of Maturity," *Sustain. Sci. Pract. Policy*, vol. 12, no. 5, p. 1795, 2020.
- [19] L. Winks, N. Green, and S. Dyer, "Nurturing innovation and creativity in educational practice: principles for supporting faculty peer learning through campus design," *Higher Education*, vol. 80, no. 1, pp. 119–135, 2020.
- [20] M. Değerli, ve M. Tolon, "Teknoloji transfer ofisleri için kritik başarı faktörleri," *Bilişim Teknolojileri Dergisi*, vol. 9, no. 2, pp. 197–220, 2016.
- [21] P. Serdyukov, "Innovation in education: what works, what doesn't, and what to do about it?" *Journal of Research in Innovative Teaching & Learning*, vol. 10, no. 1, pp. 4–33, 2017.
- [22] K. Lindvig, C. Lyall, and L. R. Meagher, "Creating interdisciplinary education within monodisciplinary structures: the art of managing interstitiality," *Studies in Higher Education*, vol. 44, no. 2, pp. 347–360, 2019.
- [23] M. Groulx, N. Nowak, K. Levy, and A. Booth, "Community needs and interests in university–community partnerships for sustainable development," *Int. J. Sustainability Higher Educ.*, vol. 22, no. 2, pp. 274–290, 2020.
- [24] S. N. Barringer, E. Leahey, and K. Salazar, "What Catalyzes Research Universities to Commit to Interdisciplinary Research?," *Res. High. Educ.*, vol. 61, no. 6, pp. 679–705, 2020.
- [25] M. Cavallone, M. V. Ciasullo, J. Douglas, and R. Palumbo, "Framing higher education quality from a business perspective: Setting the conditions for value co-creation," *Studies in Higher Education*, vol. 46, no. 6, pp. 1099–1111, 2021.
- [26] V. J. Miller, E. R. Murphy, C. Cronley, N. L. Fields, and C. Keaton, "Student experiences engaging in interdisciplinary research collaborations: A case study for social work education," *J. Soc. Work Educ.*, vol. 55, no. 4, pp. 750–766, 2019.
- [27] K. Jæger, "New-Style Higher Education: Disciplinarity, Interdisciplinarity and Transdisciplinarity in the EHEA Qualifications Framework," *Higher Education Policy*, vol. 34, no. 1, pp. 155–174, 2021.
- [28] R. Belwal, S. Belwal, A. B. Sufian, and A. Al Badi, "Project-based learning (PBL): Outcomes of students' engagement in an external consultancy project in Oman," 10.1108/et-01-2020-0006.
- [29] T. A. Björklund, T. Keipi, S. Celik, and K. Ekman, "Learning across silos: Design factories as hubs for co-creation," *European Journal of Education*, vol. 54, no. 4, pp. 552–565, 2019.
- [30] G. Kligyte, A. Buck, B. Le Hunte, S. Ullis, A. McGregor, and B. Wilson, "Re-imagining transdisciplinary education work through liminality: Creative third space in liminal times," *Aust Educ Res*, vol. 49, no. 3, pp. 617–634, 2022.
- [31] R. Brandenburg, J. Smith, A. Higgins, and J. Courvisanos, "The genesis, development and implementation of an interdisciplinary university Cross-School Research Group," *Aust Educ Res*, vol. 49, no. 3, pp. 489–510, 2022.
- [32] L. B. Bertel, M. Winther, H. W. Routhe, and A. Kolmos, "Framing and facilitating complex problem-solving competences in interdisciplinary megaprojects: an institutional strategy to educate for sustainable development," *Int. J. Sustainability Higher Educ.*, vol. 23, no. 5, pp. 1173–1191, 2022.
- [33] P.-S. Seow, G. Pan, and G. Koh, "Examining an experiential learning approach to prepare students for the volatile, uncertain, complex and ambiguous (VUCA) work environment," *The International Journal of Management Education*, vol. 17, no. 1, pp. 62–76, 2019.

- [34] J. Hannon, C. Hocking, K. Legge, and A. Lugg, “*Sustaining interdisciplinary education: Developing boundary crossing governance*,” Higher Education Research & Development, vol. 37, no. 7, pp. 1424–1438, 2018.
- [35] B. Rienties and Y. Hélot, “*Enhancing (in)formal learning ties in interdisciplinary management courses: a quasi-experimental social network study*,” Studies in Higher Education, vol. 43, no. 3, pp. 437–451, 2018.
- [36] S.-I. Minato, **Binary Decision Diagrams and Applications for VLSI CAD**. Springer, US, 2011.

# Yapay Zekânın Maliyet Etkililiğini İnceleyen Yayınların Bibliyometrik, Kelime Bulutu ve Duygu Analizi

*Araştırma Makalesi/Research Article*

 Gülçin ÇALIŞKAN,  Songül ÇINAROĞLU

Sağlık Yönetimi, Hacettepe Üniversitesi, Ankara, Türkiye  
[gulcinaliskann@gmail.com](mailto:gulcinaliskann@gmail.com), [cinaroglus@hacettepe.edu.tr](mailto:cinaroglus@hacettepe.edu.tr)  
(Geliş/Received:31.10.2022; Kabul/Accepted:29.03.2023)  
DOI: 10.17671/gazibtd.1197021

**Özet**— Bu çalışma ile, bir karar destek sistemi olarak kullanılan yapay zekânın sağlık sorunlarının tespitinde ortaya koyduğu yöntemin mevcut yöntemle göre maliyet etkililiğini tespit eden yayınların ayrıntılı olarak incelenmesi, konuyla ilgili küresel ilginin açığa çıkarılması, yayınların zaman içindeki eğilimlerinin ve hangi konuların daha çok araştırıldığı belirlenmesi amaçlanmıştır. Bununla birlikte bu çalışmanın diğer bir amacı bu yayınlarda en çok tekrar edilen kelimeleri vurgulamak ve yayınları duygu durumuna göre sınıflandırmaktır. Karar destek sistemi olarak kullanılan yapay zekânın sağladığı teşhis ya da tedavi yönteminin klasik teşhis ya da tedavi yöntemine göre maliyet etkililiği ile ilgili literatür Ağustos 2022'ye kadar Web of Science veri tabanında taranmıştır. Dışlama kriterleri uygulandıktan sonra literatür taramasında ulaşılan 24 yayın üzerinden bibliyometrik analiz, kelime bulutu ve duygu analizleri yapılmıştır. Araştırmada çok az sayıda çalışmaya ulaşıldığı ancak son yıllarda konuyla ilgili üretilen yayınların sayısında artış olduğu ve metinlerde en çok tekrar edilen anahtar kelimelerin sırasıyla yapay zekâ, maliyet etkililik, tarama ve makine öğrenimi olduğu tespit edilmiştir. Ayrıca hastalıkların teşhisinde yapay zekâ kullanılarak tanı koymanın klasik tanı koymaya göre maliyet etkililiğini tespit eden çalışmaların en fazla diş çürüğü, atriyal fibrilasyon ve diyabetik retinopati hastalıkları ile ilgili olduğu gözlenmiştir. Bununla birlikte kelime bulutunda en sık tekrar edilen kavramın “tarama” olduğu; duygu analizinde ise genel olarak pozitif duygunun daha ağır bastığı sonucuna ulaşılmıştır.

**Anahtar Kelimeler**—yapay zekâ, maliyet etkililik, diyabetik retinopati, atriyal fibrilasyon, diş çürüğü, tarama, bibliyometrik analiz

## Bibliometric, Word Cloud and Sentiment Analysis of Publications on Cost-effectiveness of Artificial Intelligence

**Abstract**—The aim of this study is to examine the publications that determine the cost-effectiveness of the current method compared to the method revealed by artificial intelligence, which is used as a decision support system in the determination of health problems, to reveal the global interest in the topic, to determine the trends of the publications over time and examine which topics are more researched. In addition, another aim of this study is to highlight the most frequently used words in these publications and to classify the publications according to their sentiment. The literature on the cost-effectiveness of the classical diagnosis or treatment method compared to the diagnosis or treatment method provided by artificial intelligence, which is used as a decision support system, was searched in the Web of Science database until August 2022. After the exclusion criteria was applied, bibliometric analysis, word cloud and sentiment analysis were performed on 24 publications reached during the review of the literature. It has been determined that there is a limited number of studies in the research, but there has been an increase in the number of publications on the topic in recent years, and the most frequently used keywords in the texts are artificial intelligence, cost-effectiveness, screening and machine learning, respectively. In addition, it was observed that the studies that determined the cost-effectiveness of diagnosing with artificial intelligence compared to diagnosing with classical method were mostly related to dental caries, atrial fibrillation and diabetic retinopathy diseases. Additionally, the most frequently used word in the word cloud is "screening"; In sentiment analysis, it was concluded that positive sentiment outweighs in general.

**Keywords**— artificial intelligence, cost-effectiveness, diabetic retinopathy, atrial fibrillation, dental caries, screening, bibliometric analysis

## 1. GİRİŞ (INTRODUCTION)

Yapay Zekâ (YZ) teknolojisi, anlama, akıl yürütme, planlama, iletişim gibi insan yeteneklerinin yazılım tarafından daha etkili, verimli ve düşük maliyetle üstlenilmesini sağladığı için önemlidir (Strusani ve Hougbonon, 2019). YZ teknolojileri, tüm otomatik sistemler gibi farklı türdeki teknolojik yeniliklerin işleyişi için temel oluşturmuştur. Yaygın kullanım alanlarından biri olan sağlık hizmetlerinin yanı sıra YZ, hava tahmini, robotların kullanımı, lojistik, inşaat mühendisliği ve işletme gibi alanlarda kullanılmaktadır (Buntak vd., 2021; Uzun, 2020; Borana, 2016).

YZ sağlık alanında uygun eylemlerin ve tedavilerin her hastaya göre uyarlanmasını sağlayarak klinik karar vermeyi destekler (Suri, 2022). Bununla birlikte, YZ maliyetleri azaltılabilir ve hastalara ve tıp uzmanlarına bakım noktasında ve gerçek zamanlı olarak daha hızlı, daha tutarlı bilgi sağlayabilir. Ayrıca YZ bazı hastaların uzmanlara gitme ihtiyacını ortadan kaldırarak klinisyen kapasitesini ve hasta erişimini geliştirebilir (GAO, 2022). YZ, hayat kurtaran ilaçların yaratılmasını hızlandırma potansiyeline sahiptir ve böylece sağlıklı ekosistemlerin sürdürülmesine yatırım yapılabilecek milyarlarca dolar tasarruf sağlar (Jena, 2022).

YZ teknolojilerinin kullanımı hayatımızın her alanında artmasına rağmen YZ söz konusu olduğunda, yöneticiler bunları sınırlı kapsamlı ve kısa vadeli projeler olarak düşünmektedir. Ayrıca karar vericiler YZ'yi geliştirme ve uygulamanın yüksek maliyetli olduğunu düşünerek kurumlar için YZ'nin doğru bir tercih olduğu konusunda şüphe duymaktadır. (Appen, 2022). Hastalıklara tanı koyarken ya da hastalıkları tedavi ederken YZ teknolojilerini tercih etme noktasında şüphelerin ortadan kalkması gerekir. Bunun yolu ise YZ ile geliştirilen sağlık teknolojilerinin mevcut yöneme göre maliyetin etkili olup olmadığını ortaya koymaktır. Maliyet etkililik analizi, kaynakların etkin kullanılması noktasında fayda sağlarken aynı zamanda maliyeti daha uygun olan yöntemin seçilmesini sağlamaktadır. Bu sebeple karar vericiler için YZ teknolojilerinin mevcut uygulamaya göre maliyet etkili olup olmadığını tespit etmek YZ teknolojilerinin kullanımının yaygınlaşması için önem arz etmektedir (Davenport ve Kalakota, 2019).

YZ teknolojilerinin kullanımı arttıkça konuyla ilgili yapılan bilimsel yayın sayıları da artmaktadır. Ancak genellikle literatürde yapılan çalışmaların içerikleri sağlık alanında YZ'nin önemine, sağlık sorunlarını tespit etmede YZ algoritmalarının geliştirilmesi ve uygulanmasına yöneliktir. Hastalıklara tanı koymada YZ teknolojilerinin sağladığı marjinal iyileştirmelerin sonuçlarda marjinal bir iyileşmeye dönüşebileceği ve YZ'nin maliyetleri azaltılabileceği düşünülmektedir (Rossi vd., 2022). Bu noktada sağlık sorunlarını tespit etmede YZ teknolojilerinin mevcut yöneme göre maliyet etkililiğini ortaya koyan çalışmalara ihtiyaç duyulmaktadır. Ancak sağlık alanında kullanılan YZ teknolojilerinin mevcut yöntem ile kıyaslamalı olarak maliyet etkililiğini analiz

eden kısıtlı sayıda çalışma bulunmaktadır. Sağlık sektöründe YZ teknolojilerinin kullanımının yaygınlaştırılması için YZ'nin mevcut uygulamaya göre maliyetinin daha uygun ve etkili olması gerekmektedir. Sağlık sorunlarını tespit etmede kullanılan YZ teknolojilerinin maliyetleri azaltması, sonuçları iyileştirmesi ve mevcut kaynaklar üzerindeki etkisini artırmış olması sağlık alanında YZ'nin önemini ve kullanımını artırmaktadır (Rossi, 2022; Huang, 2022; Xie, 2019). Ancak sağlık alanında kullanılan YZ teknolojilerinin maliyetleri azalttığı ve sonuçları iyileştirdiği ile ilgili daha fazla kanıtı ihtiyaç duyulmaktadır. Bu nedenle gelecekteki çalışmalar, YZ'nin uygulandığı ve uygulanmadığı durumları kaynak kullanımında verimlilik ve teşhis sürecinde farklılık sağlanan faydalar bakımından karşılaştırmalı olarak incelemelidir. Ayrıca hizmet sunucuların yeni teknolojileri benimsemesine ilişkin bilinçli ve rasyonel kararlar vermelerine yardımcı olacaktır.

Bu sebeple bu araştırmanın amacı; literatürde sağlık sorunlarının tespitinde karar destek sistemi olarak kullanılan yapay zekâ teknolojilerinin sağladığı yöntemin mevcut yöneme göre maliyet etkililiğini tespit eden çalışmaları ayrıntılı olarak inceleyerek gelecek araştırma eğilimlerine katkı sağlamaktır. Bu amaçla ilk aşamada maliyet etkililik analizi ile YZ kavramlarından ve bunların öneminden bahsedilecektir. İkinci aşamada araştırmada kullanılan bibliyometrik analiz, kelime bulutu ve duygu analiz yöntemleri hakkında literatür incelenecek ardından araştırmanın yöntemi ile araştırma bulguları sunulacaktır. Çalışmanın son bölümünde ise tartışma ve sonuç bölümlerine yer verilecektir.

### 1.1. Maliyet Etkililik Analizi ve Yapay Zekâ Teknolojilerinin Sağlık Sonuçlarının İyileştirilmesinde Karar Destek Sistemi Olarak Kullanımı (The Cost Effectiveness Analysis and Use of Artificial Intelligence Technologies as a Decision Support System in Improving Health Outcomes)

Maliyet-etkililik analizi (MEA), farklı müdahalelerin göreceli maliyetlerini ve sonuçlarını (etkilerini) karşılaştıran bir ekonomik değerlendirme yöntemidir (Drummond vd., 2015). MEA, karar vermeye yardımcı bir araç olmakla birlikte kaynakları etkisiz müdahalelerden etkili müdahalelere tahsis etmekte ve daha uygun maliyetli olan müdahalelere yönlendirmektedir. MEA, sağlık hizmetleri alanında sıklıkla kullanılmakta ve nispeten ucuz olan ancak hastalık yükünü önemli ölçüde azaltma potansiyeline sahip müdahaleleri vurgulayarak, ihmal edilen fırsatların belirlenmesine yardımcı olmaktadır (Shillcutt vd., 2009). Bu analizler, yöneticilerin belirsizlikleri yönetmelerine yardımcı olmanın yanı sıra maliyetleri ve sonuçları değerlendirmek için teorik bir temele sahip tutarlı ve açık bir yaklaşım sağlar. Sonuç olarak, MEA ile politika yapımcılar en az maliyetle en iyi sağlık sonuçlarını verecek en iyi tanı ve tedavileri seçebilecektir (Drummond, 2019).

YZ, bir problem çözmeyi sağlamak için bilgisayar bilimi ve sağlam veri kümelerini birleştiren bir alan olarak tanımlanmaktadır. YZ, insan beyninin nasıl düşündüğünü

ve insanların bir sorunu çözmeye çalışırken nasıl öğrenip karar verdiğini incelemektedir. YZ bu incelemeler sonucu elde ettiği sonuçları akıllı yazılım ve sistemler geliştirmenin temeli olarak kullanılmaktadır. YZ, Makine Öğrenimi (MÖ) tekniklerinin ve Derin Öğrenmenin (DÖ) bir birleşimidir (Mueller ve Massaron, 2018). YZ, üretkenlikte artış, zaman ve maliyet verimliliği, insan hatalarının azaltılması, daha hızlı iş kararları, müşteri tercihi tahmini ve satışların en yüksek seviyeye getirilmesi gibi bazı temel avantajlara sahiptir (Soni vd., 2020). YZ endüstrisinde sağlık, en hızlı büyüyen alt sektörlerden biridir ve önümüzdeki beş yıl içinde YZ'nin sağlık alanında dünya çapında 10 milyar dolarlık gelire ulaşacağı tahmin edilmektedir (Collier vd., 2017). YZ, günümüzde sadece sağlık alanında değil, finans, sosyal medya, imalat, tarım, eğitim, e-ticaret gibi birçok alanda etkisini göstermekte ve kullanımı artmaktadır (Bezboruah ve Bora, 2020). Bu hızlı büyüme, yeni teknolojilerin gelişmesinden kaynaklanmaktadır. YZ'yi benimseyen yöneticilerin çoğu, YZ'nin kullanıldığı iş alanlarında gelir artışı sağladığını ve maliyetleri düşürdüğünü belirtmektedir (McKinsey, 2019).

Yeni teknolojiler, sağlık hizmetlerinde çeşitli uygulamalar aracılığıyla uzun vadeli maliyet tasarrufu sağlarken sonuçları iyileştirme potansiyeline de sahiptir (Piccinini, 2019). YZ aracılığıyla sağlanan klinik destek, tıbbi görüntüleme tanı süreçlerini güçlendirecektir. Ayrıca, hastane iş akışları için YZ çözümlerinin kullanılması, hizmet sunumunu iyileştirecektir. Genel olarak baktığımızda, YZ'nin tedavi maliyetlerini %50'ye kadar azalttığı, sonuçları ise %30-40 oranında iyileştirme potansiyeline sahip olduğu gözlenmiştir (Frost & Sullivan, 2016). Sağlık alanında ilerleme yalnızca potansiyel artan etkinliğe değil, aynı zamanda kaynakların ve yeni teknolojilerin maliyet etkin kullanımına da bağlıdır. Sağlık ekonomisi açısından yeni teknolojilerin uygulanmasıyla ilgili olarak, özellikle uzun vadeli maliyetler ve sonuçlarla ilgili daha fazla çalışmaya ihtiyaç duyulmaktadır (Piccinini, 2019).

### 1.2. Bibliyometrik Analiz, Kelime Bulutu ve Duygu Analizi (Bibliometric, Word Cloud and Sentiment Analysis)

Mevcut çalışmada kullanılan bibliyometri, farklı belgelerin alıntılarını temsil eden bir ağ şeklinde, atıf yapılan bilimsel yayınların nicel olarak analizidir. Bibliyometrik analiz, bir araştırma konusuyla ilgili zengin literatürün makroskobik bir genel görünümünü sağlamakta ve zaman içinde etkili çalışmaları, yazarları, dergileri, kuruluşları ve ülkeleri belirlemek için kullanılmaktadır (van Eck ve Waltman, 2010). Bununla birlikte, bibliyometri, bir dizi araştırmacının, belirli bir araştırma alanındaki makalelerinin etkisini kapsamlı bir şekilde araştırmak için de kullanılmaktadır (Fan vd., 2020).

Bibliyometrik analiz, okumaya başlamadan önce literatürü inceleyerek araştırmacının çalışmasına nasıl yön vereceğini belirlemede, en etkili çalışmaları göstermekte ve bazı karşılaştırmalar yapmak için yararlı olabilmektedir (Zupic ve Čater, 2015). Bibliyometrik

analiz performans analizi ve bilimsel alan haritalama olmak üzere iki şekilde yapılmaktadır. Performans analizi, bireylerin ve kurumların araştırmalarını ve yayınlarını kaynakçadaki veriler temelinde değerlendirmeyi hedefleyen bir analiz tekniğidir. Performans analizinde derlenen çalışmalar üzerinden atıf analizi ve derlenen çalışmaların kaynakçaları üzerinden atıf analizi şeklinde iki teknik kullanılmaktadır (Al vd., 2012). Atıf analizi, araştırılan konuyla ilgili derlenen makalelere ne sıklıkla atıf yapıldığını belirleyerek araştırmacının sonraki araştırmaları ne kadar etkilediğini göstermektedir. Ayrıca atıf analizi, hangi dergilerin, kuruluşların ve hatta ülkelerin farklı araştırma alanlarında yüksek etkiye sahip olduğunu da ortaya koymaktadır (McBurney ve Novak, 2002). Bir araştırma alanından çıkarılan bilginin, ilişki ağlarına dayalı olarak analizi ve görselleştirilmesini sağlayan bilimsel alan haritalama analizinde ise ortak-atıf, kaynakça eşleşmesi, ortak-yazar analizi, ortak-kelime analizi gibi teknikler kullanılmaktadır (Noyons vd. 1999; Noyons, 2001; Kurutkan ve Orhan, 2018). Kaynakça eşleşmesi, farklı iki kaynaktan aynı yayına atıf yapılmasını; ortak-atıf analizi, bir kaynaktan farklı iki yayına atıf yapılmasını; ortak kelime analizi, dokümanlarda yer alan anahtar kelimeler, özet ya da tam metindeki aynı kelimelerin analiz edilmesini; ortak-yazar analizi ise iki ya da daha fazla yazarın dahil olduğu çok yazarlı çalışmaların analiz edilmesini ifade etmektedir (Al vd., 2012; Gulmez vd., 2020).

Bu çalışmada sağlık alanında YZ teknolojilerinin sağladığı yöntemin mevcut yöntemlere göre maliyet etkililiğini araştıran yayınlarda en çok tekrar edilen kelimeleri ortaya koymak amacıyla kullanılan tekniklerden birisi de kelime bulutudur. Kelime bulutu, kelime sıklığının görsel olarak sunulmuş halidir. Bir terim, analiz edilen metinde ne kadar yaygın olarak görünürse, oluşturulan görüntüde kelime o kadar büyük görünmektedir. Kelime bulutları, yazılı materyalin odağını belirlediği için kelime bulutlarının kullanımı artmaktadır. Kelime bulutlarının siyaset, iş dünyası, eğitim sağlık ve teknoloji gibi birçok alanda kullanıldığı görülmektedir (Atenstaedt, 2012). Bir kelime bulutu ilgili araştırmada en sık kullanılan kelimeleri dikkate almakta ve bunları sıklıklara bağlı olarak farklı boyut ve renklerde anahtar kelimeleri tanımlayan çekici bir görsel halinde sunmaktadır. Bir araştırmacı bir kelime bulutundaki sıklıkları inceleyerek metin verilerinde kullanılan belirli kelime ve ifade kalıplarını veya bunların eksikliğini ortaya koyabilir. Bununla birlikte kelime bulutları, denemeler, kısa cevaplar, anket veya fikir sorularına verilen yazılı cevaplar dahil olmak üzere her türlü metin verisini analiz etmede fayda sağlamaktadır (Depaolo ve Wilkinson, 2014).

Bu çalışmada sağlık alanında YZ teknolojilerinin sağladığı yöntemin mevcut yöntemlere göre maliyet etkililiğini araştıran yayınların duygu durumunu belirlemek amacıyla kullanılan bir diğer teknik ise duygu analizidir. Fikir madenciliği olarak da adlandırılan duygu analizi, insanların ürünler, hizmetler, organizasyonlar, bireyler, sorunlar, olaylar, konular ve bunların nitelikleri



gibi varlıklara yönelik fikirlerini, değerlendirmelerini, tutumlarını ve duygularını analiz eden çalışma alanıdır (Liu, 2012). Duygu analizi, yazılı metinde ifade edilen (olumlu veya olumsuz) görüşün gücünü bulmamıza yardımcı olan bir analizdir (Reis, 2015; Godbole, 2007). Duygu analizi doğal dil işlemede en aktif araştırma alanlarından biri olmakla birlikte veri madenciliği, web madenciliği ve metin madenciliği alanlarında da yaygın olarak kullanılmaktadır. Duygu analizi kullanımının artmasında, araştırmalar, forum tartışmaları, bloglar, mikro bloglar, Twitter ve sosyal ağların büyümesi etkili olmaktadır. Fikirler neredeyse tüm insan faaliyetlerinin merkezinde yer aldığı ve davranışlarımızın kilit unsurları olduğu için, hemen hemen her iş alanında ve sosyal alanda duygu analizi yapılmaktadır (Liu, 2012).

Duygu sözcüklerinin manuel olarak belirlenmesi zaman alan bir süreçtir. Duygu analizi sürecini otomatikleştirmek için kullanılan iki popüler yaklaşım vardır. İlk yaklaşım, sözlük tabanlı bir yaklaşımken ikinci yaklaşım, MÖ yaklaşımlarına dayanmaktadır. Sözlük tabanlı yöntem, fikir sözcükleriyle duygu sözlüğü kullanmakta ve kutupluluğu belirlemek için bunları verilerle eşleştirmektedir. Bu yaklaşım sözlükte yer alan kelimelerin ne kadar pozitif, negatif ve nötr olduğunu açıklayan görüş kelimelerine duygu puanları atamaktadır. İkinci yaklaşım olan MÖ tabanlı yaklaşım ise metni analiz etmek için sınıflandırma tekniğini kullanmaktadır (Aung ve Myo, 2017). MÖ tabanlı yaklaşımda sınıflandırma modeli oluşturabilmek için geniş açıklamalı bir metin olması gerekmektedir. Ayrıca modelin doğruluğu ek olarak yapılan açıklamanın kalitesine bağlı olmakta ve eğitim süreci çok zaman almaktadır. Bununla birlikte algoritmayı başka bir alana uyguladığımızda sonuç genellikle başarılı olmamaktadır. Ancak sözlük tabanlı bir yaklaşım, büyük verilerin depolanmasını ve eğitim süreci gerektirmeyen duygu sözlüklerini kullanır (Chamass vd., 2018). Sözlük tabanlı yöntemler, metindeki her bir kelime için bireysel duygu puanlarını toplayarak bir metin parçasının toplam duygusunu bulurken olumlu ya da olumsuz duygu etiketleriyle eşlenen sözcüklerin bir listesini oluşturmaktadır. Duygu sözlükleri, duygu analizi yapmak için popüler bir araç haline gelmiştir. (Sohangir vd., 2018; Bonta ve Janardhan, 2019; Choi vd., 2020; Bulut, 2022). Literatürde yer alan başlıca duygu sözlükleri Bing, Afinn, Loughran, NRC, Syuzhet' dir. Bing Liu ve arkadaşları (2005) tarafından oluşturulmuş olan Bing sözlüğünde duygular basit bir şekilde negatif ve pozitif olarak sınıflandırılmıştır. Bing sözlüğü, yalnızca -1 ve +1 olmak üzere ikili bir kategorizasyona sahiptir. Finn Årup Nielsen (2011) tarafından 2009 ve 2011 yılları arasında oluşturulan Afinn sözlüğü -5 (negatif/olumsuz duygular) ve +5 (pozitif/olumlu duygular) arasında bir tamsayı değeri alan ve manuel olarak derecelendirilen İngilizce terimlerin bir listesidir. Afinn sözlüğü her kelimeyle ilişkili bir duygu puanına sahip 3300'den fazla kelime içermektedir. Saif Mohammad ve Peter Turney (2010) tarafından oluşturulan NRC sözlüğü kelimelerin sekiz duyguya (öfke, beklenti, öğrenme, korku, sevinç, üzüntü, şaşkınlık, güven) ek olarak 2 duyguyla (pozitif ve negatif) ilişkilendirilen gerçek değerli yoğunluk

puanlarına sahip İngilizce kelimelerin bir listesidir. Loughran ve McDonald'ın (2011) ortaya koyduğu Loughran sözlüğü, finansal belgelerde kullanılmak amacıyla oluşturulmuştur. Bu sözlük, finansal bağlamlarda önemli olan altı olası duyguyla kelimeleri etiketler: "negatif", "pozitif", "ihtilafçı", "belirsizlik", "kısıtlayıcı" veya "gereksiz". Syuzhet sözlüğü ise, Matthew Jockers (2020) tarafından geliştirilen Syuzhet paketinin bir parçasıdır. Syuzhet, esas olarak dört standart sözlükten (Syuzhet, Bing, Afinn ve NRC) yararlanarak oluşturulan bir pakettir. Nebraska Edebiyat Laboratuvarı'nda geliştirilen Syuzhet, edebi metinler için bir duygu analiz aracı olarak tanımlanmaktadır. (Kim, 2022). 10748 kelime içeren bu sözlükte her bir duygu kelimesi -1 ile +1 duyarlılık değerleri arasında değişmekte ve daha spesifik duyarlılık değerlerine (-1.0, -0.8 -0.75, -0.6, -0.5, -0.4, -0.25, 0.1, 0.25, 0.4, 0.5, 0.6, 0.75, 0.8, 1.0) sahip olmaktadır (Jockers 2020).

Literatürde hastalıkları teşhis ve tedavi etmede YZ algoritmalarının geliştirilmesine yönelik birçok çalışma bulunmaktadır. Ancak YZ kullanılarak hastalıkları teşhis ve tedavi etmek için geliştirilen yöntemlerin klasik yöntemler ile maliyet etkililiğini karşılaştıran çalışmalar açısından literatürde boşluklar yer almaktadır. Bu sebeple bu çalışma, ilgili alan yazında ne kadar yayın yapıldığını, yapılan yayınlar arasındaki ilişkileri, bu alanda araştırma yapılmasına duyulan ihtiyacı ve araştırma yapılırken yararlanılacak anahtar kelimeleri ortaya koyarak literatüre katkı sağlayacaktır. Bununla birlikte bu çalışma, bu yayınlarda en çok tekrar edilen kelimeleri vurgulayarak çalışmalarda odaklanılan hastalıkların ve konuların belirlenmesini sağlayacaktır.

Tıbbi teknoloji kullanımı ile ilgili konularda bilimsel kararlar alınmasını sağlayan Sağlık Teknolojisi Değerlendirmesi (STD), sağlıkta YZ uygulamalarının tanıtılması ve YZ teknolojilerinin sağlık alanında bir karar destek sistemi olarak kullanılmasının anahtarıdır. Bu çalışmanın, literatürdeki araştırma eğilimlerini ortaya koyarak STD sürecine ve bu alandaki araştırmalara ışık tutmaya imkân sağlayacağı düşünülmektedir.

## 2. YÖNTEM (METHOD)

Bu araştırmanın amacı, hastalıkları teşhis ve tedavi etmede YZ algoritmalarıyla geliştirilen modellerin geleneksel teşhis ve tedavi modellerine göre maliyet etkili olup olmadığını ortaya koyan çalışmaların detaylı bir incelenerek çalışmaların zaman içindeki eğiliminin belirlenmesidir. Bununla birlikte hangi konularda daha fazla araştırma yapıldığının incelenmesi ve konuyla ilgili küresel ilginin açığa çıkarılması çalışmanın diğer amaçlarını oluşturmaktadır. Bununla birlikte bu çalışmada sağlık alanında YZ teknolojilerinin sağladığı yöntemin mevcut yöntemlere göre maliyet etkililiğini araştıran yayınlarda en çok tekrar edilen kelimeleri vurgulamak ve bu çalışmaları duygu durumuna göre sınıflandırmak amaçlanmıştır. Bu sebeple bu araştırmada, belirli bir alanda üretilmiş yayınların ve bu yayınlar arasındaki

ilişkilerin sayısal olarak analizini ortaya koyan bibliyometrik analiz; metin verilerini görselleştirmenin en kolay yolu olan ve yapılan çalışmalarda en fazla tekrar edilen kelimeleri ortaya koyan kelime bulutları; bir metin belgesini olumlu veya olumsuz bir görüş ya da duygu durumuna göre sınıflandırmayı amaçlayan duygu analizi kullanılmıştır.

### 3. VERİ TOPLAMA VE İŞLEME (DATA COLLECTION AND PROCESSING)

Bibliyometrik analiz için Ağustos 2022'de Web of Science (WoS) veri tabanında literatür taraması yapılmıştır. Araştırma Ağustos 2022 yılı ve öncesi bütün yayınları kapsamaktadır. Literatür taraması sonucunda bu alanda yapılan çalışmaların 2018 yılında başladığı gözlenmiştir. Literatür taramasında yer alan anahtar kelimeler; "Cost Effectiveness of Artificial Intelligence" ya da "Cost Effectiveness of Machine Learning" ya da "Cost-Utility Analysis of Artificial Intelligence" dir. Literatür taramasında başlık kategorisine göre yapılan ilk aramada anahtar sözcüklerle eşleşen 31 belgeye ulaşılmıştır. Bu 31 belge arasından, çalışma amaçlarımıza uymayan yayınları temizlemek için dışlama kriterleri kullanılmıştır. Çalışmada araştırma hedefimize uyan araştırma makaleleri ve bildiriler (Meeting Abstract) çalışmamıza dahil edilmiştir. Araştırma hedefimize uymayan konferans makaleleri (Proceeding Paper) ve derleme çalışması şeklinde olan bazı bildiriler araştırmadan çıkarılmıştır. Bütün belgelerin dili İngilizce'dir. Bu dışlama kriterlerini uyguladıktan sonra, bibliyometrik analiz için toplamda 24 çalışmaya (14 makale+10 bildiri) ulaşılmıştır. Bu çalışmada bibliyometrik haritalama için VOSviewer yazılımı (van Eck ve Waltman, 2020) kullanılarak veriler analiz edilmiştir. Çalışmanın bibliyometrik analiz kısmında yıllara göre yayınlanan toplam yayın ve atıf sayısı, yazarların anahtar kelime oluşum ağı, yazarların iş birliği ağı ve en çok atıf alan makaleler sunulmuştur.

Kelime bulutu ve duygu analizi için 24 çalışmanın referans kısımları çıkartılarak çalışma tek bir pdf haline getirilmiştir. İlk olarak pdf metnini temizlenip kullanıma hazır hale getirmek için metin ön işleme aşamasından geçirilmiştir. Bu aşamada, metinde üç harften daha büyük kelimeler metne dahil edilmiş; metinde bulunan zamir, bağlaç, edat, sıfat gibi kelimeler metninden çıkarılmış; araştırmada istenmeyen formattaki bazı kelimelerin yerine yenisi atanmış; büyük harflerin tamamı küçük harfe dönüştürülmüş, noktalama işaretleri kaldırılmış; metinde iç içe geçmiş rakam ve sayılar temizlenmiş; metinden rakam ve sayılar çıkarılmıştır. Bu işlemlerden sonra metin "tibble" tablo düzenine dönüştürülmüştür. Araştırmada kelime bulutlarının yıllara göre nasıl farklılık gösterdiğini analiz etmek için 2018-2020 yılları arasındaki çalışmalar bir pdf'de; 2021 ve 2022 yıllarındaki çalışmalar ise başka bir pdf'de birleştirilerek yukarıdaki ön işleme aşamasından geçirilmiştir. Daha sonra yıllara göre kelime bulutları ve en sık kullanılan 20 kelimenin bulunduğu tablolar oluşturulmuştur. Bu çalışmada duygu analizi için sözlük tabanlı yöntem uygulanarak Bing, Afinn, Loughran ve NRC olmak üzere dört farklı sözlükten

yararlanılmıştır. Kelime bulutları ve duygu analizi için R yazılım programı kullanılarak veriler analiz edilmiştir (Bali, 2017).

### 4. BULGULAR (RESULTS)

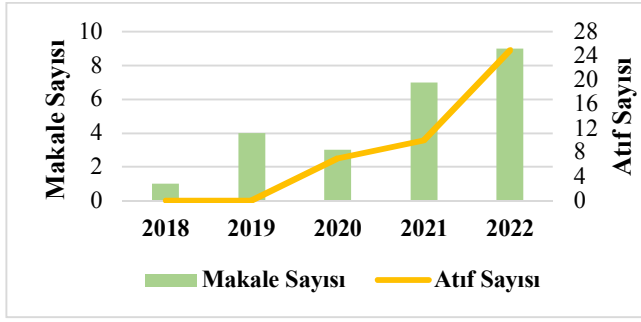
Çalışma kapsamında elde edilen yayınların bibliyometrik analizi, kelime bulutları ve duygu analiz bulguları aşağıda tartışılmaktadır.

#### 4.1. Bibliyometrik Analiz (Bibliometric Analysis)

Çalışmanın bibliyometrik analiz kısmında elde edilen yayınların, yayın ve atıf sayıları; yazarların anahtar kelime oluşum ağı; yazarlar ve iş birlikleri; en çok atıf alan makalelerin bulguları aşağıda sunulmuştur.

##### 4.1.1. Yayın ve Atıf Sayıları (Publication and Citation Numbers)

Yıllara göre toplam yayın sayısı, belirlenmiş bir araştırma alanındaki araştırma eğilimlerinin bir envanterini gösterebilmektedir. Toplam yayın sayısı bu alandaki eğilimlerin yakın gelecekte bir özetini sunmakta bununla birlikte belirli bir zaman çerçevesi içinde bir araştırma konusunun büyümesini de göstermektedir (Noman vd., 2022). Hastalıkları tedavi ya da teşhis etmede YZ teknolojilerinin ortaya koyduğu yöntemle geleneksel yöntemlerin maliyet etkililiğini karşılaştıran ilk çalışmalar 2018 yılında yayınlanmıştır. Bu çalışma, YZ tabanlı yazılım programlarının kullanılmasıyla tıbbi testlerin maliyetlerinin azaltılabileceğini göstermiştir (Gönel & Koyuncu, 2018). Şekil 1' de bu çalışmada yıllara göre yayınlanan araştırmalar ve araştırmaların aldığı atıf sayısı sunulmaktadır. Şekil 1' de 2020 yılından itibaren sağlık alanında YZ'nin teşhis ya da tedavi kabiliyetini ve maliyetini mevcut teşhis ya da tedavi yöntemine göre karşılaştıran çalışmalarda kayda değer bir büyüme görülmektedir. Son yıllarda yayınların artış göstermesi dünya çapındaki araştırmacıların bu alana ilgi duyduğunu belirtmektedir. Yayın sayısının çok az olması sebebiyle 2018 ve 2019 yıllarındaki çalışmalara atıf yapılmamıştır. Şekil 1' de, 2020 yılından itibaren atıf sayısında ciddi bir artış görülmektedir. Atıf sayılarındaki bu artış, YZ teknolojilerinin ortaya koyduğu yöntemle geleneksel yöntemlerin maliyet etkililiğini karşılaştıran çalışmaların dünya çapında araştırmacılar tarafından geniş çapta kabul görmeye başladığını göstermektedir.

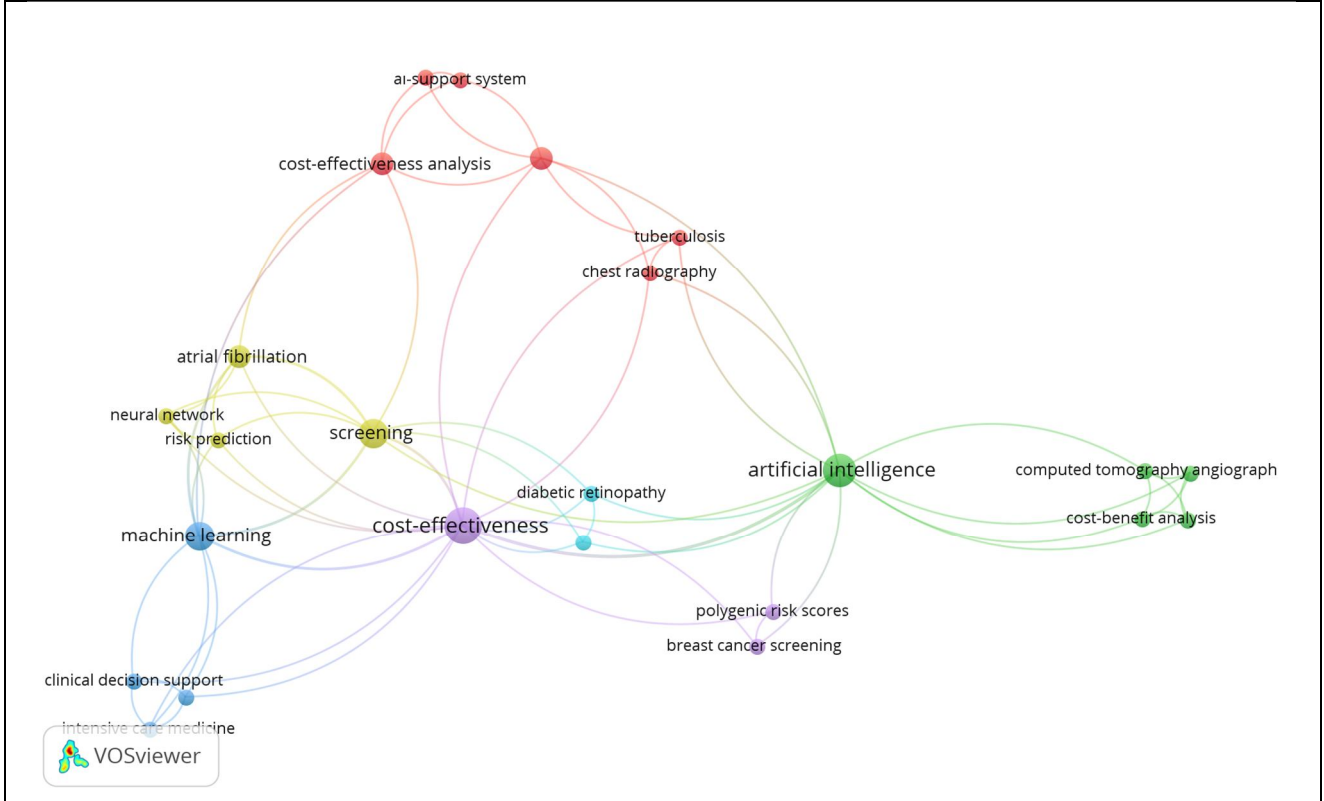


Şekil 1. Yıllara Göre Makale ve Atıf Sayıları  
(Number of Articles and Citations by Years)

#### 4.1.2. Yazarların Anahtar Kelime Oluşum Ağı (Keyword Co-Occurrence Network of Authors)

Anahtar kelimeler, bir makalede özet kısmından sonra belirtilen kelimeler ve belge ile ilgili diğer bazı dizin

terimleridir. Yazarların anahtar kelime oluşum ağı ise yazarların sadece özet bölümünden sonra kullandığı bu alandaki teknik anahtar kelimeleri sunmaktadır. Çalışmada toplam anahtar kelime sayısı 36 ve doküman sayısının az olması sebebiyle bir anahtar kelimenin minimum tekrar sayısı 1 olarak belirlenmiştir. Bu durum, tüm dokümanlarda bir kez geçen anahtar kelimelerin Şekil 2’de görüleceği anlamına gelmektedir. Şekil 2, 36 anahtar kelimenin 24 tanesinin birlikte oluşum grafiğini temsil etmektedir. Beklendiği gibi, YZ ve maliyet etkililik en çok kullanılan anahtar kelimelerdir ve YZ 20 defa, maliyet etkililik ise 15 defa geçmektedir. Ardından MÖ, tarama ve DÖ üçüncü, dördüncü ve beşinci sırada yer almaktadır. Bu çalışmada yalnızca İngilizce dilde yazılmış makaleler inceleme kapsamına alınmıştır. Çalışmadaki metinler İngilizce olduğu için analiz sonuçlarındaki görseller İngilizce olarak sunulmuştur ancak şeklin altında İngilizce kelimelerin Türkçesi yer almaktadır.



Kelimelerin Türkçe karşılıkları: artificial intelligence-yapay zekâ, cost effectiveness- maliyet etkililik, machine learning-makine öğrenimi, screening-tarama, diabetic rethinopathy- diyabetik retinopati, atrial fibrillation-atriyal fibrilasyon, neural network-sinir ağı, chest radiography-göğüs radyografisi, polygenic risk scores-poligenik risk puanları, tuberculosis-tüberküloz-risk prediction-risk tahmini, breast cancer screening-göğüs kanseri taraması, clinical decision support-klinik karar desteği, intensive care medicine-yoğun bakım ilacı

Şekil 2. Yazarların Anahtar Kelime Oluşum Ağı  
(Keyword Co-Occurrence Network of Authors)

#### 4.1.3. Yazarlar ve İş Birlikleri (Co-authorship)

Bu çalışmada incelenen toplam 24 makale 155 yazar tarafından yazılmıştır. Tablo 1, WoS veri tabanında yer

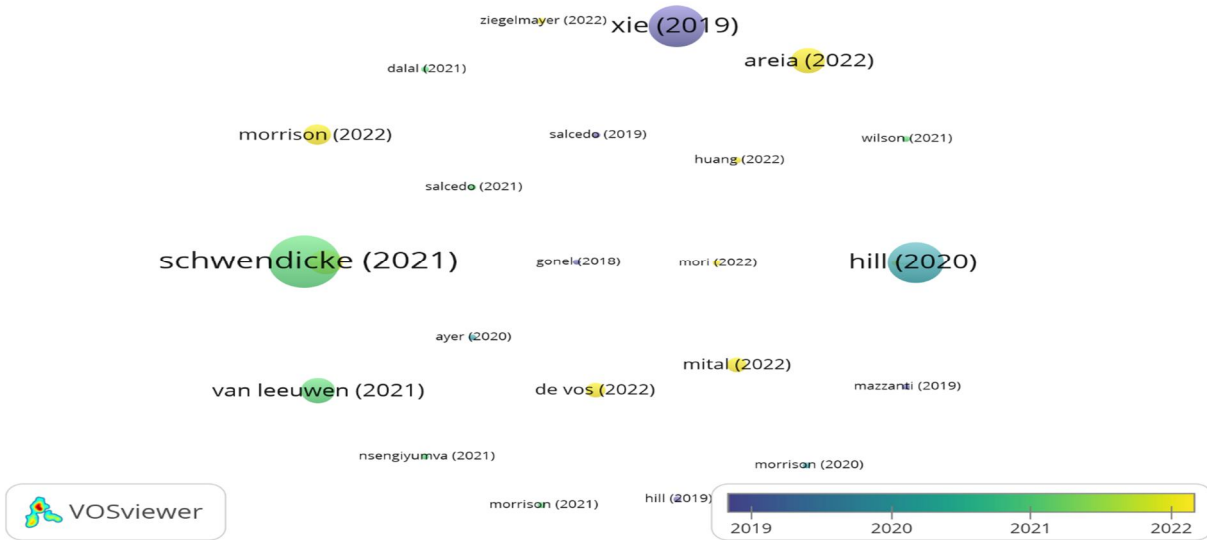
alan dergilerde yayın yapan en üretken ilk 7 yazarı göstermektedir. Yedi yazar da eşit sayıda yayın yapmıştır. En çok yayın yapan 7 yazardan Sandler, B., Lister, S., Boyce, R. Farooqui, Gordon, J. vd'nin birlikte yapmış



hekimlerine göre daha yüksek ve YZ yönteminin diş hekimlerinden önemli ölçüde daha duyarlı olduğu sonucuna ulaşmıştır.

Ayrıca makalede diş çürüğü tespitinde YZ klasik yöntemle göre tüm vakaların çoğunda (>%77) daha az maliyetli ve daha etkili olduğu tespit edilmiştir. AF vakaları genellikle asemptomatik olduğundan, hastaların komplikasyonlar (örn. inme) ortaya çıkana kadar teşhis edilmesi zor olmaktadır. Bu yüzden ikinci makale, AF' li hastaları bir MÖ risk tahmin algoritması tarafından taramanın maliyet etkinliğini değerlendirmeyi amaçlamıştır. Bu çalışmada Hill ve diğerleri hibrid öğrenme yaklaşımına dayanan bir karar ağacı ve Markov modeli kullanarak maliyet-etkililik analizi yapmıştır. Hill ve diğerleri AF' yi tespit etmede geleneksel sistematik AF tarama stratejilerinin maliyet ve sonuçlarını bir MÖ risk tahmin algoritma taramasının maliyet ve sonuçlarıyla karşılaştırmıştır. Çalışma sonuçları, MÖ tarama stratejisinin, değerlendirilen tüm

senaryolarda uygun maliyetli olduğunu göstermiştir. İkinci makale ile aynı sayıda atıfa sahip olan Xie (2019) ve diğerleri tarafından oluşturulan bir diğer makale "Cost-Effectiveness Analysis of an Artificial Intelligence-Assisted Deep Learning System Implemented in the National Tele-Medicine Diabetic Retinopathy Screening in Singapore" başlıklı makaledir. Bu makalede, Diyabetik Retinopati taramasında bu hastalığın evrelerini manuel derecelendiren sisteme kıyasla DÖ sistemli bir modelle derecelendirmenin maliyet etkinliği incelenmiştir. DÖ sistemli tarama modeli hasta başına 135 S\$ (135 Singapur doları)'lık ömür boyu maliyet tasarrufu sağlamaktadır. 2025 yılına kadar yaklaşık 250.000 hastaya tarama yapılacağı göz önüne alındığında, DÖ sistemli taramanın gelecekteki maliyet tasarruflarının mevcut değerinin 33.8 milyon dolar olduğu tahmin edilmektedir. Araştırma sonuçları, DÖ sistemli tarama modelinin, mevcut manuel derecelendirme sisteminden daha az maliyetli olduğunu göstermiştir.



Şekil 4. En Çok Atıf Alan Makaleler  
(Most Cited Articles)

Tablo 2. En Çok Atıf Alan İlk 3 Makale  
(Top 3 Most Cited Articles)

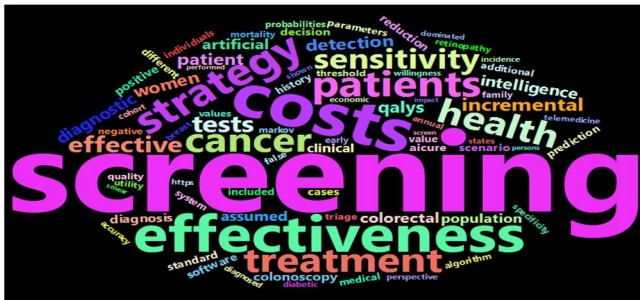
Makale Başlığı	Yazar	Dergi Adı	Yayın Yılı	Atıf Sayısı
Cost-effectiveness of Artificial Intelligence for Proximal Caries Detection	Schwendicke, F vd.	Journal of Dental Research	2021	13
Cost-effectiveness of targeted screening for the identification of patients with atrial fibrillation: evaluation of a machine learning risk prediction algorithm	Hill, NR vd.	Journal of Medical Economics	2020	8
Cost-Effectiveness Analysis of an Artificial Intelligence-Assisted Deep Learning System Implemented in the National Tele-Medicine	Xie, Y. vd.	Investigative Ophthalmology & Visual	2019	8

Diabetic Retinopathy Screening in Singapore		Science		
---	--	---------	--	--

#### 4.2. Kelime Bulutları

Hastalıklar teşhis ve tedavi etmede YZ teknolojilerinin sağladığı yöntemin mevcut yöntemle göre maliyet etkinliğini inceleyen çalışmalardan elde edilen kelime bulutları oluşturulmuş olup Şekil 5' de bütün makalelerin, Şekil 6 ve 7'de ise yıllara göre ayrılmış makalelerin kelime bulutları yer almaktadır. Bütün makalelerin yer aldığı kelime bulutunda en sık tekrar edilen kavram "tarama (screening)"dir. Hastalıklara tanı koymak için YZ

teknolojileri kullanılarak tarama yapmanın YZ teknolojileri kullanılmayan yöntem ile kıyaslanması tarama kavramının çalışmalarda en sık kullanılan kavram olmasını sağlamıştır. Bununla birlikte, kelime bulutunda maliyetler, etkililik, strateji gibi kavramlar sık kullanılmaktadır. Tablo 3 ise sıklığı en yüksek ilk 20 kelimeyi göstermektedir. Genel olarak Tablo 3’ de en sık kullanılan kelimeler arasında hastalar, sağlık, kanser, testler, tedavi gibi kavramların olduğu görülmektedir. Tablo 3, hastalıklara tanı koymada YZ teknolojilerinin klasik yöntemle göre maliyet etkililiğini inceleyen çalışmaların daha çok kanser hastalığının teşhisinde ve tanı testleri üzerinde yoğunlaştığını göstermektedir. Ayrıca Tablo 3, tanı testleri aracılığı ile YZ teknolojilerinin geliştirildiğini ve YZ teknolojilerinin hastalıkların tedavi aşamasında kullanıldığını belirtmektedir.



Şekil 5. İlgili Bütün Makalelerin Kelime Bulutu  
(Word Cloud of All Related Articles)  
(\*Kelimelerin Türkçeleri Tablo 3’ de yer almaktadır.)

Tablo 3. Sıklığı En Yüksek İlk 20 Kelime  
(Top 20 Words with the Highest Frequency)

No	Kelime	Sıklık
1	tarama (screening)	983
2	maliyetler (costs)	541
3	etkililik (effectiveness)	447
4	strateji (strategy)	340
5	hastalar (patients)	309
6	tedavi (treatment)	293
7	sağlık (health)	288
8	kanser (cancer)	287
9	duyarlılık (sensitivity)	243
10	testler (tests)	203
11	etkili (effective)	197
12	ilave (incremental)	151
13	tespit etme (detection)	139
14	qalys (qalys)	138
15	kadın (women)	137
16	hasta (patient)	136
17	tanı (diagnostic)	132
18	zeka (intelligence)	128
19	kolorektal (colorectal)	122
20	yapay (artificial)	121

Şekil 6 ve Şekil 7’de yapılan çalışmalar yıllara göre ayrılmıştır. Şekil 6’da 2018- 2020 yılları arasında ilgili alan yazındaki makalelerin kelime bulutu gösterilmektedir. Bu yıllar arasında toplamda 9 çalışma mevcuttur. Tablo 4, bu yıllar arasında çalışmalarda en çok

kullanılan 20 kelimeyi göstermektedir. Tablo 3 ile benzer şekilde Tablo 4’te tarama, test, hastalar, etkililik, maliyetler gibi kelimeler sıklıkla kullanılmaktadır. Ancak bu yıllar arasında yapılan çalışmaların daha az olması sebebiyle bu kelimelerin sıklık değerleri daha düşüktür.



Şekil 6. 2018-2020 Yılları Arasındaki İlgili Makalelerin  
Kelime Bulutu  
(Word Cloud of Related Articles for 2018-2020)  
(\*Kelimelerin Türkçeleri Tablo 4’ de yer almaktadır.)

Tablo 4. Sıklığı En Yüksek İlk 20 Kelime (2018-2020  
Yılları) (Top 20 Words with the Highest Frequency (2018-  
2022Years))

No	Kelime	Sıklık
1	tarama (screening)	167
2	testler (tests)	119
3	hastalar (patients)	68
4	etkililik (effectiveness)	57
5	maliyetler (costs)	53
6	strateji (strategy)	48
7	fırsatçı (opportunistic)	42
8	sistemik (systematic)	38
9	laboratuvar (laboratory)	37
10	yazılım (software)	35
11	algoritma (algorithm)	33
12	bilirubin (bilirubin)	33
13	tahmin (prediction)	31
14	tedavi (treatment)	30
15	doğrudan (direct)	25
16	etkili (effective)	25
17	linik (clinical)	24
18	duyarlılık (sensitivity)	24
19	zeka (intelligence)	22
20	vakalar (cases)	21

Şekil 7 ise 2021 ve 2022 yıllarında yapılan 15 çalışmanın kelime bulutunu göstermektedir. Tablo 5’deki 2021 ve 2022 yıllarında yapılan çalışmaların sıklık değerleri Tablo 4’deki 2018-2020 yılları arasındaki çalışmaların sıklık değerlerine göre daha yüksektir. Bu durum, 2021 ve 2022 yıllarında YZ’nin sağlık problemlerinin çözümünde ortaya koyduğu yöntemin mevcut yöntemle göre maliyet etkililiğini inceleyen yayınların sayısında artış olduğunu göstermektedir. Bununla birlikte, Şekil 7’ de kanser



Ortanca	Ort.	SH.ort	GA.ort.0.95	Var	Std.	Var.kelime
1	0.308	0.036	0.070	3.456	1.859	6.045

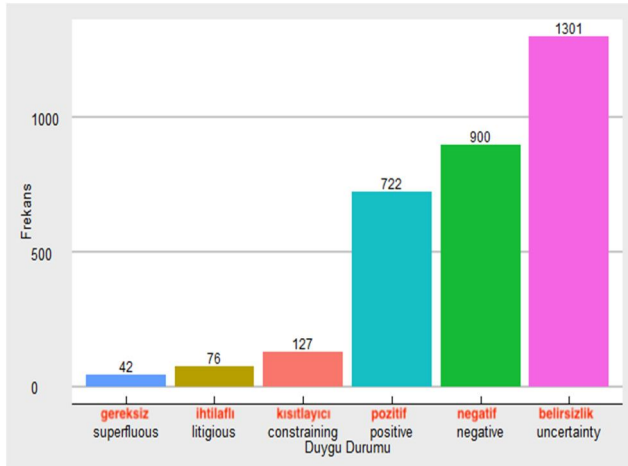
(Tablo açıklamaları: Top: veri kümesindeki toplam değer sayısı; Ol.: var olmayan değerlerin sayısı; Ek.say: eksik değer sayısı; Min: kümesindeki minimum değer; Maks: veri kümesindeki maksimum değer; fark: minimum ve maksimum arasındaki Fark top: kümesindeki değerlerin toplamı; Ort: aritmetik ortalama; orta: ortanca, SH.ort: ortalamının standart hatası; GA.ort.0.95: ortalama %95 güven aralığı; var: Varyans; Std.: standart sapma; Var. varyasyon katsayısı)

#### 4.3.3. Loughran Sözlüğü (Loughran Lexicon)

Tablo 8’ de Loughran sözlüğünün duygu sınıflarına göre tanımlayıcı istatistikleri verilmiştir. Şekil 9’ da ise Loughran sözlüğünde yer alan “negatif”, “pozitif”, “ihtilaflı”, “belirsizlik”, “kısıtlayıcı” veya “gereksiz” şeklindeki altı duygunun sıklık değeri grafik şeklinde gösterilmiştir. Tablo 8 ve Şekil 9’ da görüldüğü gibi çalışmada %41.07 sıklık oranıyla belirsizlik duygusu hakimdir. Çalışmada belirsizlik duygusunu negatif ve pozitif duygular izlemiştir.

Tablo 8. Loughran Sözlüğü Tanımlayıcı İstatistikleri (Loughran Lexicon Descriptive Statistics)

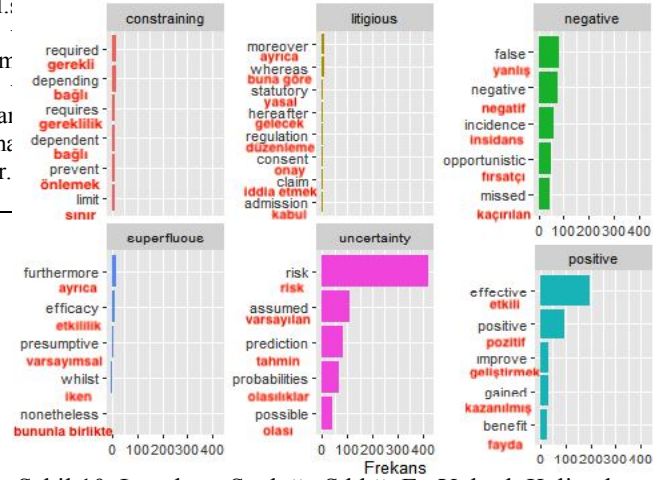
Duygu Durumu	Toplam	Oran
belirsizlik (uncertainty)	1301	41.07
negatif (negative)	900	28.41
pozitif (positive)	722	22.79
kısıtlayıcı (constraining)	127	4.01
ihtilaflı (litigious)	76	2.40
gereksiz (superfluous)	42	1.33



Şekil 9. Loughran Sözlüğüne Göre Duygu Kategorilerinin Sıklık Değerleri

Şekil 10’da ise, Loughran sözlüğündeki duygu durumlarına göre en çok kullanılan kelimeler gösterilmektedir. Şekilde “belirsizlik” duygusunda en sık kullanılan kelimenin “risk” olduğu görülmektedir. Bing ve NRC sözlüğündeki sonuçlarla benzer şekilde belirsizlik duygusunda en sık kullanılan “risk” kelimesinin sıklığı 400’den fazladır. Şekilde pozitif duygu durumunda “etkili” kelimesinin, negatif duygu

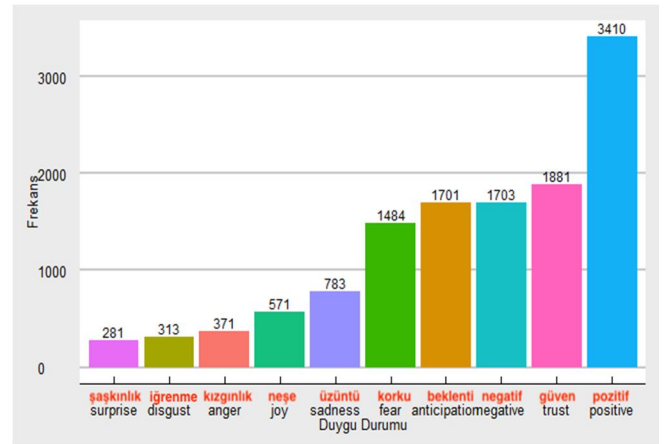
durumunda ise “yanlış” kelimesinin sıklığı en yüksek olan kelime olduğu görülmektedir.



Şekil 10. Loughran Sözlüğü Sıklığı En Yüksek Kelimeler (Loughran Lexicon Top 10 Words With The Highest Frequency)

Tablo 9. NRC Sözlüğü Tanımlayıcı İstatistikleri (NRC Lexicon Descriptive Statistics)

Duygu Durumu	Toplam	Oran
pozitif (positive)	3410	27.28
güven (trust)	1881	15.05
negatif (negative)	1703	13.63
beklenti (anticipation)	1701	13.61
korku (fear)	1484	11.87
üzüntü (sadness)	783	6.27
neşe (joy)	571	4.57
kızgınlık (anger)	371	2.97
iğrenme (disgust)	313	2.5
şaşkınlık (surprise)	281	2.25



Şekil 11. NRC Sözlüğüne Göre Duygu Kategorilerinin Sıklık Değerleri

(Frequencies of Sentiment Categories by NRC Lexicon)

#### 4.3.4. NRC Sözlüğü (NRC Lexicon)

NRC sözlüğünde, metin duyguları “öfke”, “beklenti”, “iğrenme”, “korku”, “neşe”, “negatif”, “pozitif”, “şaşkınlık” ve “güven”i içermektedir. Tablo 9’da NRC sözlüğünün duygulara göre tanımlayıcı istatistikleri verilmiştir. Şekil 11’de ise duygu durumlarının sıklık



değeri grafik şeklinde gösterilmiştir. NRC sözlüğünde pozitif duygunun 3410 sıklık ve %27,28 sıklık oranı ile çalışmanın genelinde baskın olduğu görülmektedir. Çalışmada pozitif duyguyu güven ve negatif duygular takip etmiştir.

Şekil 12’de ise, NRC sözlüğünde yer alan her bir duygu durumu için sıklığı en yüksek 5 kelime gösterilmektedir.



Şekil 12. NRC Sözlüğü Sıklığı En Yüksek İlk 5 Kelime  
(NRC Lexicon Top 5 Words With The Highest Frequency)

## 5. TARTIŞMA (DISCUSSION)

Literatürde birçok farklı konu ile ilgili bibliyometrik analiz çalışmaları yapılmıştır. Bu çalışmada ise ilk kez bir karar destek sistemi olarak kullanılan yapay zekânın sağlık sorunlarının tespitinde ortaya koyduğu yöntemin mevcut yöntemlere göre maliyet etkililiğini tespit eden çalışmaların bibliyometrik analizi yapılmıştır. Çalışma sonuçları, bu konuyla ilgili oldukça kısıtlı (24) yayına ulaşıldığını ancak önceki yıllara kıyasla son yıllarda (2021-2022) konuyla ilgili üretilen yayınların sayısında artış olduğunu göstermiştir. Çalışmada yer alan metinlerde en çok tekrar edilen anahtar kelimelerin YZ, maliyet etkililik, tarama ve MÖ olduğu sonucuna ulaşılmıştır. Ayrıca çalışma bulguları, yazarların iş birliği ağında, iki ana yazar grubuna (küme) ulaşmıştır. Her iki kümede AF'nin MÖ ile tahmin etmenin maliyet etkililiğini ortaya koyarken bir kümenin daha çok birinci basamak sağlık hizmetlerini alan hastalar üzerinde yapılmış olması ve tanı testleri kullanılarak bir algoritma geliştirmesi kümeleri birbirinden ayırmaktadır. Çalışmanın atıf analiz bulguları, dış çürüğünü teşhis etmede YZ ile görüntü değerlendiren yöntemin klasik görüntü değerlendirme yöntemine göre maliyet etkililiğini inceleyen makalenin en çok atıf aldığı ortaya koymuştur. Bununla birlikte konuyla ilgili yapılan çalışmaların daha çok AF ve diyabetik retinopati hastalıklarına tanı koyma ile ilgili olduğu görülmüştür.

Şekilde “pozitif”, “güven” ve “negatif” duygu durumlarında sırasıyla en sık kullanılan kelimelerin “etkili” ve “risk” olduğu görülmektedir. Bing sözlüğündeki sonuçlarla benzer şekilde negatif duyguda en sık kullanılan “risk” kelimesinin sıklığı 400’den fazla iken pozitif duyguda “etkili” kelimesinin sıklığının 200 olduğu sonucuna ulaşılmıştır.

Bu çalışmanın gücünü artırmak ve daha kapsayıcı sonuçlar elde edilebilmek için bir karar destek sistemi olarak kullanılan YZ'nin sağlık sorunlarının tespitinde ortaya koyduğu yöntemin mevcut yöntemlere göre maliyet etkililiğini inceleyen çalışmalarda ön plana çıkan kelimeler vurgulanmıştır. Bu kelimeler arasında en çok ön plana çıkan kavramın “tarama” olduğu sonucuna ulaşılmıştır. Tarama kavramının diğer kelimelere oranla daha sık kullanılması, sağlık alanında YZ teknolojileri geliştirilerek hastalıkların taranmasının önemini ortaya koymaktadır.

Çalışmada yıllara göre oluşturulan kelime bulutlarında ise kanser kelimesinin ön plana çıktığı görülmüştür. Çalışmanın duygu analiz bulguları incelendiğinde Loughran sözlüğüne göre belirsizliğin hâkim olduğu bir duygu durumu söz konusu iken Bing, Afinn ve NRC sözlüğü için çalışmada pozitif duyguların daha ağır bastığı gözlenmiştir. Bu sözlüklerde pozitif duygunun daha çok ağır basması sağlık sorunlarının tespitinde karar destek sistemi olarak kullanılan YZ teknolojilerinin sağladığı yöntemin mevcut yöntemlere göre maliyet etkililiğini tespit eden çalışmaların olumlu bir izlenime sahip olduğunu ifade etmektedir. Kısaca, çalışmada elde edilen maliyet etkililik ile ilgili yayınlar hem YZ teknolojilerinin daha etkili hem de daha uygun maliyetli olduğu noktasında pozitif düşünceler sunarak gelecekte benzer çalışmaların yapılmasının yeni teknolojilerin kişilerin sağlıklarına ve ülke ekonomisine sağladığı

faydaları gösterme noktasında önemli katkılar sunacaktır. Bununla birlikte, konuyla ilgili WoS veri tabanından az sayıda çalışmaya ulaşımları Loughran sözlüğünde belirsizlik duygusunun ağır basmasına sebebiyet verebilmektedir. Bu sebeple bu çalışmaların sayısının artırılması bu konudaki duygu durumlarının daha net olarak ifade edilmesinde ve bu çalışmalara olan yaklaşımı ortaya koymada önem arz etmektedir.

Guo ve diğerleri (2020)'nin YZ ile sağlık hizmetlerinin ilişkili olduğu makaleler üzerinden yaptığı bibliyometrik analiz sonuçları, konuyla ilgili araştırılan yayınların yılda ortalama %17,02 arttığını gösterirken bu çalışma, bir karar destek sistemi olarak kullanılan YZ'nin sağlık sorunlarının tespitinde ortaya koyduğu yöntemin mevcut yöneme göre maliyet etkililiğini analiz eden çalışmaların literatürde çok sınırlı (24 adet) olduğunu göstermektedir. Ancak Guo ve diğerlerinin (2020) çalışmalarıyla benzer şekilde son yıllarda (2021-2022) sağlık alanında YZ teknolojilerinin sağladığı yöntemin mevcut yöneme göre maliyet etkililiğini araştıran yayınların sayısı da artmıştır.

Dinakaran ve Anitha (2018) YZ teknolojilerini kullanan temel hastalık alanları arasında kanser, nöroloji ve kardiyoloji bulunduğunu belirtmiştir. Erken teşhisin çok önemli olduğu kanser gibi bazı hastalık türleri için, hastalığın YZ aracılığıyla taranarak erken teşhis edilmesi birçok kişinin hayatının kurtaracaktır. Çalışmada özellikle kanser kelimesinin ön plana çıkması kanser alanında YZ teknolojilerinin sağladığı yöntemin mevcut yöneme göre maliyet etkililiğini araştıran yayınların mevcut olduğunu göstermekte ancak çalışmaların sayısını az olması bu çalışmalara daha fazla ihtiyaç duyulduğunu ortaya koymaktadır.

Guo ve diğerleri (2020) YZ araştırmalarında incelenen başlıca sağlık sorunlarının kanser, depresyon, Alzheimer hastalığı, kalp hastalıkları ve diyabet olduğunu belirtmiştir. Kalp ve damar hastalıkları, dünyada başlıca ölüm nedenleri arasında yer almaktadır. Dünya Sağlık Örgütü 2018 yılında, bu hastalıklardan dolayı 18 milyon kişinin hayatını kaybettiğini ve 2030 yılında kalp ve damar sorunlarına bağlı ölümlerin 23 milyondan fazla olacağını belirtmiştir. Kalp ve damar hastalıklarına yakalanma riskini tahmin etmek için YZ teknolojilerinin kullanılması ve YZ teknolojileri ile hastalığın erken teşhis edilebilmesi bu hastalığa bağlı ölüm oranlarını azaltacaktır. Çalışmada kalp hastalıklarından biri olan AF hastalığına yakalanma riskini tahmin etmede YZ ile geliştirilen yöntemin mevcut yöneme göre maliyet etkililiğini araştıran yayınlar bulunmaktadır. Ancak bütün kalp ve damar hastalıkları için bu tür çalışmaların yapılması YZ ile geliştirilen yöntemin sağlık hizmeti sunucuları tarafından benimsenmesini sağlayacaktır.

YZ araştırmalarına konu olan ve yaygın bir halk sağlığı sorunu olan diyabet tüm dünyada artış göstermektedir. Uluslararası diyabet atlasına göre 2021 yılında 537 milyon diyabetli hasta bulunmaktadır. Bu çalışmada yapay zekanın maliyet etkililiğini inceleyen çalışmalar arasında diyabetik retinopati hastalığında YZ taramasının

göz doktorları tarafından fundus görüntülerinin değerlendirildiği oftalmolog taramasına göre maliyet etkililiğini inceleyen çalışmaların bulunması araştırmalarda diyabet gibi kronik rahatsızlıklara odaklanıldığını göstermektedir. Diyabete bağlı olarak görme kaybına sebep olan diyabetik retinopati hastalığını erken evrede tespit etmek hastalığın ilerleme riskini azaltarak diyabetli hastaların gelecekte görme kaybı yaşamamasına ve hayat kalitelerinin artmasına yardımcı olacaktır. Bu sebeple, gelecekte hastalıkları teşhis ve tedavi etmede YZ teknolojilerinin sağladığı yöntemle klasik yöntemlerin maliyet etkililiğini analiz eden çalışmaların diyabete bağlı hastalıklarda ve erken teşhisin hayati önem taşıdığı diğer kronik hastalıklarda önemli ölçüde katkılar sağlayacağı ve ön plana çıkacağı düşünülmektedir.

Bu çalışma WoS veri tabanında indekslenen YZ teknolojilerinin sağladığı yöntemle klasik yöntemlerin maliyet etkililiğini inceleyen çalışmalar incelenmiş ve bu çalışmaların duygu durumlarına genel bir bakış sunulmuştur. Gelecekteki çalışmalar bu araştırmanın bulgularından yola çıkarak gelişmiş ve gelişmekte olan ülkelerde hastalık yükü giderek artan kronik hastalıklara odaklanarak YZ'nin sağlık sorunlarının tespitinde ortaya koyduğu yöntemin mevcut yöneme göre maliyet etkililiğinin tespitine odaklanmalıdır.

## 6. SONUÇ (CONCLUSION)

Sağlık hizmetlerinde kullanılacak yeni teknolojiler, uzun vadede maliyet tasarrufu sağlamakla birlikte sağlık hizmeti çıktılarını iyileştirme potansiyeline sahiptir. Bu yüzden sağlık ekonomisi perspektifinden yeni teknolojilerin uygulanması için sağlık teknolojilerinin değerlendirilmesi (STD) gerekmektedir. STD yöntemi, sağlık teknolojilerinin klinik ve maliyet etkililiği başta olmak üzere çok yönlü değerlendirmesini yapmak suretiyle politika belirleyicilere karar alma ve uygulama sürecinde önemli bir girdi ve destek sağlamaktadır. Sağlık hizmetlerinde yeni teknolojilerin uygulanmasının bazı yararlarını ve sakıncalarını maliyet etkililiği perspektifinden göstermek önem arz etmektedir. Bu sebeple gelecekteki çalışmalar, sağlık politikalarında en iyi değeri elde etmek için bilimsel kanıtları artıracak, farklı yöntemler arasında maliyet ve etkililik açısından uygulanabilir bir dengenin sağlanıp sağlanamayacağını gösterecek ve sağlık hizmeti sağlayıcılarının yeni teknolojilerin benimsenmesine ilişkin bilinçli kararlar vermelerine yardımcı olacaktır.

Sonuç olarak, yapay zekân uygulamaları, küresel bazda önemli araştırma konularından biri hâline gelmiş olmakla birlikte diyabet ve kalp gibi kronik hastalıklara bağlı ortaya çıkan hastalıkları erken dönemde teşhis ya da tedavi etmek için bir karar destek sistemi olarak kullanılan YZ'nin sağlık sorunlarının tespitinde ortaya koyduğu yöntemin mevcut yöneme göre maliyet etkililiğini inceleyen bilimsel çalışmaların yapılması gelecekteki araştırmacılar için önerilmektedir. Bununla birlikte farklı alanlarda yeni teknolojilerin getireceği

faydaları değerlendirmek için bu tür çalışmaların kapsamının genişletilmesi de önem arz etmektedir.

## KAYNAKLAR (REFERENCES)

- [1] D. Strusani, G.V. Hounbonon, "The Role of Artificial Intelligence in Supporting Development in Emerging Markets", World Bank Group. 1-8, 2019.
- [2] K. Buntak, M. Kovačić, M. Mutavdžija, "Application of Artificial Intelligence in The Business", International Journal for Quality Research. 15, 403-416, 2021.
- [3] T. Uzun, "Yapay Zeka Ve Sağlık Uygulamaları", İzmir Katip Çelebi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 3, 1, 80-92, 2020.
- [4] J. Borana, "Applications of Artificial Intelligence & Associated Technologies", **Proceeding of International Conference on Emerging Technologies in Engineering, Biomedical, Management and Science [ETEBMS-2016]**, 2016.
- [5] J.S. Suri, Mind of An Innovator. Stalk. Artic. 1, 2022.
- [6] GAO, Artificial Intelligence in Health Care Benefits and Challenges of Machine Learning Technologies for Medical Diagnostics. *National Academy of Medicine*, 1-84, 2022.
- [7] B. Jena, S. Saxena, G.K. Nayak vd. Tumor Characterization Using Radiogenomics in Artificial Intelligence Framework. *J. Cancers* 14, 4052, 2022.
- [8] Internet: Appen, How Artificial Intelligence Data Reduces Overhead Costs for Organizations 2022 State Of AI And Machine Learning Report, <https://appen.com/blog/how-artificial-intelligence-data-reduces-overhead-costs-for-organizations/>, 20.09.2022.
- [9] T. Davenport, & R. Kalakota "The Potential for artificial Intelligence in Healthcare", *Future Healthcare Journal*, 6(2), 94–98, 2019.
- [10] J. Gomez Rossi, N. Rojas-Perilla, J. Krois, & F. Schwendicke, "Cost-effectiveness of Artificial Intelligence as a Decision-Support System Applied to the Detection and Grading of Melanoma, Dental Caries, and Diabetic Retinopathy" *JAMA Network Open*, 5(3), 1-15, 2022.
- [11] X.M. Huang, B.F. Yang, W.L. Zheng, et al. "Cost-Effectiveness of Artificial Intelligence Screening for Diabetic Retinopathy in Rural China" *BMC Health Serv Res*, 22, 260, 1-12, 2022.
- [12] Y. Xie, Q. Nguyen; V. Bellemo, et. all, "Cost-Effectiveness Analysis of an Artificial Intelligence-Assisted Deep Learning System Implemented in the National Tele-Medicine Diabetic Retinopathy Screening in Singapore", *Investigative Ophthalmology & Visual Science*, Vol.60, 5471, 2019.
- [13] M. F. Drummond, M. J. Sculpher, K. Claxton, G. L. Stoddart, & G. W. Torrance, **Methods for the Economic Evaluation of Health Care Programmes.** (4th ed., Oxford University Press, 2015.
- [14] S.D. Shillcutt, D. G. Walker, C. A. Goodman, A. J. Mills, "Cost Effectiveness in Low- And Middle-Income Countries: A Review of The Debates Surrounding Decision Rules", *PharmacoEconomics*, 27(11), 903–917, 2009.
- [15] M. F. Drummond, P. J. Neumann, S. D. Sullivan , F. U. Fricke, S. Tunis, O. Dabbous, &M. Toumi, "Analytic Considerations in Applying a General Economic Evaluation Reference Case to Gene Therapy. Value in Health", *The Journal Of The International Society for Pharmacoeconomics and Outcomes Research*, 22(6), 661–668, 2019.
- [16] J.P. Mueller, L. Massaron, "Artificial Intelligence for Dummies", John Wiley & Sons, Inc., Hoboken, New Jersey, 2018.
- [17] N. Soni, E.K. Sharma, N. Singh, A. Kapoor, "Artificial Intelligence in Business From Research and Innovation to Market Deployment" **Procedia Computer Science**, 167: 2200-2210, 2020.
- [18] Internet: M. Collier, R. Fu, L. Yin, et al. Artificial Intelligence: Healthcare's New Nervous System Dublin, Ireland: Accenture, <https://www.accenture.com/au-en/insights/health/artificial-intelligence-healthcare> 11.09.2022
- [19] T. Bezboruah, A. Bora, "Artificial intelligence: The Technology, Challenges and Applications", *Transactions on Machine Learning and Artificial Intelligence*, 8 (5), 45-51, 2020.
- [20] T. Bezboruah, A. Bora, "Artificial intelligence: The Technology, Challenges and Applications", *Transactions on Machine Learning and Artificial Intelligence*, 8 (5), 45-51, 2020.
- [21] Internet: McKinsey & Company, Global AI Survey: AI Proves its Worth, But Few Scale Impact, <https://www.mckinsey.com/featured-insights/artificial-intelligence/global-ai-survey-ai-proves-its-worth-but-few-scale-impact>, 2 Ekim 2022.
- [22] C. Piccininni, "Cost-Effectiveness of Robotics and Artificial Intelligence in Healthcare", *University of Western Ontario Medical Journal*, 87, 49-51, 2019.
- [23] Internet: Frost & Sullivan. (2016). From \$600 M to \$6 Billion, Artificial Intelligence Systems Poised for Dramatic Market Expansion in Healthcare Mountain View, Calif. <https://www.frost.com/news/press-releases/600-m-6-billion-artificial-intelligence-systems-poised-dramatic-market-expansion-healthcare/> 2.10.2022.
- [24] N.J. Van Eck, L. Waltman, "Software Survey VOSviewer, a Computer Program for Bibliometric Mapping", *Scientometrics*, 84(2), 523–538, 2010.
- [25] J. Fan, Y. Gao, N. Zhao, R. Dai, H. Zhang, X. Feng, G. Shi, J. Tian, C. Chen, B.D. Hambly. S. Bao, Bibliometric Analysis on COVID-19: A Comparison of Research Between English and Chinese Studies. *Front, Public Health*, 8, 477, 1-10, 2020.
- [26] I. Zupic, & T. Čater, "Bibliometric Methods in Management and Organization", *Organizational Research Methods*, 18(3), 429-472, 2015.
- [27] U. Al, U. Sezen, & İ. Soydal, *Hacettepe Üniversitesi Bilimsel Yayınlarının Sosyal Ağ Analizi Yöntemiyle Değerlendirilmesi*, *Hacettepe Üniversitesi Edebiyat Fakültesi Dergisi*, 29(1), 2012.

- [28] M. McBurney, P. Novak, “What is Bibliometrics and Why Should You Care?”, **IEEE International Professional Communication Conference**, 108 - 114. 10.1109/IPCC.2002.1049094, 2022.
- [29] E.C.M. Noyons, “**Bibliometric Mapping as A Science and Research Management Tool**”, DSWO Press, Leiden University, 1999.
- [30] E. Noyons, Bibliometric Mapping of Science in A Policy Context. *Scientometrics* 50, 83–98, 2001.
- [31] M.N. Kurutkan, F. Orhan, “**Sağlık Politikası Konusunun Bilim Haritalama Teknikleri ile Analizi**”, İksad Yayınevi, Ankara, 2018.
- [32] D. Gulmez, K. Ozteke, İ.G. Sedat, “*Uluslararası Dergilerde Yayımlanan Türkiye Kaynaklı Eğitim Araştırmalarının Genel Görünümü: Bibliyometrik Analiz*”, *Eğitim ve Bilim*, 1-27, 2020.
- [33] R. Atenstaedt, “*Word Cloud Analysis of The BJGP*”, *Br J Gen Pract*, 62(596), 148, 2012.
- [34] C. Depaolo, K. Wilkinson, “*Get Your Head into the Clouds: Using Word Clouds for Analyzing Qualitative Assessment Data*”, *TechTrends*. 58, 38-44, 2014.
- [35] B. Liu, “*Sentiment Analysis and Opinion Mining*”, *Synthesis Lectures on Human Language Technologies*. 5:1, 1-167, 2012.
- [36] J. Reis, P. Olmo, F. Benevenuto, H. Kwak, R. Prates, J. “*An, Breaking The News: First Impressions Matter on Online News*”, In *ICWSM '15*, 2015.
- [37] N. Godbole, M. Srinivasaiah, S. Sekine, Large-Scale Sentiment Analysis for News and Blogs. In *International Conference on Weblogs and Social Media*, Denver, CO, 2007.
- [38] K. Z. Aung, N. N. Myo, “Sentiment Analysis Of Students' Comment Using Lexicon Based Approach”, **2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)**, 149-154, 2017.
- [39] S. Chamass, H. Hazimeh, J. Makki, E. Mugellini, & O.A. Khaled, “*Lexicon-Based Sentiment Analysis Approach for Ranking Event Entities*”, *International Journal of Services and Standards*, 12 (2), 126 – 139, 2018.
- [40] S. Sohangir, N. Petty, & D. Wang, “Financial Sentiment Lexicon Analysis”, In: **2018 IEEE 12th International Conference on Semantic Computing (ICSC)**, IEEE, 286–289, 2018.
- [41] V. Bonta, N. K. N. Janardhan, “*A Comprehensive Study on Lexicon Based Approaches For Sentiment Analysis*”, *Asian Journal Of Computer Science And Technology*. 8 (S2), 1–6, 2019.
- [42] S. Choi, H. Park, J. Yeo, S.W. Hwang, “Less is More: Attention Supervision With Counterfactuals for Text Classification”, In: **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, 6695–6704, 2020.
- [43] Internet: T. Bulut, R'da Duygu Analizi Üzerine Vaka Çalışmaları: Case Studies on Sentiment Analysis in R. <https://tevfikbulut.net/rda-duygu-analizi-uzerine-vaka-calismalari-case-studies-on-sentiment-analysis-in-r/>
- [44] B. Liu, M. Hu, & J. Cheng, “Opinion Observer: Analyzing and Comparing Opinions on The Web”, In **Proceedings of the 14th international conference on World Wide Web**, 342-351, 2005.
- [45] F.A. Nielsen, “A new ANEW: Evaluation of A Word List For Sentiment Analysis in Microblogs”, **Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big Things Come in Small Packages 718 in CEUR Workshop Proceedings**, 93-98, 2011.
- [46] S. Mohammad, P. Turney, “Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion
- [46] S. Mohammad, P. Turney, “Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon”, In **Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text**, LA, California, 2010.
- [47] Loughran, Tim & Mcdonald, Bill, “*When Is a Liability NOT a Liability? Textual Analysis, Dictionaries, and 10-Ks*” *The Journal of Finance*, 66, 35 – 65, 2011.
- [48] M.L. Jockers, T. Rosamond, “**Text Analysis with R: For Students of Literature**”, Springer Nature Switzerland, 2nd ed. Cham, Switzerland, 2020.
- [49] H. Kim, “*Sentiment Analysis: Limits and Progress of the Syuzhet Package and Its Lexicons*”, *Digital Humanities Quarterly*, 16 (2), 1-40, 2022.
- [50] R. Bali, D. Sarkar, T. Sharma, “**Learning Social Media Analytics with R**”, Packet Publishing Ltd, Birmingham-Mumbai, 2017.
- [51] A. Gönel, İ. Koyuncu, “*Elimination of Clinical Biochemistry Laboratory Tests Through Artificial Intelligence Programs To Increase Cost-Effectiveness*”, *Journal of Clinical and Analytical Medicine*, 346-349, 2018.
- [52] K. Van Nunen, J. Li, G. Reniers, & K. Ponnet, “*Bibliometric Analysis of Safety Culture Research*”, *Safety Science*, 108, 248–258, 2018.
- [53] Y. Guo, Z. Hao, S. Zhao, J. Gong, F. Yang, “*Artificial Intelligence in Health Care: Bibliometric Analysis*”, *J Med Internet Res*, 22(7), e18228, 2020.
- [54] S. Dinakaran, P. Anitha, “*A Review and Study on AI in Health Care Issues. International Journal of Scientific Research in Computer Science*”, *Engineering and Information Technology*, 281-288, 2018.