ISSN: 1309-1581

AJIT-e

Academic Journal of Information Technology

Volume 16 • Issue 1 • Winter 2025

1 – 31
Araştırma Makalesi
Researh Article

DOI: 10.5824/ajite.2025.01.001.x
The Analysis of The Impact of Companies' Logistics 4.0 Awareness on Digital Transformation and Logistics Capabilities
Nazlıcan DİNDARİK, Yahya FİDAN

32 – 51
Derlem Makale
Review Article

DOI: 10.5824/ajite.2025.01.002.x
A Study of Academic Integrity in Online Distance Education
Bahaddin ERDEM, Songül KARABATAK, Murat KARABATAK,

52 – 68
Araştırma Makalesi
Researh Article

DOI: 10.5824/ajite.2025.01.003.x
Sentiment Analysis On Social Media During Crisis Events: The Case Of Kahramanmaraş Earthquake
Emir YETİMOĞLU, Gökçe KARAHAN ADALI

69 – 87
Araştırma Makalesi
Researh Article

DOI: 10.5824/ajite.2025.01.004.x
Bankacılıkta Müşteri Verilerini Anlama: CRISP-DM Yaklaşımı ve Makine Öğrenimi Uygulamaları
Sezai TUNCA

Supported by
ABA
Akademik Bilişim Araştırmaları Derneği

ISSN: 1309-1581

AJIT-e

Academic Journal of Information Technology

Volume 16
Cilt

Issue 1
Sayı

Winter 2025
Kış

www.ajit-e.org

Doç. Dr. Fatime Neşe KAPLAN İLHAN *Marmara Üniversitesi* *İletişim Fakültesi* *Radyo, Televizyon ve Sinema Bölümü* *Sinema Anabilim Dalı* *Istanbul, Turkey*	**Dr. Öğr. Üyesi Ali Barış KAPLAN** *İstanbul Üniversitesi* *İletişim Fakültesi* *Radyo, Televizyon ve Sinema Bölümü* *Radyo-Televizyon Anabilim Dalı* *Istanbul, Turkey*	**Dr. Öğr. Üyesi Yusuf BUDAK** *Kocaeli Üniversitesi* *İletişim Fakültesi* *Gazetecilik Bölümü* *Bilişim (Bilgisayar Teknikleri ve İletişim)* *Ana Bilim Dalı* *Kocaeli, Turkey*
Dr. Öğr. Üyesi Ali ÖZCAN *Gümüşhane Üniversitesi* *İletişim Fakültesi* *Gazetecilik Bölümü* *Bilişim Enformasyon Teknolojileri Ana Bilim Dalı* *Gümüşhane, Turkey*	**Dr. Öğr. Üyesi Ahmet ÖZTÜRK** *Manisa Celâl Bayar Üniversitesi* *Gördes Meslek Yüksekokulu* *Pazarlama ve Dış Ticaret Bölümü* *Halkla İlişkiler ve Tanıtım Pr.* *Manisa, Turkey*	

International Board of Overseers Uluslararası Danışma Kurulu		
Prof. David Gunkel *Northern Illinois University* *Department of Communication* *Media Studies* *Illinois, USA*	**Prof. Thomas Bauer** *University of Münster* *Islamic and Arab Studies* *Münster, Germany*	**Prof. Umit Sezer Bititci** *Heriot-Watt University* *School of Social Sciences* *Edinburgh Business School* *School of Social Sciences* *Edinburgh, Scotland*
Prof. Ian Ruthven *University of Strathclyde* *Computer and Informati* *on Sciences* *Scottish Informatics and Computer Science Alliance* *Glasgow, Scotland*	**Prof. Angappa Gunasekaran** *California State University* *School of Business and Public Administration (BPA)* *Bakersfield, California*	**Prof. Amjad Hadjikhani** *Uppsala University* *Department of Business Studies* *Uppsala, Sweden*
Prof. Maria Manuela Cruz da Cunha *Escola Superior de Tecnologia - IPCA* *Tecnologias* *Barcelos, Portugal*	**Prof. Sayed Abdul Muneem Pasha** *Jamia Millia Islamia* *Department of Political Science* *Social Sciences* *New Delhi, India*	**Prof. David Benyon** *Edinburgh Napier University* *School of Computing* *Edinburg, Scotland*
Assoc. Prof. Anvarjon Ahmedov Ahatjonovich *Universiti Malaysia Pahang* *Faculty of Industrial Sciences & Technology* *Pahang, Malaysia*	**Dr. Ismet Anitsal** *Missouri State University* *Marketing* *Springfield, USA*	PhD. Charalambos Tsekeris *National Centre for Social Research* *Researcher on* *Digital Sociology* *Athens, Greece*
PhD. Tim Marsh *Griffith University* *Griffith Film School* *Brisbane, Australia*	**PhD. Ayse Goker** *Co-founder, Director at AmbieSense* *Aberdeen, United Kingdom*	**PhD. Charalambos Tsekeris** *National Centre for Social Research* *Researcheron* *Digital Sociology* *Athens, Greece*
	PhD. David Fernández Quijada *Manager of Media Intelligence Service at European Broadcasting Union* *Geneva Area, Switzerland*	

Dergide yayınlanan makalelerde belirtilen görüşler ve fikirler sadece yazar(lar)ın görüşüdür. Yayınlanan içeriklerle ilgili bütün sorumluluklar yazar(lar)a aittir. Yayınlanan eserlerde yer alan tüm içerik kaynak gösterilmeden kullanılamaz.

The opinions and ideas stated in the articles published in the journal are only the opinion of the author (s). All responsibilities regarding the published content belong to the author (s). The published contents in the articles cannot be used without being cited.

Tüm makaleler DOI ve Crossmark ile kayıt altına alınmaktadır.

All articles are registered with DOI and Crossmark.

Tarandığımız İndeksler – Indexes

AJIT-e - Academic Journal of Information Technology
Address: *Kazım Ozalp Sk. No: 15 Kat 2 34740 Şaşkınbakkal / Suadiye / KADIKÖY / ISTANBUL / TURKEY*
Tel: *+90 216 355 56 19*
Faks: *+90 216 368 43 30*
Email*: editor@ajit-e.org*

Supported by
ABA
Akademik Bilişim Araştırmaları Derneği

www.ajit-e.org • **www.abilar.org**

ISSN: 1309-1581

AJIT-e
Academic Journal of
Information Technology

Volume 16 ● Issue 1 ● Winter 2025

Yeni iletişim ortamları hız ve yayın süreçleri açısından yazılı basına göre çok daha avantajlı olduğundan, akademik yayıncılığın geleceği, Internet gibi yeni iletişim ortamları etrafında şekillenmeye başlamıştır. Makaleler dergilerin basılı versiyonlarından önce yayınlanabilmektedir. AJIT-e de iletişim ve bilişim alanına ilgi duyan araştırmalar için bir kaynak ve yayın ortamı sağlamak amacıyla 2010 yılında yayın hayatına başlamıştır.

AJIT-e, uluslararası hakemli bir dergidir. Türkçe ve İngilizce, iki dilde yılda dört sayı yayınlanır. AJIT-e yayın alanları arasında başlıca şu konular yer alır:

Yeni Medya ve İletişim Bilimleri, Teknoloji, Adli Bilişim, Belge ve Kayıt Yönetimi, Bilgi Güvenliği, Bilgi Yönetimi, Bilişim Etiği, Bilişim Hukuku, Dağıtık Bilişim Sistemleri, E-Öğrenme, E-Dönüşüm, E-Devlet, E-Pazarlama, E-Reklam, E-Scm, E-Yayımcılık, E-Yayıncılık, E-Yönetim, Tıp Bilişimi, Karar Destek Sistemleri, Sayısal Eğlence ve Oyun, Sayısal Hak Yönetimi, Sosyal Ağlar, Tedarik Zinciri Yönetimi, Telekomünikasyon, Veri Madenciliği, Veritabanları, Yapay Zekâ, Yönetim Bilişim Sistemleri

●●●

As new communication environments are much more advantageous than print media in terms of speed and broadcast processes, the future of academic publishing has begun to take shape around new communication environments such as the Internet. Articles can be published long before the printed versions of journal. AJIT-e started publication in 2010 to provide a resource and publication environment for research interested in the field of communication and informatics.

AJIT-e is an international refereed journal. It is published four times a year in both languages, in Turkish and English. AJIT-e publication areas include the following topics:

New Media and Communication Sciences, Technology, Computer Forensics, Document and Records Management, Information Security, Information Management, Information Ethics, Distributed Information Systems, E-Learning, E-Transformation, E-Government, E-Marketing, E-Advertisement, E-Scm, E-Publishing, E-Management, Medical Informatics, Decision Support Systems, Digital Entertainment and Gaming, Digital Rights Management, Social Networks, Supply Chain Management, Telecommunications, Data Mining, Databases, Artificial Intelligence, Management information systems

Prof. Dr. Özhan TINGÖY
Editor-in-Chief

ISSN: 1309-1581

AJIT-e
Academic Journal of Information Technology

Volume 16 ● Issue 1 ● Winter 2025

Contents

İçindekiler

1 – 31 **Araştırma Makalesi** **Researh Article**	DOI: 10.5824/ajite.2025.01.001.x **The Analysis of The Impact of Companies' Logistics 4.0 Awareness on Digital Transformation and Logistics Capabilities** *Nazlıcan DİNDARİK, Yahya FİDAN*
32 – 51 **Derlem Makale** **Review Article**	DOI: 10.5824/ajite.2025.01.002.x **A Study of Academic Integrity in Online Distance Education** *Bahaddin ERDEM, Songül KARABATAK, Murat KARABATAK,*
52 – 68 **Araştırma Makalesi** **Researh Article**	DOI: 10.5824/ajite.2025.01.003.x **Sentiment Analysis On Social Media During Crisis Events: The Case Of Kahramanmaraş Earthquake** *Emir YETİMOĞLU, Gökçe KARAHAN ADALI*
69 – 87 **Araştırma Makalesi** **Researh Article**	DOI: 10.5824/ajite.2025.01.004.x **Bankacılıkta Müşteri Verilerini Anlama: CRISP-DM Yaklaşımı ve Makine Öğrenimi Uygulamaları** *Sezai TUNCA*

AJIT-e Academic Journal of Information Technology
10.5824/ajite.2025.01.001.x

2025 . Winter - Kış . Volume – Cilt: 16 . Issue – Sayı: 1

The Analysis of The Impact of Companies' Logistics 4.0 Awareness on Digital Transformation and Logistics Capabilities*

Nazlıcan DİNDARİK, *Kastamonu Üniversitesi, Uluslararası Ticaret ve Lojistik Bölümü, Dr. Öğr. Üyesi, ndindarik@kastamonu.edu.tr, 0000-0003-4313-6576*

Yahya FİDAN, *İstanbul Ticaret Üniversitesi, Lojistik Yönetimi Bölümü, Prof. Dr., yfidan@ticaret.edu.tr, 0000-0002-5012-3629*

ABSTRACT

The study aims to determine the awareness of Logistics 4.0 and examine its impact on businesses' digital transformation and logistics capabilities. A survey was distributed to businesses in the Istanbul Chamber of Commerce's transportation and logistics services committee using a sampling method, and 401 responses were received. It was found that Logistics 4.0 awareness has a significant impact on digital transformation. The sub-dimensions of Logistics 4.0 awareness—business model products and services, process management, information systems infrastructure, and evaluation variables—were found to have significant effects on digital transformation. However, the marketing and customer outreach variables did not have a significant effect on digital transformation. It was also concluded that Logistics 4.0 awareness has a significant effect on logistics capabilities. When examining the impact of the sub-dimensions of Logistics 4.0 awareness on logistics service differentiation and logistics innovation, it was found that marketing and customer outreach, and process management had no significant effect on logistics service differentiation, but did affect logistics innovation. On the other hand, business model products and services, information systems infrastructure, and evaluation variables were found to have a significant effect on logistics service differentiation but no effect on logistics innovation. The study highlights that businesses' awareness of Logistics 4.0 plays a critical role in both accelerating digital transformation processes and developing logistics capabilities. The research contributes theoretically and methodologically by emphasizing that digitalization is not only a technological innovation but also a strategic tool for businesses to enhance their logistics capabilities and gain a competitive advantage.

Keywords : Logistics 4.0, Digital Transformation, Logistics Capabilities, Logistics Service Differences, Logistics Innovation

* This article is derived from the doctoral thesis titled " The Role of Innovation Focus in the Impact of Logistics 4.0 Awareness of Businesses on Digital Transformation and Logistics Capabilities" completed at the Istanbul Ticaret University in 2024.

Araştırma Makalesi / Research Article
Date Received: 28.12.2024
Publishing Date: 28.02.2025

AJIT-e Academic Journal of Information Technology
2024 Winter/Kış – Cilt/Vol: 16 - Sayı/Issue: 1
10.5824/ajite.2025.01.001.x

İşletmelerin Lojistik 4.0 Farkındalığının Dijital Dönüşüm ve Lojistik Yeteneklere Etkisinin Analizi

ÖZ

Çalışmada Lojistik 4.0 farkındalığının belirlenebilmesi ve bu farkındalığın işletmelerin dijital dönüşümü ile lojistik yeteneklerine olan etkisinin incelenmesi amaçlanmıştır. İstanbul Ticaret Odası taşımacılık ve lojistik hizmetleri komitesinde yer alan işletmelere amaca yönelik örnekleme yöntemi ile anket formu ulaştırılmış 401 kişiden geri dönüş alınmıştır. Lojistik 4.0 farkındalığının dijital dönüşüm üzerinde anlamlı bir etkisi olduğu bulunmuştur. Lojistik 4.0 farkındalığının alt boyutları olan iş modeli ürün ve hizmetler, süreç yönetimi, bilgi sistemleri altyapısı ve değerlendirme değişkenlerinin dijital dönüşüm üzerinde anlamlı etkileri olduğu ancak pazarlama müşteriye ulaşma değişkeninin dijital dönüşüm üzerinde anlamlı bir etkisi olmadığı belirlenmiştir. Lojistik 4.0 farkındalığının aynı zamanda lojistik yetenekler üzerinde de anlamlı bir etkisi olduğu sonucuna varılmıştır. Lojistik 4.0 farkındalığının alt boyutlarının lojistik hizmet farklılıklarına ve lojistik inovasyona etkisi incelendiğinde, pazarlama müşteriye ulaşma ve süreç yönetiminin lojistik hizmet farklılıklarına anlamlı bir etkisi olmadığı ancak lojistik inovasyona etkisi olduğu görülmüştür. Diğer yandan, iş modeli ürün ve hizmetler, bilgi sistemleri altyapısı ve değerlendirme değişkenlerinin lojistik hizmet farklılıklarına anlamlı bir etkisi olduğu ancak lojistik inovasyona etkisi olmadığı sonucuna varılmıştır. Çalışma, işletmelerin lojistik 4.0 farkındalığının hem dijital dönüşüm süreçlerini hızlandırmada hem de lojistik yeteneklerin geliştirilmesinde kritik bir rol oynadığını ortaya koymaktadır. Araştırma, dijitalleşmenin yalnızca teknolojik bir yenilik değil, aynı zamanda işletmelerin lojistik yeteneklerini geliştirip rekabet avantajı elde etmede stratejik bir araç olduğuna vurgu yaparak teorik ve metodolojik bir katkı sunmaktadır.

Anahtar Kelimeler : Lojistik 4.0, Dijital Dönüşüm, Lojistik Yetenekler, Lojistik Hizmet Farklılıkları, Lojistik İnovasyon

INTRODUCTION

Developments and transformation processes in various industrial fields have deeply impacted logistics sector activities, just as they have in other sectors. The logistics sector and industrial revolutions mutually shape and influence each other. In particular, the concept of Logistics 4.0, which involves the integration of digital transformation technologies into the logistics sector under the influence of Industry 4.0, has emerged as a significant development. The ongoing industrial revolution brings numerous innovations and advancements, forcing businesses to adapt to these inevitable changes. Digitalization is not only advancing in the manufacturing sector but also gradually progressing across other sectors, including logistics, becoming a key force shaping the future of business. It is evident that businesses that fail to adapt to these changes will fall behind in competition. Therefore, it is crucial for businesses in Turkey's logistics sector to evaluate their awareness of Logistics 4.0. Companies that effectively manage their resources and adapt to digital transformation trends will be better positioned to gain a competitive advantage on both local and global scales. On the other hand, businesses that fail to comply with or effectively integrate into the digital transformation

process will find it increasingly difficult to maintain their competitive strength and contribute to economic development.

Given that the logistics sector is heavily technology-driven, it is essential to assess it within a technological context. Managing the digital transformation process and adapting to it is not merely a necessity, but rather a goal to achieve success by combining creativity and innovation to present the technologies of the future. Digital transformation can be realized as labor-intensive processes are replaced by technology-based, mechanical processes. When considering the reasons for implementing digital transformation, the process's functioning, the challenges encountered, and potential solutions, the changes in procurement processes, the rapid adaptation of competitive and agile businesses, the fragmentation of market shares, the expansion of competitive areas, and customers' increasing expectations for personalized services make digital transformation a necessity. In this context, adapting to the process is crucial for gaining a competitive advantage with the aim of catching up with the future vision. As competition increases, the role of dynamic capabilities in the digital transformation technologies used by logistics businesses becomes increasingly important.

This study aims to examine the relationship between businesses' awareness of Logistics 4.0 and their digital transformation processes, with a particular focus on its impact on logistics capabilities. The study emphasizes the critical role of Logistics 4.0 awareness in promoting digital transformation and logistics capabilities. The primary research question of this study is: How does awareness of Logistics 4.0 influence the development of digital transformation and logistics capabilities in businesses? This research highlights that understanding and integrating the principles of Logistics 4.0 is a crucial strategic tool for innovation, growth, and long-term sustainability within the logistics sector. The study makes a theoretical and methodological contribution by investigating the impact of Logistics 4.0 awareness on digital transformation and logistics capabilities.

Although there has been much discussion regarding the digitalization of logistics and technological advancements, there is limited research directly examining the relationship between Logistics 4.0 awareness and digital transformation and logistics capabilities. By investigating this relationship, this study contributes both theoretically and practically to the logistics sector and academia. Ultimately, this research provides valuable insights into how logistics companies can manage their digital transformation process and its impact on logistics capabilities. Moreover, the study offers recommendations for logistics companies to effectively engage in the digitalization process, shedding light on future trends in the sector.

1. LITERATURE REVİEW

1.1.Logistics 4.0

Industrial revolutions, beginning in the 18th century, represent a series of significant transformations that continuously evolve. These revolutions have brought about radical

10.5824/ajite.2025.01.001.x

changes in production processes and led to profound shifts in the economic, social, and cultural structures of societies. The latest stage of these industrial revolutions is the period we are currently experiencing, known as Industry 4.0 (Kagermann et al., 2011, p. 2). This period has led to the emergence of production methods by enabling the transformation and renewal of business models to increase efficiency in industrial production, ensure flexibility and enhance competitiveness.

The business world requires adapting to new demands in a dynamic and customer-oriented manner and developing innovative strategies to achieve logistical goals. Real-time visibility in logistics has become critically important in order to respond quickly and flexibly to instant demands and changes (Kache & Seuring, 2017, p. 11). The ability to adapt to changing demands based on customer needs and requirements serves as a cornerstone and sustainable driving force in forecasting the evolution of logistics.

The concept of Logistics 4.0 was introduced in 2011 to respond to changing customer needs and develop sustainable logistics solutions (Winkelhaus and Grosse, 2020, p. 18). This new paradigm aims to make the logistics flow more efficient, flexible, and agile through the combination of technologies used in the digitalization process. In this context, the framework of ideas for Logistics 4.0 is detailed in the figure.

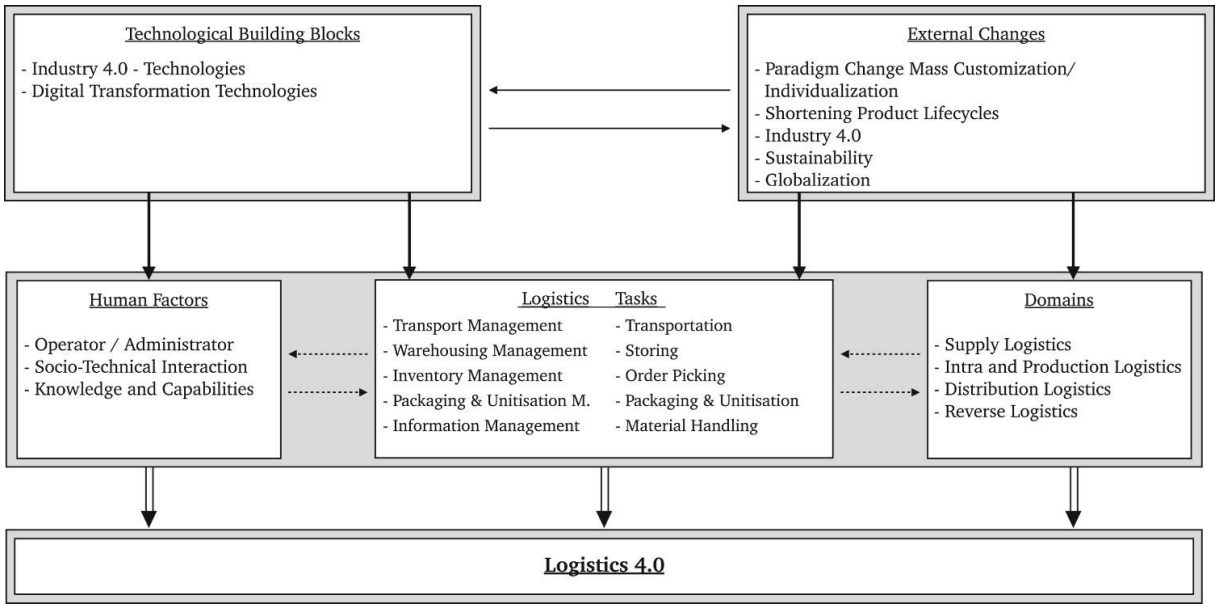


Figure 1: Conceptual Framework of Logistics 4.0 (Winkelhaus and Grosse, 2020, p. 20).

Logistics 4.0 provides certain levels of integration and coordination. It ensures that the information technologies used in different departments within the enterprise are accessible and coordinated at the same time. It also provides an important benefit in terms of cooperation between different businesses. Finally, it is an important management mechanism in

integrating stakeholders, devices and products around the enterprise in a holistic structure (Strandhagen et al., 2017, p. 344).

Other digital transformation technologies utilized through Logistics 4.0, especially blockchain technology, focus on the customer throughout the chain, making the product traceable at every stage and transparently accessible. Thus, the customer can easily access information about where and at what stage the product is, which allows for an increase in the trust relationship along the chain (Thoben et al., 2017, p. 4)

Businesses that see customer satisfaction as a prerequisite are those that have grasped the importance of the concept of logistics 4.0 and have included the customer in decision-making processes by giving the necessary value to customer services in this context. Businesses that aim for effective, efficient, sustainable management with a holistic approach throughout the chain that sees every feedback from the customer as an improvement can control the competitive factor more (Ejsmont et al., 2020, p. 5).

The logistics sector is continually seeking new digital solutions to enhance traceability and predictability. Parallel to technological advancements, the widespread use of mobile sensors has made it possible to track the status and location of loads in real time within logistics processes. This has greatly facilitated the monitoring of trucks and ships, improving operational efficiency. Furthermore, the opportunities provided by cloud technology have made innovative business models like virtual transport feasible. This enables logistics processes to be managed more flexibly and scalable. 3D printers have the potential to bring revolutionary changes to the logistics sector. These printers allow the rapid production of various products and spare parts and their immediate distribution. Thus, delays in the supply chain can be reduced, and changes in demand can be responded to more quickly. Autonomous vehicles also contribute significantly to the digital transformation in logistics. These vehicles increase reliability and minimize human errors by being used in both land and sea transportation processes. The integration of these technologies makes it possible to offer faster, safer, and more cost-effective services in the logistics sector (Gökırmak, 2019, p. 76).

1.2. Digital Transformation

Digital transformation reshapes existing operations and ways of doing business, offering a more efficient, innovative, and customer-centric approach (Verina & Titko, 2019, p. 10). With digital transformation, the concept of "smart" has increased the importance of technology-based solutions. Smart logistics positively impacts the logistics supply chain by offering specialized and value-added services in this context. Smart logistics systems provide fast and effective solutions to the changing demands of customers, preventing the stagnation of the process along the supply chain and making it adaptable (Erdal, 2024, p. 63).

Businesses implement the transformation process by utilizing various applications in the digitalization process. Suppliers, investors, business employees, and managers are among

the individuals using information systems in this process. Each of them supports business processes by using information systems in different ways and levels. Businesses strive to integrate their digitalization strategies in areas such as business processes, social and mobile applications, big data, cloud technologies, and enterprise applications. This integration strengthens not only the internal processes of businesses but also their interactions with external stakeholders and markets. Thus, businesses succeed in expanding their boundaries and effectively participating in a broader ecosystem (Schwertner, 2017, p. 388).

To detail the digital transformation technologies integrated with Logistics 4.0 technologies, they are briefly mentioned below:

The Internet of Things (IoT) provides businesses with the ability to continuously track and manage their products throughout logistics processes. This technology constantly monitors the flow of goods, collecting and analyzing data produced at each stage. Through the data obtained, businesses can optimize the logistics flow and provide real-time information to all stakeholders in the process. As a result, transparency in logistics operations increases, and the efficiency of processes is improved (Demir et al., 2020, p. 16). *Artificial Intelligence (AI)* is a technology that has the ability to imitate human intelligence and learn according to given commands. As intelligent machines develop, AI can analyze data, make predictions, and make decisions based on these analyses. This process involves processing data using statistical models and algorithms to obtain meaningful results. AI, combined with big data and cloud computing infrastructure with high computing power, represents a technology that allows machines to direct critical information. The integration provided achieves more accurate and effective results (Bouanba et al., 2022, p. 445). *Big Data* refers to data that has emerged with the development of information and technological infrastructure, characterized by large volumes, rapid changes, and various types of data. This data is generated by sensors and systems and is processed through operations such as collection, storage, cleaning, visualization, analysis, and interpretation. As a result, data obtained from different sources is analyzed, enabling the use of this information for strategic decision-making (Akdoğan & Akdoğan, 2018, p. 4). *Autonomous vehicles* are a technology integrated with different sensors that enable them to recognize situational changes in their environment. They analyze the surrounding data to make decisions and move independently without human intervention. These vehicles aim to increase safety, improve operational efficiency, and optimize traffic flow through these technologies (Park & Han, 2023, p. 2). *Autonomous robots* can perceive their environment and move without being directly controlled by an operator or restricted to fixed, predetermined paths. These robots, using AI applications, have advanced computing hardware and software. Due to these features, they can make independent decisions in production processes, act on these decisions, communicate with other robots, and exchange data. These capabilities enable more flexible and efficient production processes

(Banger, 2018, p. 46). *Augmented Reality (AR)* is a technology that can be used in many processes, such as logistics and production. AR systems integrate data created in a digital environment with the real world, providing support to employees. For example, the AR-based order-picking method known as pick-by-vision is an innovative application of this technology. This method makes order picking faster and more effective while shortening working time. Thus, AR increases efficiency in logistics processes and optimizes workflows (Demir et al., 2020, p. 16). *Cloud computing* is a vital technology for autonomous control applications in logistics. This technology supports autonomous logistics solutions using intelligent software tools and is based on scalable IT infrastructure provided by cloud service providers. This approach eliminates the need for companies to make significant investments in IT infrastructure, leaving infrastructure management to cloud service providers. Thus, businesses can perform more efficient and cost-effective logistics operations (Demir et al., 2020, p. 16). *Blockchain technology* has a decentralized management structure, supporting real-time and synchronized operations while ensuring the immutability, anonymity (multi-stakeholder structures), and transparency of data. Blockchain is a distributed network accessible to different parties, where transactions such as contract management, shipment tracking, payment processing, and data transfer are conducted. Activities can be monitored at every step, and transparency is maximized through continuous data distribution among parties. This structure enhances security and guarantees the accuracy of transactions (Partala, 2018, p. 2). *Cybersecurity* encompasses measures to protect against unauthorized access and attacks conducted via computers and the internet. Given the high value and volume of data in digital environments, establishing an effective cybersecurity infrastructure is critically necessary. This infrastructure protects data confidentiality, integrity, and accessibility while providing proactive defense against cyber-attacks. For institutions, governments, and individuals, cybersecurity minimizes the impacts of potential threats and creates a secure digital environment by protecting digital assets. Therefore, comprehensive and continuously updated cybersecurity strategies are of great importance (Kökhan, 2021, p. 97).

In conclusion, digital transformation is an inevitable development process for both individuals and businesses. Organizations that fail to adapt to this transformation process are often unable to sustain their existence successfully. The rapidly changing nature of digital technologies compels organizations to keep up with these innovations; otherwise, they risk losing their competitive advantage and being left out of the market. Therefore, for businesses, effectively managing and adapting to the digital transformation process is an important determinant of survival.

1.3. Logistics Capabilities

Logistics capabilities are considered fundamental elements that enable businesses to effectively manage and optimize their logistics processes. A detailed analysis of logistics capabilities provides critical insights into achieving competitive advantage and operational

efficiency. Therefore, research on logistics capabilities holds significant importance in both academic and applied fields.

Numerous studies have been conducted to measure logistics capabilities, resulting in the development of various definitions and different scales. Comprehensive studies and reviews on logistics capabilities have led to the creation of scales that relate these capabilities to competitive advantage and firm performance (Defee & Fugate, 2010, p. 191). As innovative understanding and developments continue to advance, it becomes inevitable for companies to enhance their capabilities and differentiate in logistics service delivery. This requirement is accepted as a goal that must be continuously maintained for companies to remain competitive and adapt to market demands. To succeed in this dynamic environment, companies must continuously innovate and differentiate their services.

In the global trade system, the supply chains of companies are interconnected and span across the globe. Therefore, effective management of the supply chain is a highly valuable area for each company. Logistics companies aim to generate information, share this information with other stakeholders within the organization, and shape their future strategies. Businesses that embrace differentiation to add value to the process and implement it sustainably have greater power.

Businesses that stand out with their logistics capabilities use these skills to build stronger relationships with the end customer and enhance customer loyalty. They also reduce product costs, thereby increasing their profit margins. These advantages help companies gain a competitive edge. Having high logistics capability offers a significant strategic advantage in terms of enhancing customer satisfaction and optimizing costs (Bakan & Şekkeli, 2015, p. 398).

The variability in global competition and economic differences require businesses to make the most extensive use of their existing logistics capabilities in order to remain competitive. (Karagöz & Akgün, 2015, p. 23). Companies can facilitate the enhancement of their logistics capabilities to a higher level by broadening their perspectives and adopting innovation as a benchmark.

Logistics capability consists of two sub-dimensions. After explaining the concept of logistics capability, its sub-dimensions, namely logistics innovation and logistics service differentiation, will be elaborated on.

1.3.1. Logistics Service Differentiation

The impact of globalization, the diversification of economic partnerships and relationships, the spread of trade between countries, and the increased interdependence among nations have made logistics service differences a prominent concept both nationally and globally. Companies can achieve sustainable competitive advantage only by conducting a proper capability analysis and identifying service differentiations that adapt to logistics

processes. This adaptation and differentiation process makes it possible to manage logistics processes effectively. Additionally, comparing the results of logistics activities with competitive factors is also considered within the scope of logistics service differentiation (Langley et al., 1992, p. 2).

Since the logistics sector has a dynamic, developing, and complex structure, businesses are expected to be different from their competitors in the sector and to provide a service that provides added value. In this direction, businesses that want to gain competitive advantage should read their capabilities correctly, be open to discovery, and be based on sustainable innovation. Only if they take these elements into account, they should be aware that they can add value by offering logistics service differences to the customer. In a changing demand environment, the more open to change and the wider the service possibilities, the more it is expected to be ready to offer differences (Kozan, 2019, p. 20).

Services provided at the time, place, and conditions determined based on customer demands and needs show a direct correlation with customer satisfaction. In line with the "seven rights" of logistics, effectively responding to customer demands, managing processes with quality, and ensuring complete and damage-free deliveries lead to variations in how service quality is demonstrated, depending on how the service is presented. E-commerce, the improvement of logistics services, and the offering of personalized solutions increase the perceived value for customers and highlight the quality brought by service differentiation. These differentiated logistics services are considered a key pillar of customer satisfaction and provide businesses with a competitive advantage.

1.3.2. Logistics Innovation

Among the logistics capabilities, the ability to innovate in logistics refers to the skills shaped by innovative ideas, inventions, and thoughts related to logistics activities within the enterprise. These innovations can be implemented in various ways, such as developing a new product, differentiating existing products, or making products and services more beneficial. Such innovations offer a structure open to improvement, not only increasing efficiency but also addressing unresolved issues and better responding to customer needs. The ability to innovate in logistics creates unique and hard-to-imitate capabilities, enhancing the likelihood of achieving sustainable competitive advantage. Therefore, businesses are encouraged to develop this skill area and conduct research within this scope (Kallio et al., 2012; Yang et al., 2009).

The ability to innovate in logistics is the ability of firms to make changes in their logistics services to gain competitive advantage, increase efficiency, and raise profitability. This capability includes the skills to bring innovations in areas such as planning, strategy development, program creation, method determination, selection of transportation modes, cost management, inventory control, and information technologies (Bakan & Şekkeli, 2015, p. 402). Innovation is critically important for the success of many enterprises, including logistics companies, as these processes enable firms to remain competitive and respond effectively to market demands.

10.5824/ajite.2025.01.001.x

Logistics innovation is a function that promotes performance-enhancing operational processes aimed at the market, benefiting both service providers and consumers while creating value with a sustainable efficiency understanding. This innovation aims to reduce costs while also serving as a differentiating factor for customers in today's highly competitive environment (Demirdöğen et al., 2018, p. 145).

Innovation is considered by businesses as a valuable capability and when combined with logistics activities, it makes it more difficult to deal with similar practices or imitations by competitors (Yang et al., 2009, p. 4-5). Therefore, when logistics innovation is correctly perceived and implemented, it emerges as a unique factor that provides added value and competitive advantage to businesses.

2. METHODOLOGY

2.1.Hypotheses Development and Research Model

The two main hypotheses analyzed in the current study, along with eighteen sub-hypotheses and their justifications, are as follows:

H1: Awareness of Logistics 4.0 has a significant impact on digital transformation.

There are numerous studies in the literature regarding the impact of Logistics 4.0 awareness on digital transformation. Öztemel and Gürsev (2018) emphasize the importance of digital transformation technologies while stating that the structure of the enterprise should be considered when selecting appropriate technologies. In this context, awareness of Logistics 4.0 enables businesses to improve their processes by influencing which digital transformation strategies they adopt. Additionally, researchers such as Witkowski (2017) highlight that data and IoT elements enhance the innovation potential in the logistics sector, demonstrating that Logistics 4.0 awareness is a supporting factor for digital transformation. Şekkeli and Bakan (2018) reveal the potential of Logistics 4.0, underline its critical importance, and discuss the benefits of its proper implementation. Büyüközkan and Güler (2019) define the existing technologies within the scope of Logistics 4.0, list the expectations related to digital transformation technologies, and present a method for analyzing these technologies. Lagorio et al. (2020) conducted a literature review aimed at identifying gaps to understand how technological trends and developments are progressing in the logistics sector. Studies conducted within this scope indicate that businesses' perspectives on Logistics 4.0, their recommendations regarding their current situations, and the digital technologies they initially consider investing in are subjects of ongoing research.

H1a: Logistics 4.0 has a significant impact on digital transformation.

H1b: Business models, products, and services have a significant impact on digital transformation.

H1c: Marketing and customer outreach have a significant impact on digital transformation.

H1d: Process management has a significant impact on digital transformation.

H1e: Information systems infrastructure has a significant impact on digital transformation.

H1f: Evaluation has a significant impact on digital transformation.

H1a, b, c, d, e, f: A literature review on the effects of the sub-hypotheses of Logistics 4.0 awareness on digital transformation reveals the study by Taş and Alagöz (2021), which aimed to determine the Industry 4.0 awareness levels of logistics firms. Their study found that firms, regardless of whether they actively use Logistics 4.0 technologies, possess a certain level of awareness but are not at the desired level in terms of implementation. Based on this, they provided recommendations for improvement. Similarly, in his study, Çiçekli S. (2020) assessed the awareness and implementation levels of Logistics 4.0 at the Ankara Logistics Base, identified deficiencies, and offered suggestions to enhance the level of implementation. Atlı (2022) conducted an exploratory case study in logistics companies to examine the perception of Logistics 4.0, its current stage, awareness levels, and key considerations regarding the topic. His findings indicated that logistics companies have begun to increase their awareness, are engaging in technological transformation efforts, are establishing R&D centers, and have recognized the significance of technological adaptation training.

H2: Awareness of Logistics 4.0 has a significant impact on logistics capabilities.

The effect of Logistics 4.0 awareness on logistics capabilities, in relation to the H2 hypothesis, is supported by studies such as those by Çetinkaya (2021) and Sarıhan (2021). Çetinkaya's findings show that Industry 4.0 awareness increases innovation. Sarıhan (2021) has determined that supply chain performance is most affected by logistics capabilities and buyer-supplier relationships. This indicates that Logistics 4.0 awareness enhances logistics capabilities, thereby strengthening the competitiveness of businesses. In his study, Kozan (2019) focused on the concept of logistics capabilities and concluded that its impact on performance from a logistics perspective is significant. Based on these studies, it has been determined that the relationship between logistics capabilities and various variables, especially technological advancements, has been investigated. However, no study measuring the effect of Logistics 4.0 awareness on logistics capabilities has been found.

H2a: Logistics 4.0 has a significant impact on logistics service differentiation.

H2b: Business model, products, and services have a significant impact on logistics service differentiation.

H2c: Marketing and customer outreach have a significant impact on logistics service differentiation.

10.5824/ajite.2025.01.001.x

H2d: Process management has a significant impact on logistics service differentiation.

H2e: Information systems infrastructure has a significant impact on logistics service differentiation.

H2f: Evaluation has a significant impact on logistics service differentiation.

H2a,b,c,d,e,f: The subdimensions of Logistics 4.0 awareness have a significant impact on logistics service differentiation. In the research examining the effect of Logistics 4.0 awareness on logistics service differentiation, studies such as Karadeniz and Başaran (2014) emphasize the impact of information systems on service perception. This highlights the role of information systems infrastructure in determining logistics service differentiation. Additionally, Yılmaz (2020) addressed the importance of the business model in the perception of products and services, which aligns with the current findings.

H2g: Logistics 4.0 has a significant impact on logistics innovation.

H2h: Business model, products, and services have a significant impact on logistics innovation.

H2ı: Marketing and customer outreach have a significant impact on logistics innovation.

H2i: Process management has a significant impact on logistics innovation.

H2j: Information systems infrastructure has a significant impact on logistics innovation.

H2k: Evaluation has a significant impact on logistics innovation.

H2g,h,ı,i,j,k: The subdimensions of Logistics 4.0 awareness have a significant impact on logistics innovation. The effect of Logistics 4.0 awareness on logistics innovation is supported by the study of Çetinkaya (2021). This study shows that some subdimensions of innovation are effective, while the current research demonstrates that process management and information systems infrastructure contribute to logistics innovation.

In the current research, a causal research model has been used to investigate the effects between variables based on the hypotheses developed from the literature.

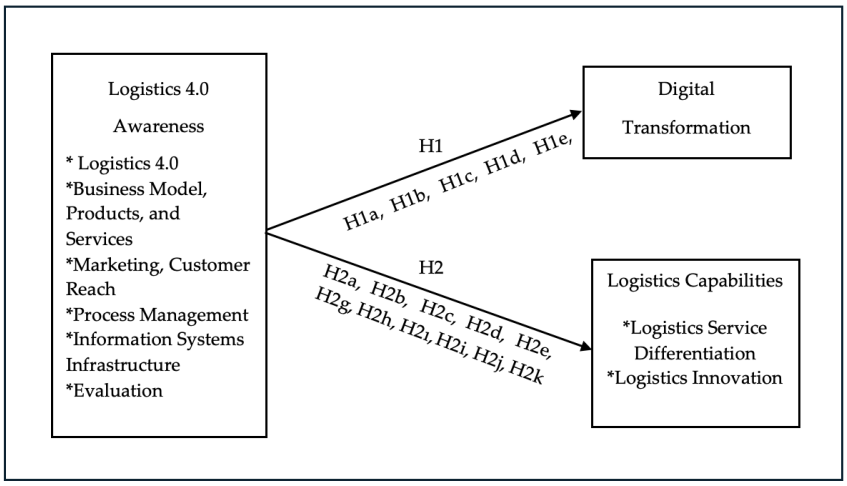


Figure 2: Research Model

The theoretical explanations regarding the impact of Logistics 4.0 awareness on digital transformation and logistics capabilities are provided in the following section. A model has been established to examine the relationship between Logistics 4.0 awareness, which has emerged from Industry 4.0, and its effect on digital transformation and logistics capabilities. The study includes not only theoretical analyses but also the development of a model that will allow for the testing of the proposed hypotheses. In this way, the strategic importance of Logistics 4.0 awareness is emphasized, and the effective role of digital transformation technologies in logistics processes is addressed within a holistic framework of logistics capabilities. The research model designed to carry out these analyses is shown in Figure 2.

2.2.Sampling and Data Collection Method of the Research

Within the scope of the study, a specially prepared survey form was created by conducting a literature review to collect data. The survey form contains pre-determined questions that meet the needs of the research and cover the scales of Logistics 4.0, digital transformation, and logistics capabilities. The prepared questionnaire was directed to participants working in different positions at businesses in the Logistics Services Committee of the Istanbul Chamber of Commerce. To evaluate factors such as the comprehensibility of the questionnaire, the order of the questions, response time, and reliability, a pre-test was first applied to 20 logistics businesses. As a result of the pre-test, minor adjustments were made to the questionnaire, and it was finalized. Based on the population of the conducted study, the large number of logistics businesses operating in Turkey meant that only businesses that are members of the Istanbul Chamber of Commerce Logistics Committee were selected, which defines the scope and limitations of the research. The fact that not all members provided feedback on the survey and the deficiencies in some members' contact information are among the constraints.

The population of the study consists of logistics enterprises that are members of the transportation and logistics services committee of the Istanbul Chamber of Commerce.

10.5824/ajite.2025.01.001.x

According to the information on the website of the Istanbul Chamber of Commerce, 10,982 enterprises are members of this committee. In accordance with the purpose of the research, sampling was made among these enterprises, and logistics enterprises operating in the international market with high market volume were selected. Purposeful sampling refers to selecting information-rich situations for in-depth research in line with the aim of the study. It is preferred in specific cases that possess certain characteristics and criteria (Büyüköztürk et al., 2016, p. 92). A questionnaire was used as the data collection method. The questionnaires were sent to the selected enterprises via e-mail and they were asked to fill them in online. In the research, the minimum sample size to represent the universe was calculated as 385, but this number was increased to 418. Due to missing data in the questionnaires, 17 forms were considered invalid, and the remaining 401 forms were found to be suitable for analysis. It can be said that this number is sufficient for making generalizations about the population.

Table 1: Sample Size

Population Size	**Sample Size** (95% confidence level)	**Percentage of Population in the Sample**	**Population Size**	**Sample Size** (99% confidence level)	**Percentage of Population in the Sample**
50	44	88,0	200	171	85,5
100	80	80,0	500	352	70,4
500	222	44,4	1.000	543	54,3
1.000	286	28,6	2.000	745	37,2
5.000	370	7,4	5.000	960	19,2
10.000	385	3,9	10.000	1061	10,6
100.000	398	0,4	20.000	1121	5,6
1.000.000	400	0,04	50.000	1160	2,3
10.000.000	400	0,004	100.000	1173	1,2

Source: Prepared using Statistics Canada, 2010, p. 155; Neuman, 2010, p. 351.

As seen in the table, the minimum sample size required to represent the population within a 95% confidence interval has been calculated as 385. The purposive sampling technique was used in the study, and employees working in logistics businesses were selected as survey participants.

As a result of the literature review, it was determined that different scales have been used to measure Logistics 4.0 awareness. In this study, the Industry 4.0 Awareness Survey developed by Veysi İşler (2017) for measuring Industry 4.0 awareness in the manufacturing sector was used as the basis. The dimensions in the scale were adapted to the logistics sector and adjusted to fit the purpose of the research. The questionnaire consists of a total of 33 questions and includes 6 subdimensions: • Logistics 4.0: 5 questions • Business model, products, and services: 4 questions • Marketing and customer outreach: 4 questions • Process management: 6 questions • Information systems infrastructure: 7 questions • Evaluation: 7 questions. The statements in the scale were adapted to the logistics sector context for the

purpose of analyzing Logistics 4.0 awareness. The questions asked to the participants were evaluated using a 5-point Likert scale (1: Strongly Disagree – 5: Strongly Agree).

The Turkish adaptation of the Digital Transformation Scale, created with 12 expressions and a 5-point Likert scale based on various studies (Gudergan & Mugge, 2017; Jafarzadeh et al., 2015; Kane et al., 2016; Lansiti & Lakhani, 2014; Svahn et al., 2017) by Nadeem et al., (2018) and conducted by Sağlam (2021), was used in the study. The questions were prepared with a 5-point Likert scale.

To measure logistics capability, the Logistics Innovation Scale -5 question- (Anderson & West, 1998) and the Logistics Service Differentiation Scale -4 question- (Song & Parry, 1997; Adaptation Lynch et al., 2000) were used. The questions were prepared with a 5-point Likert scale.

3. FINDINGS

3.1.Descriptive Statistics Regarding Participant Businesses

In determining the research sample, demographic information regarding the participants' position in the business, work experience, education level, and gender has been provided. Additionally, information about the business's duration of operation, the digital transformation technologies used in the business, and the areas of activity where these technologies are effective has also been included. The distribution of variables in this regard is shown in the tables.

Table 2: Distribution of Participants According to Demographic Variables

Demographic Variables		Frequency	Percentage (%)
POSITION	Company Owner	45	11,2
	General Manager	68	17,0
	Logistics Employee	184	45,9
	Import/Export	44	11,0
	R&D	14	3,5
	Manager	15	3,7
	Other	31	7,7
Total		**401**	**100,0**
WORK EXPERIENCE	1-5 years	258	64,3
	6-10 years	65	16,2
	11-15 years	39	9,7
	Over 15 years	39	9,7
Total		**401**	**100,0**
EDUCATION	High School	18	4,5
	Associate Degree	54	13,5
	Bachelor's Degree	225	56,1
	Postgraduate	104	25,9
Total		**401**	**100,0**
GENDER	Female	92	22,9
	Male	309	77,1
Total		**401**	**100,0**

In the research, when examining the participants' positions within the company, it was found that a large portion, 45.9%, are logistics employees. The second-highest percentage, 17.0%, is made up of general managers. Following them, 11.2% are business owners, and slightly behind them, 11.0% work in export/import positions. The percentage of participants in managerial positions is 3.7%. It was found that 3.5% of the participants work in R&D roles. Finally, employees working in various positions in logistics companies in the 'other' category make up 7.7% of the total.

When considering the work experience of the participants in logistics businesses, the majority, 64.3%, have been employed for 1-5 years. 16.2% have worked for 6-10 years, while 9.7% have worked for 11-15 years and over 15 years. Based on this, it can be concluded that the majority of employees in logistics companies have worked between 1 and 5 years.

Regarding the education levels of employees in logistics businesses, 4.5% have a high school education, 13.5% have an associate degree, 56.1% have a bachelor's degree, and 25.9% have a graduate degree. This shows that the majority of employees in logistics companies (95.5%) have university-level education. The number of employees with graduate degrees is also significant.

When examining the gender distribution of employees in logistics businesses, 77.1% are male, while 22.9% are female. Based on this distribution, it can be concluded that logistics businesses employ more male employees than female employees.

Table 3: Distribution of Businesses According to Their Duration of Operation

Operational Duration	Frequincy	Percentage
1-5 years	69	17,2
6-10 years	45	11,2
11-15 years	48	12,0
More than 15 years	239	59,6
Total	**401**	**100,0**

When participants were asked how many years their businesses have been operational, it was found that 17.2% of them have been in operation for 1-5 years, 11.2% for 6-10 years, 12.0% for 11-15 years, and 59.6% for more than 15 years. Based on this data, it can be seen that a large proportion of logistics businesses have been operational for more than 15 years. While the majority of employees work for 1-5 years, the businesses themselves have been operational for more than 15 years.

Table 4: Distribution of Businesses According to the Digital Transformation Technologies They Use

Digital Transformation Technologies	Value	Frequency	Percentage
Internet of Things (IoT)	Yes	188	46,9
	No	213	53,1
Big Data	Yes	173	43,1
	No	228	56,9

Cloud Computing	Yes	244	60,8
	No	157	39,2
Blockchain	Yes	83	20,7
	No	318	79,3
Cybersecurity	Yes	219	54,6
	No	182	45,4
Other	Yes	27	6,7
	No	374	93,3

When logistics business employees were asked about the digital transformation technologies used in their companies, 244 people indicated cloud computing, 219 people indicated cybersecurity, 188 people indicated the Internet of Things, 173 people indicated big data, 83 people indicated blockchain, and 27 people selected the 'other' option, referring to traditional methods. Based on this, it can be concluded that cloud computing is the most widely used digital transformation technology in businesses. Some employees selected multiple options, indicating that they use more than one technology.

Table 5: Distribution of Business Activities Affected by the Technologies Used

Affected Activity Area	Value	Frequency	Percentage
Storage	Yes	265	66,1
	No	136	33,9
Stock Management	Yes	236	58,9
	No	165	41,1
Transportation	Yes	303	75,6
	No	98	24,4
Packaging	Yes	143	35,7
	No	258	64,3
Customer Service	Yes	242	60,3
	No	159	39,7
Other	Yes	13	3,2
	No	388	96,8

When participants were asked in which logistics activity areas the technologies used in their businesses were effective, it was emphasized that they were most effective in transportation, with 303 people responding in this regard. Following that, 265 people indicated storage, 242 people indicated customer services, 236 people indicated inventory management, 143 people indicated packaging, and 13 people stated that technologies were effective in other areas. Based on the distribution of the responses, it can be inferred that the technologies used are effective in multiple activity areas.

Data Analysis In testing the developed hypotheses, regression analysis was performed using the SPSS program (Muñoz et al., 2017, p. 79). According to Hair et al. (2014), regression analysis is used to determine the degree to which independent variables explain the dependent variable. In this context, analyses were conducted for the hypotheses identified in the research (p. 204).

10.5824/ajite.2025.01.001.x

To analyze the data used in the research, the SPSS 23.0 program was utilized. Factor analysis and Cronbach's alpha analysis were performed for the reliability and validity of the scales of logistics 4.0 awareness, digital transformation, and logistics capabilities.

3.2.Reliability Analysis

To conduct the research, a survey form was prepared using three different scales and applied to employees in the logistics sector.

Table 6. Reliability Analysis of Scales

Scales	Cronbach's Alpha	Number of Items (N)
Logistics 4.0 Awareness Scale	**,954**	**33**
Digital Transformation Scale	**,971**	**12**
Logistics Capabilities Scale	**,957**	**9**

The scales used in the study are measurement tools that have been previously tested and validated for validity and reliability in the literature (İşler, 2017; Sağlam, 2021; Anderson & West, 1998; Song & Parry, 1997; Adaptation: Lynch et al., 2000). In the study, Exploratory Factor Analysis (EFA) was applied to assess the factor structure of the scales, and the results were consistent with previous studies in the relevant literature. However, since the factor structure and subdimensions of the scales were thoroughly validated in previous research, an additional Confirmatory Factor Analysis (CFA) was not performed.

Additionally, Cronbach's Alpha coefficients were calculated to determine the internal consistency of the measurement instrument. According to the reliability analysis results presented in Table 6, the Cronbach's Alpha values were calculated as 0.954 for the logistics 4.0 awareness scale, 0.971 for the digital transformation scale, and 0.957 for the logistics capabilities scale. A Cronbach's Alpha coefficient of 0.70 and above indicates that the scale is reliable (Durmuş et al., 2022, p. 89). In this regard, it was determined that all the scales used in the study were highly reliable.

3.3.Normality Analysis

An analysis was conducted to assess whether the logistics 4.0, digital transformation, and logistics capabilities scales used in the research show a normal distribution.

Table 7: Normality Analyses of Scales

Scales	Skewness	Kurtosis
Logistics 4.0 Awareness Scale	-0,186	-0,453
Digital Transformation Scale	-0,582	-0,188
Logistics Capabilities Scale	-0,476	-0,157

When the skewness and kurtosis values of the variables in the study were analyzed, it was concluded that the data were normally distributed for three scales (Tabachnick and Fidell, 2013).

3.4.Analysis of Research Hypotheses

H1: Logistics 4.0 awareness has a significant effect on digital transformation."

Table 8: Regression Analysis Results for H1 Hypothesis

Model 1	R^2	Adjusted R^2	F	Std. Error	Beta	T	P
Logistics 4.0 Awareness	,467	,466	349,659	,709	,683	18,699	0,000

Dependent Variable: Digital Transformation, (*p<0.05)

The regression analysis model conducted to measure the impact of logistics 4.0 awareness on digital transformation is found to be significant. Logistics 4.0 awareness has a significant (p<0.05) and positive effect (R^2: 46%) on digital transformation. The H1 hypothesis is accepted. The explanatory power of the created model is 0.466, indicating that 46% of the digital transformation variable can be explained by the logistics 4.0 variable, and this effect is found to be at a moderate level.

In the continuation of the study, the effect of logistics 4.0, business model products and services marketing customer access, process management, information systems infrastructure, evaluation variables, which are the sub-dimensions of logistics 4.0 awareness, which is the independent variable in the sub-hypotheses of hypothesis H1, on the digital transformation dependent variable was evaluated separately. Multiple linear regression analysis was used to evaluate this effect (Kayaalp, et al., 2015, p. 1).

During the testing of sub-hypotheses, the ANOVA table was evaluated in the first stage. As a result of the evaluation, the model was found to be significant (p:0.000), and multiple regression analysis was conducted.

Table 9: Multiple Regression Analysis Results for H1a, b, c, d, e, f Hypotheses

Model 1	R^2	Adjusted R^2	F	Std. Error	Beta	T	P	VIF	Tolerance
Logistics 4.0				,052	,172	3,207	,001	2,137	,468
Business Model, Products and Services				,072	,152	2,659	,008	2,425	,412
Marketing and Customer Reach				,065	,034	,591	,555	2,487	,402
	,471	,463	58,459						

10.5824/ajite.2025.01.001.x

Process Management	,075	,172	2,419	,016	3,769	,265
Information Systems Infrastructure	,082	,183	2,606	,010	3,663	,273
Evaluation	,063	,104	2,241	,026	1,612	,621

Dependent Variable: Digital Transformation, (*p<0.05)
Durbin-Watson: 1.940

In an undesirable situation where there is a deviation from the assumption in the regression analysis performed, there is a linear relationship between two or more independent variables in the multicollinearity problem. In this case, the regression model may give contradictory results in some cases. A method used to detect multicollinearity is the variance inflation factor (VIF). VIF measures the degree of correlation among independent variables. If the VIF value of an independent variable is 10 or greater, it is considered that there is a multicollinearity problem. In such cases, independent variables are linear, and the reliability of the regression model is questioned (Salkin and Rasmussen, 2007).

Since VIF and tolerance values indicate whether there is a multicollinearity problem, they also show the suitability of the model for testing the hypotheses. In the model, it is seen that the VIF values of the hypotheses are less than 10 and the tolerance values are greater than 0.2. Therefore, it has been found that there is no multicollinearity problem.

Another important point to consider in regression analysis is the assumption that there should be no relationship among the error terms. If this relationship exists, autocorrelation arises. The reasons for autocorrelation include incorrect model selection, accepting non-linear relationships as linear, excluding important independent variables from the model, and strong relationships among independent variables. As seen in the model, the Durbin-Watson test was conducted to detect the presence of autocorrelation. According to the test, if the DW value is between 1.5-2.5, it is concluded that there is no autocorrelation. According to the analysis, this value (1.940) is close to 2, indicating no autocorrelation (İmir, 1986). In this context, the hypotheses in the model are considered suitable for testing.

In Table 4, the effect of the logistics 4.0 variable on the digital transformation variable, which is the H1a hypothesis, was analyzed. According to the analysis results, it was found to be statistically significant (p<0.05, Beta: 0.172). In the H1b hypothesis, when the impact of the business model, products, and services variable on the digital transformation variable was analyzed, it was found to be statistically significant (p<0.05, Beta: 0.152). In the H1c hypothesis, the effect of marketing and customer reach on digital transformation (p: 0.555) was analyzed, and the findings indicate that it is not statistically significant (p>0.05). In the H1d hypothesis,

it is stated that process management (p<0.05, Beta: 0.172) has an impact on digital transformation. In the H1e hypothesis, the impact of the information systems infrastructure variable on the digital transformation variable was suggested, and according to the analysis results (p<0.05, Beta: 0.183), its effect was accepted. In the H1f hypothesis, it was found that evaluation (p<0.05, Beta: 0.104) has an impact on digital transformation.

As a result of the regression analysis, it was determined that the sub-dimensions of the logistics 4.0 awareness variable, namely logistics 4.0, business model products and services, process management, information systems infrastructure, and evaluation, statistically significantly affect digital transformation, and the H1a, H1b, H1d, H1e, H1f hypotheses were accepted. Based on the analyses, it was concluded that the marketing and customer reach variable does not statistically significantly affect digital transformation, and the H1c hypothesis was rejected.

The explanatory power of the accepted hypotheses of the created model is 0.46, indicating that 46% of the digital transformation variable can be explained by the logistics 4.0, business model products and services, process management, information systems infrastructure, and evaluation variables. This level of effect is moderate.

"**H2:** Logistics 4.0 awareness has a significant impact on logistics capabilities."

Table 10: Regression Analysis Results for H2 Hypothesis

Model 1	R^2	Adjusted R^2	F	Std. Error	Beta	T	P
Logistics 4.0 Awareness	,375	,373	238,990	,760	,612	15,459	0,000

Dependent Variable: Logistic Capabilities (*p<0.05)

Logistics 4.0 awareness was found to have a significant (p<0.05) and positive effect (R^2: 37%) on logistics capabilities. The H2 hypothesis was accepted. The explanatory power of the created model is 0.373, indicating that 37% of the logistics capabilities variable can be explained by the logistics 4.0 variable. This effect was found to be at a low level.

In the continuation of the study, the sub-hypotheses of H2 were evaluated separately for the impact of the sub-dimensions of the independent variable logistics 4.0 awareness, including logistics 4.0, business model products and services, marketing and customer reach, process management, information systems infrastructure, and evaluation, on the sub-dimensions of the dependent variable logistic capabilities, which are logistic innovation and logistic service differentiation. Multiple linear regression analysis was used to evaluate this effect. As part of the research, analyses were conducted for two sub-hypotheses related to the sub-dimensions of logistics capabilities and six sub-hypotheses related to the sub-dimensions of logistics 4.0 awareness.

10.5824/ajite.2025.01.001.x

Table 11: Multiple Regression Analysis Results for H2a,b,c,d,e,f Hypotheses

Model 1	R^2	Adjusted R^2	F	Std. Error	Beta	T	P	VIF	Tolerance
Logistics 4.0				,060	,179	3,074	,002	2,134	,468
Business Model, Products and Services				,084	,154	2,484	,013	2,425	,412
Marketing, Customer Reach				,075	,048	,757	,449	2,487	,402
	,375	,365	39,368						
Process Management				,087	,060	,772	,440	3,769	,265
Information Systems Infrastructure				,095	,189	2,475	,014	3,663	,273
Evaluation				,073	,105	2,074	,039	1,612	,621

Dependent Variable: Logistic Service Differentiation (*p<0.05)
Durbin-Watson: 1.848

To understand if there is a multicollinearity problem among the variables used in the model, the VIF and tolerance values were examined. The VIF values of the H2a, H2b, H2c, H2d, H2e, and H2f hypotheses were found to be less than 10, and the tolerance values were greater than 0.2, indicating no multicollinearity problem.

The Durbin-Watson test result (1.848) suggests that there is no autocorrelation between the error terms. Therefore, the hypotheses in the model were deemed suitable for testing.

The regression analysis results indicate that the H2a hypothesis, suggesting the impact of the logistics 4.0 variable on the logistic service differentiation variable, is significant (p: 0.002, Beta: 0.179) and was accepted. Similarly, the H2b, H2e, and H2f hypotheses, suggest that business model products and services (p: 0.013, Beta: 0.154), information systems infrastructure (p: 0.014, Beta: 0.189), and evaluation (p: 0.039, Beta: 0.105) have a significant impact on logistic service differentiation, were accepted. However, the H2c and H2d hypotheses, suggesting that marketing and customer reach (p: 0.449, Beta: 0.048) and process management (p: 0.440, Beta: 0.060) impact logistic service differentiation, were rejected.

The regression analysis concluded that the sub-dimensions of logistics 4.0 awareness, including logistics 4.0, business model products and services, information systems infrastructure, and evaluation, significantly affect logistic service differentiation, accepting the

H2a, H2b, H2e, and H2f hypotheses. Meanwhile, the marketing and customer reach and process management dimensions were found to have no significant effect on logistic service differentiation, leading to the rejection of the H2c and H2d hypotheses.

The explanatory power of the accepted hypotheses in the model is 0.365, indicating that 36% of the logistic service differentiation variable can be explained by the variables logistics 4.0, business model products and services, information systems infrastructure, and evaluation. This effect level is considered low.

The next part of the research involves analyzing the impact of the sub-dimensions of logistics 4.0 awareness on another sub-dimension of logistic capabilities, logistic innovation, with six sub-hypotheses.

Table 12: Multiple Regression Analysis Results for H2g,h,ı, i, j, k Hypotheses

Model 1	**R^2**	**Adjusted R^2**	**F**	**Std. Error**	**Beta**	**T**	**P**	**VIF**	**Tolerance**
Logistics 4.0				,060	,097	1,593	,112	2,137	,468
Business Model, Products and Services				,084	,081	1,244	,214	2,425	,412
Marketing, Customer Reach				,076	,151	2,296	,022	2,487	,402
	,313	,303	29,943						
Process Management				,088	,190	2,339	,020	3,769	,265
Information Systems Infrastructure				,096	,075	,944	,346	3,663	,273
Evaluation				,074	,070	1,311	,191	1,612	,621

Dependent Variable: Logistic Innovation (*p<0.05)
Durbin-Watson: 1.702

To understand if there is a multicollinearity problem among the variables used in the model, the VIF and tolerance values were examined. The VIF values of the H2g, H2h, H2ı, H2i, H2j, and H2k hypotheses were found to be less than 10, and the tolerance values were greater than 0.2, indicating no multicollinearity problem (Field, 2013). After confirming that the dataset meets the necessary assumptions for analysis, the measurement model was tested.

The Durbin-Watson test result (1.702) suggests that there is no autocorrelation between the error terms. Therefore, the hypotheses in the model were deemed suitable for testing.

10.5824/ajite.2025.01.001.x

The regression analysis results indicate that the logistics 4.0 sub-dimension ($p>0.05$, Beta: 0.097) and the business model, products, and services sub-dimension ($p>0.05$, Beta: 0.081) do not significantly affect logistic innovation. However, the marketing and customer reach sub-dimension ($p<0.05$, Beta: 0.151) and the process management sub-dimension ($p<0.05$, Beta: 0.190) significantly affect logistic innovation. Additionally, the information systems infrastructure ($p>0.05$, Beta: 0.075) and evaluation ($p>0.05$, Beta: 0.070) do not significantly affect logistic innovation. As a result, the H2g, H2h, H2j, and H2k hypotheses were rejected, while the H2ı and H2i hypotheses were accepted.

The explanatory power of the accepted hypotheses in the model is 0.303, indicating that 30% of the logistic innovation variable can be explained by the marketing customer reach, and process management variables. This effect level is considered low.

CONCLUSION

For businesses aiming to increase growth and market share in the logistics sector, integrating technology into business processes is an inevitable necessity. To achieve this integration, businesses must be innovative and open to change while carrying out their logistics activities. It is observed that businesses that are open to change can implement technological transformation more easily, achieving strategic efficiency by developing coordinated progress with their stakeholders. In this regard, it is important for businesses that adopt technology and take a digital-focused approach to analyze their position on digital transformation technologies. Businesses need to assess their current status and create a roadmap within this framework. In this context, evaluating logistics 4.0 awareness, analyzing their current position, and developing strategies for the future are of critical importance.

In this study, research was conducted to examine the relationships between logistics 4.0 awareness, digital transformation, and logistics capabilities. Following a brief introduction, a comprehensive literature review was conducted on the concepts of logistics 4.0, digital transformation, logistics capabilities, logistics service differentiation, and logistics innovation. The research method, limitations, sample, model, hypotheses, analyses, and findings were detailed. In the findings section, descriptive statistics of the participant businesses were presented, and data was analyzed based on demographic variables. The results show that a large portion of the participants are logistics sector workers, with the majority having work experience of 1-5 years. In terms of education level, it was found that bachelor's degree graduates form the highest proportion. When examining the business operational durations, it was determined that businesses that have been in operation for more than 15 years are dominant.

When evaluating the data on the use of digital transformation technologies, it was determined that businesses integrate multiple technologies, with cloud computing being the

most commonly used technology. The reliability of the scales used in the research was analyzed, and it was found that the logistics 4.0 awareness, digital transformation, and logistics capabilities scales were highly reliable. The factor analysis results revealed that the scales' factor structures were appropriate and that the distributions were consistent with the original scale items. The logistics 4.0 awareness scale consists of six factors, parallel to the industry 4.0 awareness scale, while the digital transformation scale has a single-factor structure. The logistics capabilities scale was considered in two sub-dimensions: logistics service differentiation and logistics innovation.

When testing the hypotheses of the research, simple linear regression and multiple linear regression analyses were performed to examine the relationships between variables, using the SPSS 23.0 program. A literature review was conducted based on the two main hypotheses and eighteen sub-hypotheses established in the study. It was determined that the hypothesis suggesting logistics 4.0 awareness affects digital transformation was supported and accepted. However, it was found that the impact of the marketing and customer outreach dimension on digital transformation was not significant. This situation may indicate that marketing strategies are not effectively implemented in businesses, or that digital transformation processes are unable to fully integrate these dimensions.

The hypothesis that logistics 4.0 awareness affects logistics capabilities has been supported and accepted. When examining the effects of the sub-dimensions of logistics 4.0 awareness on logistics capabilities, it was found that marketing and customer outreach, as well as process management, do not have a significant effect on logistics service differentiation; however, they do have an impact on logistics innovation. It was also found that the variables of logistics 4.0, business model products and services, information systems infrastructure, and evaluation affect logistics service differentiation, but do not have a significant effect on logistics innovation. These findings indicate that the variables affecting logistics service differentiation do not directly affect logistics innovation, and likewise, the variables affecting logistics innovation are not determinative for logistics service differentiation.

In conclusion, the relationships between logistics 4.0 awareness, digital transformation, and logistics capabilities have been identified consistently with the literature. The findings of the study highlight the importance of strategic planning for the digital transformation process in logistics businesses and show that they need to take necessary steps to gain a competitive advantage. It has been observed that the adaptation of logistics 4.0 facilitates businesses' processes in terms of efficiency, effectiveness, and productivity, while also enabling them to develop customer-focused logistics capabilities with a cost-effective structure.

The main limitation of the study is that it was conducted solely on logistics businesses that are members of the Istanbul Chamber of Commerce. Future studies could increase the generalizability of the results by using broader samples from different geographical regions

10.5824/ajite.2025.01.001.x

and industries. Furthermore, qualitative research could be conducted to explore the effects of marketing and customer outreach on digital transformation and logistics capabilities in greater depth. In this context, case studies or cross-sector comparisons could provide valuable contributions to the literature. Considering the dynamic nature of the digital transformation process, long-term studies could be conducted to understand how these effects change over time. This study, while contributing to the evaluation of the current position of businesses in the logistics sector regarding technological developments, may also serve as a fundamental source for future research.

Suggestions that can be made as a result of the study:

- Businesses should organize training programs to keep employees' knowledge up to date and seek support from expert consultants.
- To manage digital transformation processes effectively, businesses should develop short-, medium-, and long-term strategies and plan resource investments by managing employee skills properly during this process.
- Before making high-cost technological investments, a cost-benefit analysis should be conducted, and technologies that are most suitable for the business's core activities should be chosen.
- Businesses should conduct feasibility studies to measure the financial impacts of technology investments and carry out market research based on customer demands.
- Businesses should increase their operational flexibility by using digital technologies such as the Internet of Things (IoT), artificial intelligence (AI), automation, cloud computing, and blockchain, and quickly adapt to changing market conditions.
- Businesses should take advantage of big data management by analyzing customer behaviors and market trends and develop marketing and sales strategies through social media and digital platforms.
- Bureaucratic processes should be reduced, paper-based processes should be transferred to online platforms, and traceable, transparent, customer-oriented systems should be created.
- Logistics businesses should develop joint projects with software companies to ensure that all stakeholders involved in the supply chain process can manage financial and operational processes more efficiently.
- Logistics 4.0 and digital transformation models applied in developed countries should be examined, and similar strategies should be adapted for local businesses.

- Before making technology investments, elements such as software, hardware, human resources, and project leadership should be clarified, and budgeting errors should be avoided. The business's overall mission and vision should be aligned with digital transformation goals to ensure coordination among departments.
- Digital transformation processes should be embraced by all employees, and the benefits that the transformation will bring to the business should be explained with concrete examples.
- Businesses should train their current staff to support digital transformation processes and hire external specialist personnel if necessary.
- Businesses should set key performance indicators (KPIs) to measure employee performance and use methods like Balance Score Card and PDCA (Plan-Do-Check-Act) cycles for process improvement.

REFERENCES

Akdoğan, N., & Akdoğan, M. U. (2018). Büyük veri-bilişim teknolojisindeki gelişmelerin muhasebe uygulamalarına ve muhasebe mesleğine etkisi. *Muhasebe ve Denetime Bakış*, 55, 1-14.

Anderson, N. R., & West, M. A. (1998). Measuring climate for work group innovation: Development and validation of the team climate inventory. *Journal of Organizational Behavior*, 19(3), 235-258.

Atlı, E. Ç. (2022). *Lojistik 4.0 konusunun Türkiye'deki lojistik sektöründe keşifsel olarak incelenmesine yönelik vaka çakışması* [Yayınlanmamış yüksek lisans tezi]. Çanakkale Onsekiz Mart Üniversitesi, Lisansüstü Eğitim Enstitüsü.

Büyüközkan, G., & Güler, M. (2019). Lojistik 4.0 teknolojilerinin analizi için metodolojik yaklaşım. *Girişimcilik ve İnovasyon Yönetimi Dergisi / Journal of Entrepreneurship and Innovation Management - JEIM*, 8(1), 21-47.

Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2016). *Bilimsel araştırma yöntemleri* (22. b.). Pegem Akademi.

Bakan, İ., & Şekkeli, Z. H. (2015). Lojistik stratejilerinin lojistik yetenekler üzerine etkilerine ilişkin bir alan araştırması. *SOBİDER Dergisi*, 2(5), 398-422.

Banger, G. (2018). *Endüstri 4.0 uygulama ve dönüşüm rehberi*. Dorlion Yayınları.

Bouanba, N., Barakat, O., & Bendou, A. (2022). Artificial intelligence and agile innovation: Case of Moroccan logistics companies. *Procedia Computer Science*, 203, 444-449.

Çetinkaya, F. F. (2021). Endüstri 4.0 farkındalığının inovasyon üzerindeki etkisi. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 21(2), 571-598.

Çiçekli, S. (2020, Ocak). *Lojistik 4.0: Ankara lojistik üssünün farkındalık ve uygulama düzeyinin değerlendirilmesi* [Yayınlanmamış yüksek lisans tezi]. Ankara Hacı Bayram Veli Üniversitesi, Lisansüstü Eğitim Enstitüsü.

Defee, C. C., & Fugate, B. (2010). Changing perspective of capabilities in the dynamic supply chain area. *International Journal of Logistics Management*, 21(2), 180-206.

Demir, S., Paksoy, T., & Kochan, C. G. (2020). Logistics 4.0: SCM in Industry 4.0 Era: (Changing patterns of logistics in Industry 4.0 and role of digital transformation in SCM). In *Logistics 4.0*. CRC Press.

Demirdöğen, O., Erdal, H., & Akbaba, A. İ. (2018). The analysis of factors that affect innovation performance of logistics enterprises in Turkey. In F. Bakırcı, T. Heupel, O. Kocagöz, & Ü. Özen (Eds.), *German-Turkish perspectives on IT and innovation management challenges and approaches* (pp. 143-164). Springer Gabler.

Durmuş, B., Yurtkoru, S., & Çinko, M. (2022). *Sosyal bilimlerde SPSS'le veri analizi*. theKitap.

Ejsmont, K., Gladysz, B., & Kluczek, A. (2020). Impact of industry 4.0 on sustainability – Bibliometric literature review. *Sustainability*, 12(564), 1-29.

Erdal, N. (2024). 4.0 lojistik ve tedarik zinciri yönetiminde akıllı teknolojilerin kullanımı: Akıllı eldiven örneği. *ASOS Journal*, 148, 62-77.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage Publications.

Gökırmak, H. (2019). Lojistik sektöründe dijital dönüşüm ve İstanbul Otobüs A.Ş. akıllı ulaşım sistemleri uygulaması. *BEYDER Bilgi Ekonomisi ve Yönetimi Dergisi*, 14(1), 73-87.

Gudergan, G., & Mugge, P. (2017). The gap between practice and theory of digital transformation. In *Proceeding Hawaii International Conference of System Science* (pp. 1-15).

Hair, J., Black, W., Babin, B., & Anderson, R. (2014). *Multivariate data analysis* (7th ed.). Pearson Education Limited.

İşler, V. (2017, Ekim 16). Endüstri 4.0 farkındalık anketi. *Temmuz 2024* tarihinden https://docs.google.com/forms/d/e/1FAIpQLSdXZ0JFgBP8gUebLtMH1B5FYXs2gxvKTm04NpdEtLZ-IOL3mA/viewform?c=0&w=1 adresinden alındı.

Jafarzadeh, H., Aurum, H., D'Ambra, J., Abedin, B., & Assemi, B. (2015). Search engine advertising adoption and utilization: An empirical investigation of 203 inflectional factors. *Organizational Computing and Electronic Commerce*, 25(4), 402-427.

Kökhan, S. (2021). Dijital dönüşüm sürecinde yaşanabilecek zorluklar. In İ. E. Tarakçı & B. Göktaş (Eds.), *Dijital gelecek dijital dönüşüm* (pp. 93-118). Efe Akademi Yayınları.

Kache, F., & Seuring, S. (2017). Challenges and opportunities of digital information at the intersection of big data analytics and supply chain management. *International Journal of Operations & Production Management*, 37(1), 10-36.

Kagermann, H., Lukas, W. D., & Wahlster, W. (2011). *Industrie 4.0: Mit dem Internet der Dinge auf dem Weg zur 4. industriellen Revolution*, 13(1), 2-3.

Kallio, A., & Kujansivu, P. (2012). Locating the weak points of innovation capability before launching a development project. *Interdisciplinary Journal of Information Knowledge and Management*, 7.

Kane, G. C., Palmer, D., Phillips, A. N., Kiron, D., & Buckley, N. (2016). Aligning the organization for its digital future. *MIT Sloan Management Review*, 57(4).

Karadeniz, M., & Başaran, H. (2014). Lojistik işletmelerde kullanılan bilgi sistemlerinin müşterilerin hizmet algısı üzerine etkisinin araştırılması. *Yönetim Bilimleri Dergisi*, 12(24), 239-273.

Karagöz, İ. B., & Akgün, A. E. (2015). The roles of IT capability and organizational culture on logistics capability and firm performance. *Journal of Business Studies*, 7(2), 23.

Kayaalp, G. T., Güney, M. Ç., & Cebeci, Z. (2015). Çoklu doğrusal regresyon modelinde değişken seçiminin zootekniye uygulanışı. *Çukurova Üniversitesi Ziraat Fakültesi Dergisi*, 30(1), 1-8.

Kozan, K. G. (2019). Lojistik yeteneğin öncülleri ve lojistik yeteneğin lojistik performans üzerindeki etkisi. Yayınlanmış yüksek lisans tezi, Kahramanmaraş Sütçü İmam Üniversitesi Sosyal Bilimler Enstitüsü.

Langley, C., John, J., & Mary, C. H. (1992). Creating logistics customer value. *Journal of Business Logistics*, 13(2), 1-27.

Lansiti, M., & Lakhani, K. (2014). Digital ubiquity—How connections, sensors and data are revolutionizing business. *Harvard Business Review*.

Lynch, D. F., Keller, S. B., & Ozment, J. (2000). The effects of logistics capabilities and strategy on firm performance. *Journal of Business Logistics*, 21(2), 47-67.

Muñoz, F., José, J., Gónzalez, G., & Manuel, J. (2017, July). Mediation analysis with the PROCESS macro for SPSS. *Revista d'Innovació i Recerca en Educació*, 10(2), 79-88. https://doi.org/10.1344/reire2017.10.218109

Nadeem, A., Abedin, B., Cerpa, N., & Chew, E. (2018). Editorial: Digital transformation & digital business strategy in electronic commerce—the role of organizational capabilities. *Journal of Theoretical and Applied Electronic Commerce Research*, 13(2), 1-8.

Neuman, W. L. (2010). *Social research methods: Qualitative and quantitative approaches* (7th ed.). Allyn & Bacon.

Öztemel, E., & Gürsev, S. (2018). Türkiye'de lojistik yönetiminde endüstri 4.0 etkileri ve yatırım imkanlarına bakış üzerine anket uygulaması. *Marmara Fen Bilimleri Dergisi*, 145-154.

Park, J., & Han, S. (2023). Investigating older consumers' acceptance factors of autonomous vehicles. *Journal of Retailing and Consumer Services*, 72, 103241.

Partala, J. (2018). Provably secure covert communication on blockchain. *Cryptography*, 2(3), 2.

Salkin, N. J., & Rasmussen, K. (2007). *Encyclopedia of measurements and statistics*. Sage.

Sarıhan, A. Y. (2021, Nisan). *Lojistik yetenekler, şebeke ilişkileri, alıcı-satıcı ilişkileri, teknolojik ilerleme ve tedarik zinciri performansı arasındaki ilişkinin ihracat işlemler üzerinde çok katmanlı olarak incelenmesi* [Yayınlanmamış doktora tezi]. Gazi Üniversitesi, Sosyal Bilimler Enstitüsü.

Sağlam, M. (2021). İşletmelerde geleceğin vizyonu olarak dijital dönüşümün gerçekleştirilmesi ve dijital dönüşüm ölçeği'nin Türkçe uyarlaması. *İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi*, 20(40), 395-420.

Schwertner, K. (2017). Digital transformation of business. *Trakia Journal of Sciences*, 15(1), 388-393.

Statistics Canada. (2010). Census metropolitan area and census agglomeration influenced zones.

Song, X. M., & Parry, M. E. (1997). A cross-national comparative study of new product development processes: Japan and the United States. *Journal of Marketing*, 61(2), 1-18.

Strandhagen, J. W., Alfnes, E., Strandhagen, J. O., & Vallandingham, L. R. (2017). The fit of Industry 4.0 applications in manufacturing logistics: A multiple case study. *Advances in Manufacturing*, 5(4), 344-358.

Svahn, F., Mathiassen, L., & Lindgren, R. (2017). Embracing digital innovation in incumbent firms: How Volvo Cars managed competing concerns. *MIS Quarterly*, 41(1), 239-253.

Şekkeli, Z. H., & Bakan, İ. (2018). Endüstri 4.0'ın etkisiyle lojistik 4.0. *Journal of Life Economics*, 17-36.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. Pearson.

Thoben, K. D., Wiesner, S., & Wuest, T. (2017). Industrie 4.0 and smart manufacturing – A review of research issues and application examples. *International Journal of Automation Technology*, 11(1), 4-16.

Verina, N., & Titko, J. (2019). Digital transformation: Conceptual framework. In *Proceedings of the International Scientific Conference Contemporary Issues in Business* (pp. 9-10). Management and Economics Engineering.

Winkelhaus, S., & Grosse, E. H. (2020). Logistics 4.0: A systematic review towards a new logistics system. *International Journal of Production Research*, 58(1), 18-43.

Witkowski, K. (2017). Internet of Things, big data, Industry 4.0, innovative solutions in logistics and supply chain management. *Procedia Engineering*, 182, 763-769.

Yılmaz, N. (2020). Rekabetçi bir iş modeli: Ürün-hizmet sistemleri. *KMÜ Sosyal ve Ekonomik Araştırmalar Dergisi*, 22(39), 336-349.

Yang, C. C., Marlow, P. B., & Lu, C. S. (2009). Assessing resources, logistics service capabilities, innovation capabilities and the performance of container shipping services in Taiwan. *International Journal of Production Economics*, 122, 4-20.

AJIT-e Academic Journal of Information Technology
10.5824/ajite.2025.01.002.x

2025 . Winter - Kış . Volume – Cilt: 16 . Issue – Sayı: 1

A Study of Academic Integrity in Online Distance Education

Bahaddin ERDEM, *Bitlis Eren University, Adilcevaz Vocational School, Lecturer, bahaddin2363@gmail.com, 0000-0003-3693-0966*

Songül KARABATAK, *Fırat University, Faculty of Education, Assoc. Prof., skarabatak@firat.edu.tr, 0000-0002-1303-2429*

Murat KARABATAK, *Fırat University, Faculty of Technology, Prof., mkarabatak@firat.edu.tr, 0000-0002-6719-7421*

ABSTRACT *The global Covid-19 pandemic has accelerated the implementation of distance education and training, resulting in a shift of course activities and processes to an online environment. The introduction of online exams in this new educational landscape has raised concerns among learners regarding academic honesty. This situation has prompted worries about safeguarding the ongoing institutional identity and reputation of academic institutions, thereby raising questions about academic honesty. This study presents a systematic review of the literature on protecting academic honesty in online examinations, focusing on recent publications within the last five years. This period is marked by intensified online exam evaluations and the impact of Covid-19. Based on the literature review, a total of 50 studies were examined by three experts in the field. Descriptive and content analysis techniques were utilized to identify research trends. The findings of these studies indicated that learners tend to accept unethical behavior violations, which are also increasing. In conclusion, it is recommended that future studies consider factors such as a wider range of publication years, employed research methods, increased publication rates, success rates, dataset size, and the utilization of assistive technologies.*

Keywords : Distance Education, Online Cheating Detection, Academic Honesty, Online Academic Honesty

Çevrimiçi Uzaktan Eğitimde Akademik Dürüstlük Üzerine Bir Çalışma

ÖZ *Dünya genelinde Covid-19 salgınının ortaya çıkması uzaktan eğitim ve öğretimin çok acil bir şekilde hayata geçirilmesine ve bunun sonucunda ders etkinlikleri ile süreçleri çevrimiçi bir ortama taşınmasına neden oldu. Bu*

Derleme Makale / Review Article
Date Received: 30.10.2024
Publishing Date: 28.02.2025

10.5824/ajite.2025.01.002.x

durum akademik kurumlar nezdinde devam eden kurumsal kimliği ve itibarı korumak olarak endişelere yol açmış ve akademik bütünlük sorgulanmaya başlanmıştır. Bu makale çevrimiçi sınavlarda akademik bütünlüğü korumak adına literatürde yer alan yayınların, yayınlama amacına göre sistematik bir incelemesini ele almaktadır. Çevrimiçi sınav değerlendirmelerinin yoğunlaştığı ve Covid-19'un da etkili olduğu son beş yıldaki yayınlar dikkate alınmıştır. Literatür incelemesi sonucu 50 yayın, yayınlama amacına hizmet etmeleri dikkate alınarak analiz edilmiştir. Yapılan analizlerde yayınların kopya çekme, akademik dürüstlük incelemesi ve akademik dürüstlüğe yönelik tutum ve algıları incelemek amacıyla yayınlandıkları tespit edilmiştir. Amaç doğrultusunda çevrimiçi sınavlarda etik olmayan davranış ihlalleri, kullanılan gözetimli ve gözetimsiz teknolojilerle, sesli yardımcı teknolojilerle, çeşitli sınav senaryolarıyla, kullanılan soru tipleriyle, yapılandırılmış görüşme formları ve anket uygulamalarıyla tespit edilmeye çalışılmıştır. Yapılan çalışmalar etik olmayan davranış ihlallerinin öğrenenler tarafından kabul edildiğini ve sürekli arttığını göstermiştir. Literatürde yayınların kapsadığı yıllar geniş tutularak, kullanılan yöntemler, yayın sayısındaki artış, bunların başarım oranları, veri seti ve kullanılan yardımcı teknolojiler gibi parametreler dikkate alınması geleceğe yönelik yapılacak çalışmalarda akademik bütünlük açısından araştırmacılara fayda sağlayabilir.

Anahtar kelimeler : Uzaktan Eğitim, Akademik Dürüstlük, Çevrimiçi Akademik Dürüstlük

INTRODUCTION

Education is focused on fostering knowledge and promoting behavioral change in learners through interactive engagement between teachers and students utilizing various tools. The idea of distance in education refers to the physical separation between teachers and students. According to Kaya (2023, p. 410) in the realm of distance education, the involvement and whereabouts of instructors and administrators can generally be defined in relation to place and time, whereas students have the flexibility to access educational and training processes remotely from any location at their convenience.

Distance education is a multidisciplinary field that effectively utilizes information technologies to overcome barriers between learning resources, students, and instructors (Atasel, 2023, p. 62). The proliferation of communication and multimedia technologies that can be adapted to the educational context has expanded opportunities for lifelong learning and education, regardless of time and location (Kavrat & Türel 2013, p. 23). According to Amzalag et al. (2021, p. 243), the global outbreak of Covid-19 has further underscored the significance of distance education, prompting educational institutions to shift all their teaching and training processes to online platforms. According to Kuang Chiang et al. (2022, p. 909), today,

the demand for online courses is increasing day by day due to the opportunities offered by Internet technologies. According to Özen & Düzenli (2023, p. 316), with technological advancements in online educational environments, the number of students engaging with these platforms is steadily rising.

According to Almuhanna (2023, p. 130), with the widespread adoption of distance education, the incorporation of online assessments has become inevitable. There is a growing interest in utilizing online formative assessment activities in both distance education and traditional teaching methods. According to Kuang Chiang et al. (2022, p. 908), in the online learning process, teachers and educational institutions need to be pedagogically supported to evaluate students' academic integrity online. According to Marriott (2009, p. 237), e-assessment presents opportunities to develop innovative assessment practices that engage students and enhance their motivation to learn. Online learning and assessment should be viewed as a comprehensive system for educating students and evaluating their academic achievements. According to Gaytan & McEwen (2007, p. 130), understanding online learning and assessment is crucial in a time where educational institutions face increasing demands for accountability, growth, and excellence.

E-assessment has many advantages: flexibility in conducting exams in terms of location and timing, lower cost than traditional assessment, using multimedia elements, reviewing questions, creating exams from question banks, instant feedback and automatic grading, which also provides significant advantages for instructors (Gaytan & McEwen 2007, p. 130; James, 2016, p. ; Hebebci & Yılmaz, 2022, p. 130).

According to Hebebci & Yılmaz (2022, p. 105) and Aslan (2024, p. 121), although online assessment and evaluation methods offer more alternatives than those used in face-to-face education, providing conveniences and opportunities such as face-to-face multidimensional assessment, time and space independence, they have also brought some negative aspects. These include requirements for hardware, software, and internet infrastructure, laborious question preparation, budget and time demands for managing and operating online systems, difficulties in preventing cheating in online exams, ensuring system security, and addressing instant failures or interruptions during the exam process. These challenges require more preparation compared to face-to-face education.

Efforts to detect unethical behaviors in online distance education exam environments are crucial for protecting the academic honesty of learners and the reputation of relevant institutions. Behaviors such as cheating, plagiarism, dishonesty, lying, and making excuses are contrary to the goals of education and training processes, and pose significant problems for institutions.

The purpose of this article is to systematically review 50 publications in the literature to protect academic integrity in online distance education exams, to examine cheating,

10.5824/ajite.2025.01.002.x

academic honesty review, and attitudes and perceptions towards academic honesty. In line with the purpose of the study, publications were tried to determine unethical behavior violations in online exams, supervised and unsupervised technologies used, voice assistive technologies, various exam scenarios, question types used, structured interview forms and survey applications

1. ONLINE ACADEMIC HONESTY

In synchronous online learning, students and instructors are typically expected to participate in virtual class meetings in real time. On the other hand, asynchronous online learning does not have scheduled meeting times but does have deadlines for learning assignments. Blended courses, on the other hand, involve learners and instructors interacting in both virtual and physical classroom settings. The use of synchronous and asynchronous communication technology in educational environments has greatly improved accessibility and increased the availability of educational resources Kuang Chiang et al. (2022, p. 908). However, it is important to note that online learning is not without its challenges, particularly in terms of academic dishonesty Holden (2021, p. 1, as cited in Fishman, 2014). Academic dishonesty in online learning is common and a major concern for educational institutions and society as a whole Holden (2021, p. 1). Upholding academic honesty requires a strong commitment to core values such as trust, fairness, respect, responsibility, and courage. By upholding these values, ethical academic behavior is defined by creating a community dedicated to learning and the exchange of knowledge. Students benefit individually from this reputation and the inferences drawn from their academic achievements. According to Holden (2021, p. 2), at a broader level, understanding and adhering to the core values of academic honesty within a community establishes a shared framework for professional work, clearly demonstrating its value in terms of knowledge, skills, and abilities. According to Gehringer & Pedyycord (2013, p. 10), an online exam is a special type of assessment where students are allowed to use the internet. While this may provide an authentic experience, it can also make it challenging to determine whether students have completed the work independently. According to Holden (2021, p. 10), as online distance education continues to expand and universities move towards online education, faculty and administrators face the challenge of developing methods to effectively assess student learning in an online environment while upholding academic honesty. Although many faculty members state that the use of Internet technologies for course purposes will provide more flexibility than traditional course formats, they are also concerned about the emergence of unethical behavior (Grijalva, 2006, p. 14; McGee, 2003, p. 14; Bandyopadhyay and Barnes, 2014, p. 2; Rautela, 2022, p. 111). According to Mantecon et al. (2018, p. 1) the automated assessment of text-based assessment items such as short answers or essays is an important and ongoing research issue.

According to King and Case (2014, p. 20), with online courses offered to provide more flexibility to students, concerns about academic honesty are emerging. Academic dishonesty is always a concern in any educational setting. According to Berkeley City College (2018) academic dishonesty is defined as "any form of cheating that occurs in connection with formal academic exercises." "Cheating is defined as an attempt to give or receive assistance in a formal academic practice (such as an examination) without due approval". The provision of various technological facilities that encourage students' academic dishonesty has been a major problem in online assessment. Findings from studies show that the most common type of cheating activity reported by students is downloading documents from the internet and passing them off as the student's own work. Additionally, students report that they believe other students cheat more on homework problems than on exams, term papers, and internet projects. According to King and Case (2014, p. 20), however, the most troubling finding for educators is that e-cheating appears to be on the rise. During the coronavirus disease-2019 (Covid-19) pandemic, many higher education institutions have adopted online exam proctoring technologies to monitor ever-increasing cheating behavior and control unethical behavior. According to Lee & Fanguy (2022, p. 475); Stoesz & Eaton (2022, p. 36) although it may seem like a natural and effective solution, researchers argue that proctoring technologies arise from problematic assumptions when considering pedagogical approaches to fairly assessing student online learning performance.

2. METHOD

2.1. Research Model

A qualitative research method was employed for this study. Within the purview of the qualitative research approach, publications were scrutinized through document analysis to elucidate trends in online academic honesty.

Collecting data by examining existing records and documents is called documentary scanning. It includes the process of finding and evaluating sources for a specific purpose (Ekinci, 2019, p. 81). Document analysis includes the examination of the full text of, or excerpts from, organizational, clinical, or program records, official publications, or reports (Ekinci, 2019, p. 201). PRISMA guidelines are recommended for the systematization of documentation documents and the presentation of transparent reports (Toker, 2022, p. 321). The PRISMA flow chart explaining the methodology of the study is given in Figure 1.

10.5824/ajite.2025.01.002.x

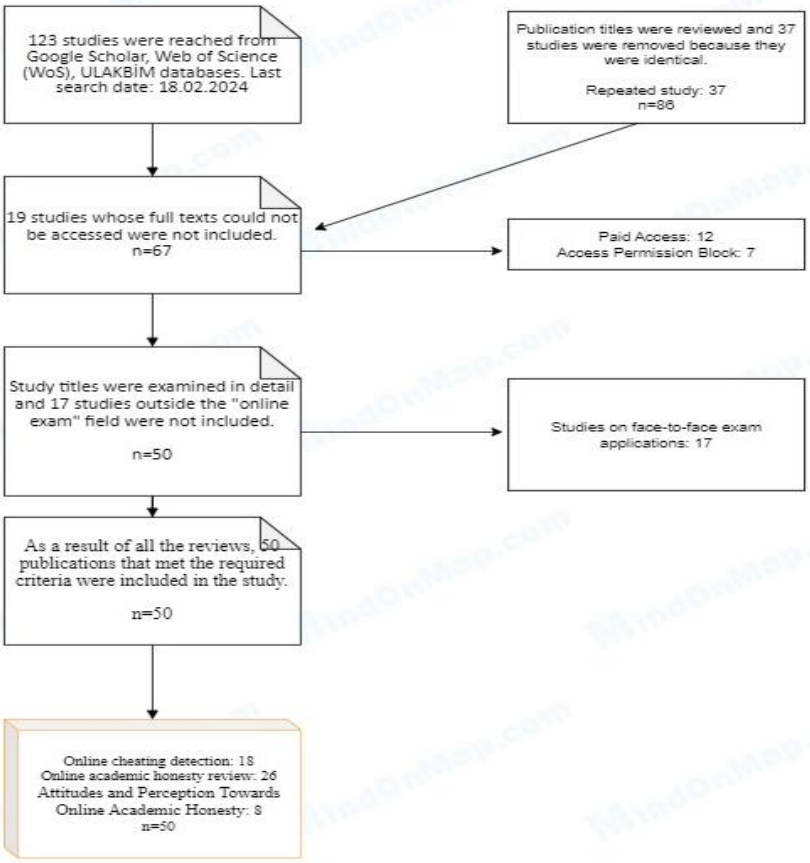


Figure 1: PRISMA flow chart

Table 1: PRISMA Model for systematic literature reviews within the scope of the study.

Study design	The study applies a literature review to progressively report studies in the existing literature in an open, transparent and rigorous manner.
Review protocol	The researchers determined the keywords and research criteria in advance to ensure neutrality.
Eligibility criteria	The studies were classified by the researchers by searching the Web of Science, Google Scholar and ULAKBİM databases.
Publication type included	The Web of Science, Google Scholar and ULAKBİM databases.
Publication time frame	2018-2022
Language	English
Search strategy	The relevant source databases were queried with the strings “online exam” AND “academic honesty” AND “ethical behavior” AND “cheating”. As a result, a large number of publications were reached. A manual review was performed to filter. As a result of the review, 50 publications that had the desired characteristics were finally obtained.

Table 1 shows the PRISMA model explaining the systematic literature review of the study in detail.

2.2. Sampling

This research examines existing studies on academic honesty in the online examination environment between 2018-2022 for the purposes of the study. To ensure a systematic approach, Google Scholar, Web of Science (WoS), and ULAKBIM databases were utilized for the review process. Relevant resources were searched using the specific keywords and phrases: "online exam," "academic honesty," "ethical behaviors," and "cheating". This search resulted in a significant number of studies, which were then manually reviewed and filtered. Ultimately, 50 studies that met the desired criteria were identified. It is worth noting that the number of studies in this field has shown a periodic increase, particularly following the Covid-19 pandemic. For detailed information about the codes, names, and researchers involved in these studies, please refer to Table 1.

Table 2 presents a comprehensive overview of the relevant studies, their respective authors, and their coding status in relation to the study's objectives. As depicted in Table 2, the majority of the studies (94%) were sourced from foreign literature.

Table 2: Details of the studies.

Study Code	Author(s)	Year
A1	Alvarado et al., 2018	2018
A2	Oravec, 2022	2022
A3	Susnjak, 2022	2022
A4	Ayoub et al., 2021	2021
A5	Zhan et al., 2022	2022
A6	Rautela et al., 2022	2022
A7	Salman et al., 2022	2022
A8	Munteanu, 2021	2021
A9	Gehlot et al., 2022	2022
A10	Jaramillo-Morillo et al., 2022	2022
A11	Abozaid and Atia, 2022	2022
A12	Duzbayeva et al., 2022	2022
A13	Boobalan, 2022	2022
A14	Hamzaoui, 2022	2022
A15	Banson and Hardin, 2022	2022
A16	Lee et al., 2022	2022
A17	Revilla and Libre, 2022	2022
A18	Pleasants and Pleasants, 2021	2021
A19	Ivashkina, 2021	2021
A20	Eramo, 2021	2021
A21	Paredes et al., 2021	2021
A22	Tanis, 2020	2020
A23	Verhoef and Coetser, 2021	2021
A24	Norris, 2019	2019
A25	Lee ve Fanguy, 2022	2022
A26	Golden and Kohlbeck, 2020	2020
A27	MacLeod and Eaton, 2020	2020

10.5824/ajite.2025.01.002.x

Study Code	Author(s)	Year
A28	Stephens et al., 2021	2021
A29	Burgason, 2019	2019
A30	Dyer et al., 2020	2020
A31	Costley, 2019	2019
A32	Chirikov et al., 2020	2020
A33	Reisenwitz, 2020	2020
A34	Janke et al., 2021	2021
A35	Tweissi et al., 2022	2022
A36	Novick et al., 2022	2022
A37	Hussain et al., 2021	2021
A38	Harton et al., 2019	2019
A39	Bilen ve Matros, 2021	2021
A40	Dendir and Maxwell, 2020	2020
A41	Lucky et al., 2019	2019
A42	Rivera-Mata, 2021	2021
A43	Vazquez et al., 2021	2021
A44	Jaramillo-Morillo et al., 2020	2020
A45	Slimi, 2020	2020
A46	Amzalag et al., 2021	2022
A47	Reedy et al., 2021	2021
A48	Fatima et al., 2020	2020
A49	Çakmak and Baysen, 2022	2022
A50	Acar Güvendir and Özer Özkan, 2021	2021

2.3.Development of the Analysis Form

A five-question form was utilized to analyze the 50 publications included in the study. The researchers collaborated with three experts who are proficient in the field of "Scientific Research Methods" to ensure the suitability of the prepared questions for the overall purpose. The experts confirmed that the questions were indeed appropriate for the intended purpose. The form consisted of the following questions: (1) How are the studies categorized according to publication type? (2) How are the studies distributed across different years? (3) How are studies distributed across different countries? (4) How are the studies categorized according to the purpose of publication? (5) How are the studies distributed according to the type of participants?

2.4. Data Analysis

In the analysis of the obtained data, descriptive analysis from qualitative research techniques was employed to analyze the data analysis techniques, research methods, data collection tools, and publication tags. Content analysis from qualitative research techniques was employed to examine the study topics of the publications. The publications were assessed by three experts in the field. Based on the literature review, the experts unanimously agreed

that there are discernible trends in the categories of "Online Cheating Detection", "Online Academic Honesty Review", and "Attitudes and Perceptions Towards Online Academic Honesty" within publications that encompass exams, assignments, citations, etc. in online environments.

3. FINDINGS

The 50 publications obtained from the studies conducted in the last five years underwent descriptive and content analysis in order to ascertain research trends.

The findings regarding the classification of studies based on publication type are presented in Table 3.

(1) How are the studies categorized according to publication type

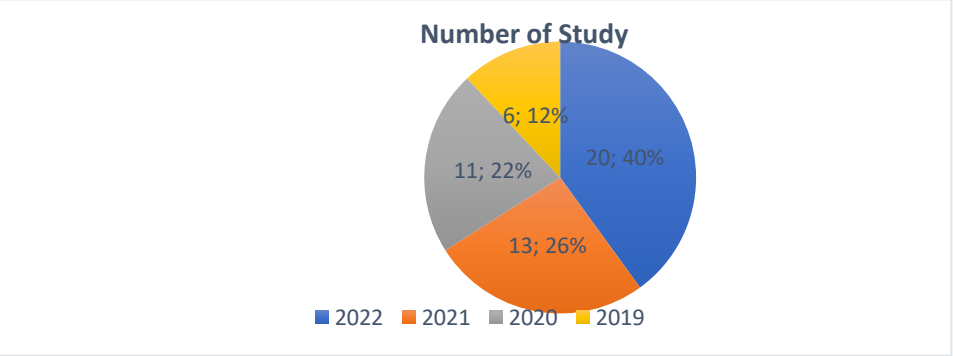
Table 3 illustrates the distribution of the studies based on their publication type. As depicted in the table, 90% of the studies were published as articles, while the remaining 10% were presented in various formats such as conferences, symposiums, projects, theses, and papers.

Table 3: Categorization of studies according to publication type

Publication Type	Study Code	N	%
Article	A2-A3-A4-A5-A6-A7-A8-A10-A12-A14-A16-A17-A18-A19-A21-A22-A23-A24-A25-A26-A27-A28-A29-A30-A31-A32-A33-A34-A35-A36-A37-A38-A39-A40-A41-A42-A43-A44-A45-A46-A47-A48-A49-A50	45	90
Other (paper-project-thesis)	A9-A11-A13-A15-A20	5	10

(2) How are the studies distributed across different years?

The publication dates and rates of studies conducted within the past five years are presented in Graph 1.



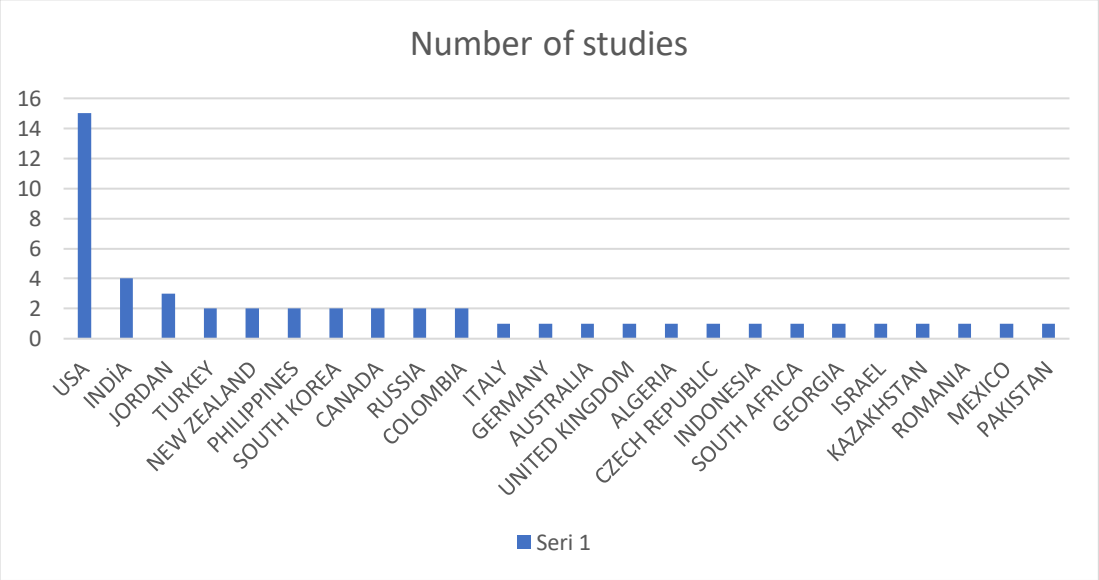
Graph 1: Number of studies per year

10.5824/ajite.2025.01.002.x

Graph 1 shows the number of publications of studies conducted in the last five years by year. As can be seen from the graph, it is observed that studies on online exams and academic honesty are constantly increasing.

(3) How are the studies distributed across different countries?

The distribution of studies by country is presented in Graph 2.

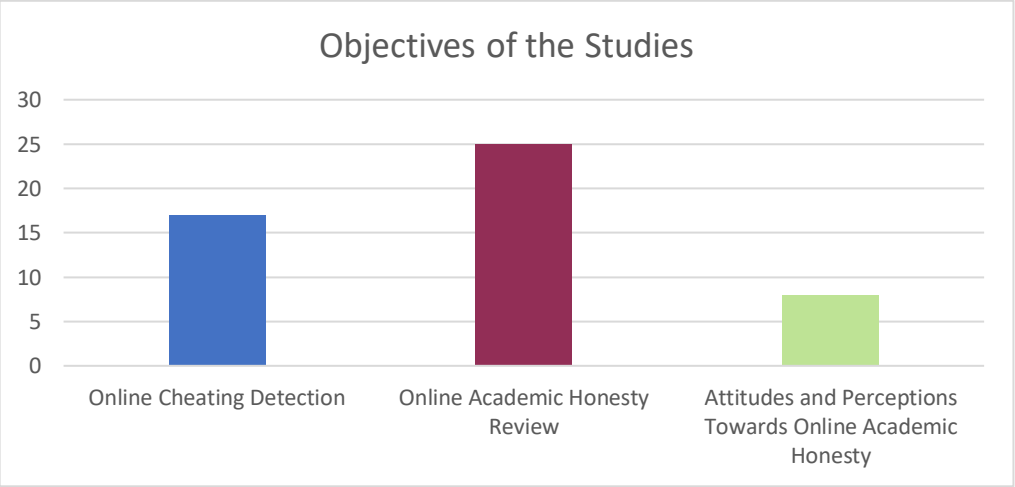


Graph 2: Number of Studies by Country

As seen in Graph 2, while the USA has the most studies with 15, India has 4, and other countries have two or one study each, indicating significant differences in research outputs.

(4) How are the studies categorized according to the purpose of publication?

Graph 3 presents the objectives of the studies.

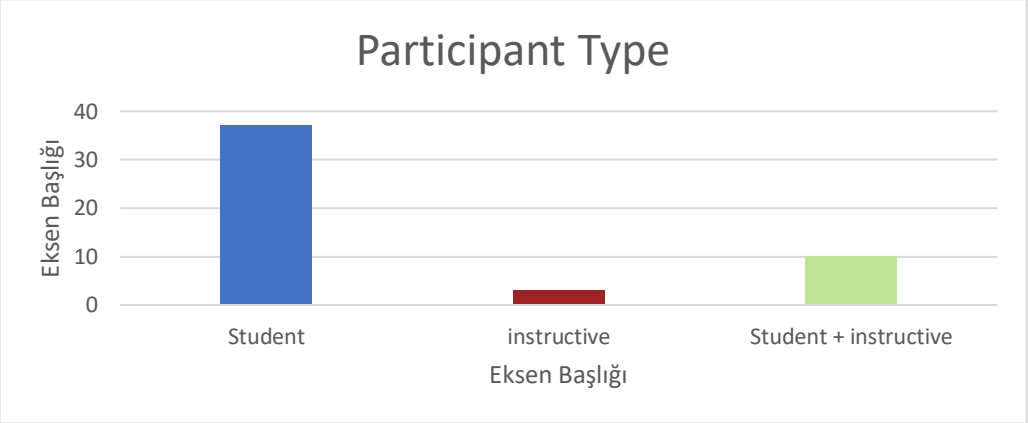


Graph 3: Objectives of Publication of Studies

Graph 3 depicts the findings of research studies on online academic honesty. Specifically, it reveals that there have been 25 studies conducted on this topic, while 17 studies have investigated online cheating. Additionally, 8 studies have focused on attitudes and perceptions related to online academic honesty.

(5) How are the studies distributed according to the type of participants?

Graph 4 presents the findings pertaining to the number of studies based on the type of participants included in the study sample.



Graph 4: Number of Studies Categorized by Participant Type

In Graph 4, it is observed that 37 studies exclusively focus on the student sample, 2 studies concentrate on the instructive sample, and 10 studies incorporate both instructor and student data. The 50 studies were analyzed based on their publication objectives and categorized into three themes: Online Cheating Detection, Online Academic Honesty Review, and Attitudes and Perceptions Towards Online Academic Honesty. The themes and corresponding publication codes, aligned with the objectives of the studies, are outlined in Table 3.

Based on the analysis, as indicated in Table 4, it was determined that the research objectives primarily focused on cheating detection to uphold academic honesty in online settings, uncovering instances of plagiarism, detecting unethical behavior among students, promoting academic honesty, and examining attitudes and perceptions towards academic honesty. Specifically, there were 17 studies on Online Cheating Detection, 25 studies on online academic honesty, and 8 studies on attitudes and perceptions towards online academic honesty. The findings derived from these studies, organized according to their respective themes, are outlined below.

10.5824/ajite.2025.01.002.x

Table 4: Analysis of the Studies Based on the Publication Objective

Themes	Study Code	N	%
Online Cheating Detection	A1, A2, A3, A8, A9, A11, A18, A19, A33, A34, A35, A39, A40, A41, A42, A43	17	34
Online Academic Honesty Review	A4, A6, A7, A10, A12, A13, A14, A15, A17, A20, A21, A22, A23, A24, A25, A26, A27, A29, A32, A36, A37, A45, A48, A49, A50	25	50
Attitudes and Perceptions Towards Online Academic Honesty	A5, A16, A28, A30, A31, A38, A46, A47	8	16

Preserving the essence of the statements in the studies, select excerpts related to each theme are presented.

A study on Online Cheat Detection has been explained as follows:

Developed using the Viva Platform, the webcam and microphone application is utilized for the detection and monitoring of suspected students. The system is comprised of various components, including a secure environment, an advanced authentication system to verify the intended candidate, a comprehensive monitoring system to observe the candidate during evaluation, and a response verification system. During the examination, students have the ability to make different movements, such as facial and eye movements, which can be directly observed via the camera. These movements provide valuable information about the student's actions through the facial landmark pattern. A correlation has been identified between students turning away from the camera and engaging in cheating behavior during the Viva session (Gehlot & Joshi, 2021, p. 1).

In the study conducted by Bilen and Matros (2021, p. 203) on Online Copy Detection, it is explained as follows:

"In a conducted study to investigate cheating behavior in online exams, data from short-answer student exams and time per question measurements were collected, with the webcam serving as a supplementary technology. The findings from the study enabled the detection of copy instances" .

In the study conducted by Pleasants and Pleasants (2021, p. 268) on Online Copy Detection, it is explained as follows:

"Another study examined exam data from unproctored exams in an introductory biology course delivered online. A feature of the distance education learning management system utilized is its ability to detect behaviors such as learners navigating away from the exam

page and accessing other materials on the computer. When exploring the relationship between cheating behavior and exam performance, it was discovered that 70% of students engaged in cheating, with many repeating this behavior throughout the exam. Moreover, appeals to learners' honesty or requesting them to pledge their integrity were found to be ineffective in preventing cheating. However, when learners were informed about the presence of technology capable of detecting cheating in the exam environment, as well as the associated penalties, it was observed that cheating behavior decreased to 15% of learners. Nevertheless, no evidence indicates that changes in cheating behavior correspond to changes in exam performance".

In the study conducted by Rautela et al. (2022, p. 111) on On Online Academic Honesty Review, it is explained as follows:

"This study aims to identify the factors contributing to unethical practices in online evaluations. It proposes a four-level model that focuses on the lack of training for both faculty and students, personal ethics and evaluation, technological barriers, interpersonal barriers, and time management design as the underlying causes of unethical behavior in online assessments. The interconnections between these factors are further elucidated through fuzzy MICMAC analysis".

In the study conducted by Verhoef and Coetser (2021, p. 1) on On Online Academic Honesty Review, it is explained as follows:

"Another study explores the phenomenon of academic honesty among university students during online assessment from the perspective of South African students. Audio data obtained from a university's institutional forum was analyzed using a phenomenological approach. The analysis results indicate that while some students engaged in dishonest behavior during the pandemic period due to related issues (such as lack of monitoring), there are other factors such as feeling overwhelmed and stressed, inadequate time management, and struggles with technology that contribute to students' dishonesty".

In the study conducted by Stephens et al. (2021, p. 1) on Attitudes and Perceptions Towards Online Academic Honesty, it is explained as follows:

"A survey was used in a research project aimed at expanding the existing body of knowledge on the effects of online academic integrity training on the perceptions of academic dishonesty held by college students, and its impact on the participants. Prior research has shown the effectiveness of such training in enhancing knowledge and deterring cheating, but contrary to expectations, the results of this study indicated a diminished understanding of academic integrity as a result of the provided training".

In the study conducted by Costley (2018, p. 205) on Attitudes and Perceptions Towards Online Academic Honesty, it is explained as follows:

10.5824/ajite.2025.01.002.x

"Academic dishonesty can reduce the quality of the learning experience for students, and reduce the validity and trust in online class's assessment. For this reason, understanding how student perceive their cheating behavior is useful for researchers. This paper looks at the interview responses of a group of students taking cyber university classes to gain insight into their perceptions and motivations for cheating. The responses were varied, but showed the ad-hoc nature of the behavior and that students felt that academic dishonesty was a natural aspect of their learning experience".

In the study conducted by Amzalag et al. (2021, p. 243) on Attitudes and Perceptions Towards Online Academic Honesty, it is explained as follows:

"This study aimed to investigate unethical behavior among students in online exam environments, as well as the underlying reasons and perceptions of academic dishonesty. A survey was conducted with both instructors and students to collect data. The findings revealed a tendency among young learners to engage in cheating. Moreover, a mutual lack of trust between learners and instructors emerged as a concerning factor".

4. DISCUSSION

The articles analyzed in this study were classified based on the study's purpose, which predominantly focused on cheating behavior, academic honesty, and attitudes and perceptions towards academic honesty. Although the objectives were similar, the findings and suggested solutions varied.

According to Gehlot and Joshi (2022, p. 2), regarding cheating detection, some studies have found that using a webcam can act as a deterrent to cheating incidents. However, according to Bozkurt and Uçar (2018, p. 753), other studies have indicated that webcams do not have an impact on cheating behavior, and that the use of assistive technologies can infringe upon privacy. Additionally, concerns about identity verification during online exams have been raised. Furthermore, biometric identification devices may not yield reliable results if they are expensive, require frequent calibration, demand continuous technical support, or fail to provide sufficient data input for pattern identification techniques. According to Güvendir and Özkan (2021, p. 23), another study in the literature revealed that prospective teachers find online exams more beneficial than in-class exams due to the flexibility of time and location, but it also highlighted that online exams are highly susceptible to cheating. According to Mantecon et al. (2018, p. 10); Oravec (2022, p. 6); Abozaid & Atia (2022, p. 1), the methods employed in studies on cheating behavior vary, with some using artificial intelligence methods through open-ended quizzes and proctored technologies, while others utilize multimodal regression, logistic regression, or statistical methods.

According to Chirikov (2020, p. 2466), studies on academic honesty tend to utilize surveys as qualitative and quantitative data collection tools. The primary aim of these studies is to bring about a change in learners' behavior by involving institutions that adopt academic honesty policies, and the training provided within the scope of these policies. It has been found that students who face severe penalties from their instructors due to unethical behavior during the educational process, students who receive academic honesty training, or students who commit to specific procedures display more honest behavior in the face of academic dishonesty. However, according to Özmen and Yurttaş (2020, p. 21), it is worth noting that these studies are conducted at different times, places, and educational institutions, which may result in unexpected behaviors among individuals due to factors such as parental pressure, academic success, homework, and peer relations.

Studies on academic honesty commonly employ artificial intelligence, statistical, and descriptive analyses. When examining perceptions and attitudes towards academic honesty, these studies typically assess satisfaction, academic honesty, attitudes, and perceptions. Statistical methods are commonly employed, and significant differences are observed when evaluating these factors in terms of demographic information, technologies used, and supervised or unsupervised online environments.

RESULTS

This study examines 50 publications published between 2018 and 2022 that explore the evaluation of online exams and unethical behavior within online applications. This topic is becoming increasingly important, especially given the rise of online applications during the Covid-19 period. In this changing landscape, learners and teachers interact in different locations and time periods, resulting in new learning experiences. The focus of this study is on the use of e-assessment systems to measure learners' behaviors in the online environment, based on various theories and approaches.

The study adopts a multidisciplinary approach, combining engineering techniques such as deep learning and fuzzy logic with statistical and descriptive methods from the social science and education fields. Through the analysis of data from online exam applications and surveys, distinct behaviors related to cheating detection, academic honesty, and learners' attitudes and perceptions are revealed. These findings can be valuable in informing policies related to academic honesty in educational institutions.

Within the scope of this study, various online exam scenarios were created using both supervised and unsupervised online exam environments. These scenarios involved generating a question pool that matched the number of students and ensured that questions were not sourced from the internet. Additionally, the scenarios covered different types of questions. These findings can provide guidance for instructors in future online exam applications.

10.5824/ajite.2025.01.002.x

In future studies, it would be beneficial for researchers to consider the relevant literature, the methods used, the dataset, the assistive technologies employed, the increase in the number of publications, their performance rates, and the years covered by the publications. Taking these factors into account will contribute to a comprehensive understanding of academic honesty.

REFERENCES

Abozaid, A., & Atia, A. (2022, July). Multi-Modal Online Exam Cheating Detection. In *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)* (pp. 1-6). IEEE.

Almuhanna, M. (2023). Improving E-Assessment Based on University Students' Experiences. *The Turkish Online Journal of Educational Technology*, 22(1), 130-143.

Amzalag, M., Shapira, N., Dolev, N. (2021). Two sides of the coin: lack of academic integrity in exams during the corona pandemic, students' and lecturers' perceptions. *Journal of Academic Ethics*, 20, 243-263.

Aslan, M. (2024). EFL Students' Views about Distance Education at Higher Education in Turkey. AJIT-E: Academic Journal of Information Technology, 15(2), 121-137. https://doi.org/10.5824/ajite.2024.02.001.x

Atasel, O.Y. (2023). Uzaktan Eğitim ile Alınan Muhasebe Dersine Yönelik Bir Araştırma: KTÜ İİBF Örneği. *Denetim ve Güvence Hizmetleri Dergisi*, 3(1), 61-87.

Ayoub/Al-Salim, M. I., Aladwan, K. (2021). The relationship between academic integrity of online university students and its effects on academic performance and learning quality. *Journal of Ethics in Entrepreneurship and Technology*, *1*(1), 43-60.

Bandyopadhyay, K., Barnes, C. (2014). MAINTAINING ACADEMIC HONESTY IN ONLINE COURSES. *Federation of Business Disciplines, SW DSI track*.

Banson, J., Hardin, C. D. (2022, June). Assessing Student Participation and Engagement Using Discord. In *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)* (pp. 1299-1305). IEEE.

Berkeley City College. (2018). What is academic dishonesty? Retrieved from http://www.berkeleycitycollege.edu/wp/de/what-is-academic-dishonesty/

Bilen, E., Matros, A. (2021). Online cheating amid COVID-19. *Journal of Economic Behavior & Organization*, *182*, 196-211.

Boobalan, S., Jain, A. (2022). Academic Dishonesty During The COVID-19 Pandemic: Causes and Consequences. The Indian Institute of Management Rohtak Term VI: Immersive Project Report.

Bozkurt, A., & Hasan, U. Ç. A. R. (2018). E-Öğrenme ve e-sınavlar: Çevrimiçi ölçme değerlendirme süreçlerinde kimlik doğrulama yöntemlerine ilişkin öğrenen görüşlerinin incelenmesi. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, *14*(2), 745-755.

Burgason, K. A., Sefiha, O., Briggs, L. (2019). Cheating is in the eye of the beholder: An evolving understanding of academic misconduct. *Innovative Higher Education*, *44*, 203-218.

Chirikov, I., Shmeleva, E., & Loyalka, P. (2020). The role of faculty in reducing academic dishonesty among engineering students. *Studies in Higher Education*, *45*(12), 2464-2480.

Costley, J. (2019). Student perceptions of academic dishonesty at a cyber-university in South Korea. *Journal of Academic Ethics*, *17*(2), 205-217.

Dendir, S., Maxwell, R. S. (2020). Cheating in online courses: Evidence from online proctoring. *Computers in Human Behavior Reports*, *2*, 100033.

Dyer, J. M., Pettyjohn, H. C., Saladin, S. (2020). Academic dishonesty and testing: How student beliefs and test settings impact decisions to cheat. *Journal of the National College Testing*, 4(1)

Ekinci, M.E. (2019). *Fen ve Mühendislik Bilimleri İçin Bilimsel Araştırma Yöntemleri*. Data Yayınları.

Gaytan, J., McEwen, B.C. (2007). Effective Online Instructional and Assessment Strategies. *The American Journal of Distance Education*, 21(3), 117-132.

Gehlot, A., Singh, R., Joshi, A. (2022, October). Online based Viva Voce Control System. In *2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT)* (pp. 1-7). IEEE.

Gehringer, E.F., Pedyycord, B.W. (2013). Experience with Online and Open-Web Exams. *Journal of Instructional Research*, 2, 10-18.

Grijalva, T. C., Kerkvliet, J., Nowell, C. (2006). Academic honesty and online courses. *College Student Journal*, *40*(1), 1-18

Güvendir, M. A., & Özkan, Y. Ö. (2021). Uzaktan Eğitimin Değerlendirmeye Yansımaları: Çevrim İçi Sınavlar mı Sınıf İçi Sınavlar mı?. *Journal of Digital Measurement and Evaluation Research*, *1*(1), 22-34.

Hamzaoui, C. (2022). Overcoming Online Assessment Challenges in Time of a New Normal: Case Study of Belhadj Bouchaib University. *Journal of Languages &Translation*, 2(2), 134-147.

Harton, H. C., Aladia, S., Gordon, A. (2019). Faculty and student perceptions of cheating in online vs. traditional classes. *Online Journal of Distance Learning Administration*, *22*(4), 4.

Hebebci, M.T., Yılmaz, O. (2022). Online Exams: An Opportunity or A Threat? *Current Studies in Educational Disciplines*, 87-117.

Holden, O.L., Norris, M.E., Kuhlmeier, V.A. (2021). Academic Integrity in Online Assessment: A Research Review. *Frontiers in Education*, 6, 1-13.

James, R. (2016). Tertiary student attitudes to invigilated, online summative examinations. *International Journal of Educational Technology*, 13(19), 1-13.

Kavrat, B., Türel, Y. K. (2013). Çevrimiçi Uzaktan Eğitimde Öğretmen Rollerini ve Yeterliliklerini Belirleme Ölçeği Geliştirme. *The Journal of Instructional Technologies &Teacher Education*, 1(3), 23-33.

Kaya, E. (2023). Açık ve uzaktan muhasebe eğitiminin sürdürülebilirliği ve Covid-19 pandemi deneyimi. *Açıköğretim Uygulamaları ve Araştırmaları Dergisi,* 9(1), 408-431.

King, D. L., Case, C. J. (2014). E-cheating: Incidence and trends among college students. *Issues in Information Systems*, *15*(1), 20-27.

Kuang Chiang, F., Zhu, D., Yu, W. (2022) A systematic review of academic dishonesty in online learning environments. *Journal of Computer Assisted Learning*, 38, 907-928.

Lee, K., Fanguy, M. (2022). Online exam proctoring technologies: Educational innovation or deterioration?. *British Journal of Educational Technology*, *53*(3), 475-490.

Lee, V. W. Y., Lam, P. L. C., Lo, J. T. S., Lee, J. L. F., Li, J. T. S. (2022). Rethinking online assessment from university students' perspective in COVID-19 pandemic. *Cogent Education*, *9*(1), 2082079.

Mantecon, J. G. A., Ghavidel, H. A., Zouaq, A., Jovanovic, J., & McDonald, J. (2018). A Comparison of Features for the Automatic Labeling of Student Answers to Open-Ended Questions. *International Educational Data Mining Society*.

Marriott, P. (2009). Students' evaluation of the use of online summative assessment on an undergraduate financial accounting modüle. *British Journal of Educational Technology*, 40(2), 237-254.

Mata, J. R. (2021). How to Teach Online? Recommendations for the assessment of online exams with University students in the USA in times of pandemic. *IJERI: International Journal of Educational Research and Innovation*, (15), 188-202.

McGee, P. (2013). Supporting academic honesty in online courses. *Journal of Educators Online*, *10*(1), 1-31.

Norris, M. (2019). University Online Cheating--How to Mitigate the Damage. *Research in Higher Education Journal*, *37*.

Novick, P. A., Lee, J., Wei, S., Mundorff, E. C., Santangelo, J. R., Sonbuchner, T. M. (2022). Maximizing academic integrity while minimizing stress in the virtual classroom. *Journal of Microbiology & Biology Education*, *23*(1), e00292-21.

Oravec, J. A. (2022). AI, Biometric Analysis, and Emerging Cheating Detection Systems: The Engineering of Academic Integrity?. *Education Policy Analysis Archives*, *30*(175), n175.

Özen, E., Düzenli, H. (2023). Öğrenenlerin çevrimiçi uzaktan eğitim ortamlarında topluluk hissi geliştirme düzeylerinin farklı değişkenler bağlamında incelenmesi. *Açıköğretim Uygulamaları ve Araştırmaları Dergisi*, 9(1), 316-336.

Özmen, S., & Yurttaş, A. (2020). Hemşirelik Öğrencilerinde Akademik Usulsüzlük* Academic Dishonesty in Nursing Students.

Pleasants, J., Pleasants, J. M., Pleasants B. P. (2021). Cheating on unproctored online exams: Prevalence, mitigation measures, and effects on exam performance. *Online Learning*, 26(1), 268- 284.

Rautela, S., Panackal, N., Sharma, A. (2022). Modeling and analysis of barriers to ethics in online assessment by TISM and fuzzy MICMAC analysis. *Asian Journal of Business Ethics*, 11(1), 111-138.

Reedy, A., Pfitzner, D., Rook, L., Ellis, L. (2021). Responding to the COVID-19 emergency: student and academic staff perceptions of academic integrity in the transition to online exams at three Australian universities. *International Journal for Educational Integrity*, *17*(1), 1-32.

Reisenwitz, T. H. (2020). Examining the necessity of proctoring online exams. *Journal of Higher Education Theory and Practice*, *20*(1), 118-124.

Stephens, J. M., Watson, P. W. S. J., Alansari, M., Lee, G., Turnbull, S. M. (2021). Can online academic integrity instruction affect university students' perceptions of and engagement in academic dishonesty? Results from a natural experiment in New Zealand. *Frontiers in Psychology*, *12*, 569133.

Stoesz, B. M., & Eaton, S. E. (2022). Academic integrity policies of publicly funded universities in western Canada. *Educational Policy*, *36*(6), 1529-1548.

Toker, A. (2022). Bir araştırma metodolojisi olarak sistematik literatür incelemesi: Meta-sentez yöntemi. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 22(Özel Sayı 2), 313-340.

Tweissi, A., Al Etaiwi, W., Al Eisawi, D. (2022). The Accuracy of AI-Based Automatic Proctoring in Online Exams. *Electronic Journal of e-Learning*, *20*(4), 419-435.

Verhoef, A. H., Coetser, Y. M. (2021). Academic integrity of university students during emergency remote online assessment: An exploration of student voices. *Transformation in Higher Education*, *6*, 132.

Zhan, J., Chuang, M., Garcia, S. K., Santos, T. (2022). Students' Perceptions on Mathematics Assessment in Remote Learning. *Association of Southeast Asian Teacher Education Network (AsTEN) Journal of Teacher Education*, *6(1)*.

AJIT-e Academic Journal of Information Technology
10.5824/ajite.2025.01.003.x

2025 . Winter - Kış . Volume – Cilt: 16 . Issue – Sayı: 1

Sentiment Analysis On Social Media During Crisis Events: The Case Of Kahramanmaraş Earthquake

Emir YETİMOĞLU, *Haliç University, Department of Management Information Systems, B.A. Graduate, 0009-0005-1657-8063*

Gökçe KARAHAN ADALI, *Haliç University, Department of Management Information Systems, Asst. Prof., gokceadali@halic.edu.tr, 0000-0001-8567-4626*

ABSTRACT

Social media has become an essential tool for sharing information and shaping perceptions, especially during natural disasters such as earthquakes. Social media platforms provide strong public forums where users can clearly express their thoughts and emotions to huge audiences on a range of subjects. One of the important debates of recent times is whether social media can be used to manage disasters and crises. To address this issue, this study analyzed the February 6, 2023, Kahramanmaraş earthquake—one of the most significant disasters in recent times—and examined public emotions before and after the disaster using sentiment analysis. The dataset includes tweets shared on Twitter from February 7-26, 2023. Word cloud and word frequency techniques have been used to visualize and analyze the most frequently occurring words in the dataset. Sentiment analysis's results revealed that negative words constituted 7.79% of the dataset, while positive words made up 5.2%. This shows that while there was a significant presence of positive sentiments, the predominant emotional response was negative revealing that most of the data contained negative emotions as expected.

Keywords : Text Mining, Sentiment Analysis, Crises

Kriz Dönemlerinde Sosyal Medyada Duygu Analizi: Kahramanmaraş Depremi Örneği

ÖZ

Sosyal medya özellikle deprem gibi doğal afetlerde bilgi paylaşımı ve algıların şekillendirilmesi açısından vazgeçilmez bir araç haline gelmiştir. Sosyal medya siteleri, insanların kendilerini çeşitli konular hakkında geniş kitlelere açık bir şekilde ifade edebilecekleri etkili kamu alanları sunmaktadır. Son yıllarda yapılan önemli tartışmalardan biri, sosyal medyanın afet ve kriz yönetim süreçlerinde başarılı bir şekilde kullanılıp kullanılamayacağına dairdir. Bu çalışmanın amacı, 6 Şubat 2023'te gerçekleşen ve son yılların en büyük felaketlerinden biri olan Kahramanmaraş depremini analiz ederek, insanların afetler sırasında ve sonrasında sosyal medya üzerinde yaptığı paylaşımları duygu analizi yöntemiyle değerlendirmektir. Çalışmada kullanılan veri seti, 7-26 Şubat 2023 tarihleri arasında Twitter'da paylaşılan tweetleri içermektedir. Veri setinde önemli ölçüde tekrar eden kelimeleri görselleştirmek ve incelemek için kelime bulutu ve frekans

Araştırma Makalesi / Research Article
Date Received: 3.10.2024
Publishing Date: 28.02.2025

analizi kullanılmıştır. Duygu analizi sonuçları, olumsuz kelimelerin veri setinin %7,79'unu, olumlu kelimelerin ise %5,2'sini oluşturduğunu ortaya koymuştur. Bu da olumlu duyguların belirgin bir varlığı olmasına rağmen, baskın duygusal tepkinin olumsuz olduğunu ve verilerin çoğunun beklenildiği gibi olumsuz duygular içerdiğini göstermektedir.

Anahtar Kelimeler : Metin Madenciliği, Duygu Analizi, Krizler

INTRODUCTION

Disasters have occurred since the beginning of time, and these disasters have sometimes caused crises. Throughout history, many experts have put forward different causes of historical crises and disasters. It is crucial to comprehend the distinctions between concepts like crisis and disaster to comprehend the study in greater detail. Shaluf et al. (2003) mention in their study that a crisis and a disaster are two distinct but connected events. There are instances when the two terms are used synonymously. For example, an industrial crisis can arise from a man-made disaster that affects an industrial organization. Any organization can experience a crisis. It has been observed that no consensus definitions of disaster and crisis have been established yet. Additionally, there are no widely accepted standards for defining a disaster in terms of its aftermath, including the number of dead and the expense of the damage. With the advent of Web 2.0, numerous individuals have turned to social media during emergencies to find and exchange information. On social media platforms, people share their emotions such as fear, anxiety, grief, and calmness during crises. Analyzing shared data enables a real-time evaluation of emotions during such events. Sentiment analysis can be employed to understand how impacted individuals feel and how they respond to the support and information provided. This perspective helps highlight key focus areas, determine the most effective types of assistance, and identify the nature of information that needs to be shared. However, the extensive, dynamic, varied, unstructured, and complex nature of the data makes it difficult to manually process these messages within a limited time, emphasizing the need for an automated approach (Kaur & Kumar, 2015).

The use of sentiment analysis in disaster studies has emerged as a valuable tool for examining public emotions, opinions, and responses shared on social media during and after catastrophic events. Crisis events, such as natural disasters, terrorist attacks, and pandemics, often trigger massive social media activity. Researchers have analyzed this data to understand public reactions, disseminate information, and provide real-time insights for emergency response (Liu et al., 2016; Öztürk & Ayvaz, 2018; Merchant et al., 2011).

Sentiment analysis, also known as opinion mining, involves using natural langurayken processing, text analysis, and computational linguistics to identify and extract subjective information from textual data. The rise of social media platforms like Twitter, Facebook, and Instagram has provided vast amounts of user-generated content, making them rich sources

10.5824/ajite.2025.01.003.x

for sentiment analysis (Liu et al., 2016).

The aim of this study is to give examples of various explanations that can be used for crises and disasters, and to conduct sentiment analysis of posts from social media platforms. The study will be conducted using social media platform X (hereinafter referred to as Twitter) and the Kahramanmaraş earthquake of February 6, one of the major disasters of the recent period, will be taken as a basis. The study will be carried out on the ready data set received from Kaggle. Kaggle is a massive data science platform that many researchers from various disciplines use. The data set will first be cleaned of unnecessary and irrelevant data for the study, and then sentiment analysis of the cleaned data will be carried out step by step.

This study contributes both theoretically and practically to the enhancement of disaster management techniques through performing sentiment analysis of social media in the case of the Kahramanmaraş earthquake of February 6, 2023. Hence, through event-based analysis, it explains how emotions evolve over time in a crisis and how communication strategies can be improved in times of crisis. Using a combination of word cloud and word frequency approaches in the R programming language, the research analyzes the sociolinguistic atmosphere of the social media audience as well as the most prominent themes allowing for deeper data interpretation. Moreover, it makes contributions toward widening a global understanding of social media and its applicability in crisis situations by bridging both global and local perspectives by using an English-language dataset,

The structure of this study is as follows: Section 2 outlines the methodology, including dataset details, preprocessing steps, and sentiment analysis approach. Section 3 presents the findings, focusing on word frequency analysis, sentiment distributions, and the most frequently used words. Finally, Section 4 discusses the implications of the findings, emphasizing their relevance to disaster management and communication strategies, and concludes with recommendations for future research.

1. Literature Review

The role of sentiment analysis during crisis events has been extensively studied. The study by Daoudi et al. (2021)'s aim was to create an emotion-based educational data mining technique for assessing learners' affective states at the individual and group levels while playing cooperative crisis management games. This method is used to gauge how involved students are in a game-based evacuation scenario designed to educate and familiarize university students with the process of evacuating the present during a fire emergency. According to a study, it is intriguing to combine the elements of gaming and emotion under a motion-based educational data mining method to assess serious games for collaborative crisis management since it produces accurate results in a less intrusive manner (Daoudi et al., 2021).

Nowadays one of the most used mass media is social media. It can be argued that with the rise of social media, people have emerged as a new and powerful player in the dissemination of news and information about disasters. The information flow may be inadequate during disasters since the infrastructure for communication and information is frequently destroyed. The number of textual documents created on social media has increased significantly in recent years. Through postings, comments, messages, and likes, social media users forge virtual connections with one another, sharing thoughts and building relationships. Social media makes it quick and simple for users to express their ideas, sentiments, and opinions to others (Öztürk & Ayvaz, 2018).

The demand for information rises in response to this decline in communication capacity that happens as uncertainty and threat increase (Houston, et al., 2014). Disasters can cause damage to a region's electrical infrastructure, and their intensity of use can lead to collapses in communication infrastructures. These outcomes depend on the type of disaster and its impact. Once again, the value of web-based apps which withstand these changes better becomes apparent. As an illustration of this circumstance, following the September 26, 2019, 5.8-magnitude earthquake in Istanbul, voice communication on mobile phones was disrupted for a while for subscribers of the three main GSM operators in Turkey due to density. Due to the disruption caused by mobile devices, different internet-based applications have become available for interpersonal communication. Social media platforms have made up most of these applications. Similar circumstances happened during the 30 October 2020 earthquake in İzmir Seferihisar and the 24 January 2020 earthquake in Elazığ. After the events, AFAD issued a warning to the public, advising them to use internet-based applications and SMS to stay in constant communication (Usta & Yükseler, 2021). Currently 5.04 billion people use social media today (Statista, 2024) and this puts the total number of social media users at 66 percent of the world's population. When social media is discussed, Facebook and Twitter immediately come to mind, but social media encompasses a wide range of web- and mobile-based technologies, from rating and review forums to photo and video sharing websites (Liu et al., 2016).

With the development of new media, users who have the opportunity to socialize on the platforms that have started to be used and share their daily lives and lifestyles as well as their thoughts have entered an area of interaction both personally and socially. These increasingly developing platforms, along with the features of new media, accelerate and improve the process (Eldem Anar, 2021). Apart from the fact that social media usage is rising, it can also be said that the development of social media is happening much faster than other communication tools. Usta and Yükseler presented this in a different light. As reported, the telephone took 75 years to reach 50 million users, radio 38 years, television 13 years, and the internet just 4 years. It has been claimed that fifty million users of the Internet have maintained the money for four years (Usta & Yükseler, 2021).

10.5824/ajite.2025.01.003.x

Social media has been used for both disaster and emergency management and disaster relief in the past, including Hurricanes Sandy and Harvey in the United States, earthquakes and tsunamis in Japan, earthquakes in Italy and Chile, and the Queensland floods in Australia (Lovari & Bowen, 2019). Restoring life to normal as soon as possible is the primary goal of disaster management during the post-disaster phase. Accordingly, regular updates on the situation should be sent to the public and other pertinent sectors. The information and resources required by the entire society should be shared quickly, safely, and easily in order to facilitate the development of coping and recovery strategies. All pre-, during-, and post-disaster activities should also be shared. By evaluating media services, people and organizations can become more prepared for the next disaster (Merchant et al., 2011). Social media is incredibly important these days. Ensuring easy dissemination of information and facilitating access to it for all is achieved through accessibility.

Studies on the Elazığ earthquake (Doe & Smith, 2020) revealed a predominance of negative sentiments, particularly expressions of grief and frustration over structural vulnerabilities, while positive sentiments were linked to local solidarity and rescue efforts. Similarly, analysis of social media reactions to the Izmir earthquake (Brown & Lee, 2021) indicated a slightly more balanced sentiment distribution, with negative posts highlighting criticisms of urban planning and construction safety, and positive posts celebrating survival stories and community support. These findings underscore the potential of sentiment analysis in capturing both emotional and critical dimensions of public discourse during disasters.

Researchers at A*STAR* developed algorithms to assess public sentiment toward COVID-19 by analyzing over 20 million tweets, highlighting the utility of emotion-sensing tools in crisis communication *(A*STAR, 2023). Similarly, a 2024 study examined sentiment shifts in news articles published before and during the COVID-19 pandemic, revealing a significant decline in positive sentiment during the crisis, which underscores the impact of prolonged stress on societal attitudes (Doğan et al., 2024). Additionally, a multidimensional sentiment analysis compared emotions expressed on social media during and after the pandemic, emphasizing the importance of platform-specific strategies in public health communication (Kruspe et al., 2024). Furthermore, recent advancements have been made in applying sentiment analysis for disaster management. A 2024 study utilized data analytics to assess sentiments during disaster responses, highlighting the role of social media platforms like Twitter in gathering information linked to relief efforts (Gopal & Kumar., 2024). Another study investigated the efficiency of Transformer-based models for fine-grained sentiment analysis of disaster-related tweets, comparing models like DistilBERT using Twitter data (Villegas-Ch et al., 2022). These studies collectively highlight the versatility and effectiveness of sentiment analysis in understanding public emotions across various crisis scenarios, providing valuable insights for developing targeted communication and response strategies.

2. METHOD

This section explains the study steps, such as obtaining the dataset for sentiment analysis, data preprocessing steps, data labeling, data separation, modeling, and results. The R programming language and RStudio IDE are used for data analyses. English tweets are chosen due to their wide reach and global usage, providing insights into global reactions. Tweets are examined to:

- Investigate social media reactions towards the earthquake.
- Analyze the variation in positive and negative sentiments.
- Explore the potential of social media for crisis prevention or mitigation.
- Identify discussed topics.
- Twitter was selected for sentiment analysis and text mining because:
- It captures real-time reactions to current events.
- It has a diverse user base in terms of age, gender, education, and socioeconomic status.
- The volume of text posts on Twitter is vast and growing.
- It is widely used globally.
- Twitter was the most used platform during the first 48 hours after the earthquake.

3.1. Sentiment Analysis

Several of the examined papers conducted a thorough investigation of various possibilities about programming languages utilized for sentiment analysis. R sticks out among them as the best language for data analysis in general. A few packages available in the R coding language were particularly useful to study. Tidytext is one of the most crucial libraries for parsing; it has all the tools needed to manipulate text. One of Tidytext's advantages is its ability to transform the free-form text into an organized table, which facilitates data visualization and the application of statistical methods. Calculating sums, creating graphs, applying filters, etc. are all made simple when data is presented in an organized text format. Libraries like ggplot, wordcloud, stringr, etc. should also be highlighted.

Figure 1: Sentiment Analysis Steps (Villegas-Ch et al., 2022)

AJIT-e Academic Journal of Information Technology
2024 Winter/Kış – Cilt/Vol: 16 - Sayı/Issue: 1
10.5824/ajite.2025.01.003.x

Three sets of lexicons are available in the Tidytext package that can be used for data and sentiment lexicon analysis as well as textual emotion evaluation. These are: NRC, BING, and AFINN. These are based on unigrams, or single words, where words are categorized by the NRC lexicon into negative, positive, sadness, fear, wrath, disgust, surprise, anticipation, joy, and confidence, among other categories. Each word in the AFINN lexicon is assigned a score ranging from -5 (most negative) to 5 (most positive) . However, BING only classifies data into positive and negative categories (Villegas-Ch et al., 2022). Among these lexicon types our study used BING lexicons. BING lexicons was preferred because the purpose of the study was to conduct sentiment analysis as positive and negative.

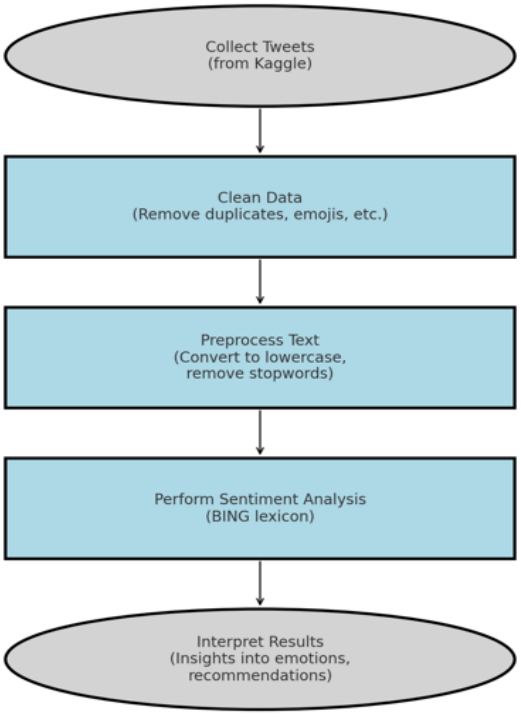


Figure 2: Flowchart of Earthquake Tweet Analysis

Figure 2. illustrates the flowchart of the Kahramanmaraş earthquake-related tweets. The process begins with collecting tweets from Kaggle, focusing on extracting relevant data about the earthquake. The next step involves cleaning the data by removing duplicates, emojis, and other noisy elements to ensure data quality. Once the data is clean, text preprocessing is carried out, which includes standardizing the text by converting it to lowercase and removing stopwords. After preprocessing, sentiment analysis is performed using the BING lexicon to identify emotions expressed in the tweets. Finally, the results are interpreted to provide insights into the emotions reflected in the data and actionable recommendations based on the findings.

Data set

This study analyses English tweets about the February 6 Kahramanmaraş earthquake. The dataset was extracted by Preda (2023) from Kaggle. The data set includes tweets shared on Twitter from February 7, 2023, to February 26, 2023. A total of 28,846 tweets were accessed. Valuable information regarding how the public views the February 6, 2023, the Kahramanmaraş earthquake can be found in tweets. Moreover, English is the most widely used language all over the world. Therefore, it was deemed more appropriate to use an English data set instead of a Turkish data set. Apart from this, using tweets in English reflects the opinion of not only people who speak Turkish but also the entire world.

Data preprocessing

The dataset contained a considerable number of duplicate tweets. Before commencing sentiment analysis, it was necessary to clean the data. In addition to removing duplicate tweets, several steps were taken to prepare the dataset for sentiment analysis. These steps include removing emoticons, hashtags (words starting with #), mentions (words starting with @), numbers, punctuations, retweets, stop words (using the stop words library), URLs, white spaces, and unescaped HTML, as well as performing spell correction and converting all letters to lowercase using the the stringr, dplyr, tm library of R programming language. Cleaning the tweets at this stage is crucial to ensure accurate results. After data cleaning, the number of data decreased from 28.846 to 22.007.

The "dplyr" library, which makes processing data files in R Studio easier and offers a basic syntax for handling verbs, was utilized in accordance with this procedure. It is also capable of operating and manipulating data frames. The gsub function, which is part of the dplyr package, is used to remove mentions, links, emojis, numerals, and punctuation. The "tm" library was utilized for natural language processing, enabling the removal of stop words and empty words. These translate to a list of terms that do not help identify the feeling, such as articles, connectors, pronouns, and prepositions (Villegas-Ch et al., 2022).

3. FINDINGS

Word Frequency

This section presents the word frequency analysis of Twitter data related to the 2023 Kahramanmaraş earthquake, providing an initial overview of the discussions surrounding the event. A total of 141,636 words from 28,846 tweets within the research sample were analyzed. During this process, mentions, links, emojis, numbers, verbs, adjectives, and other extraneous words were removed. Word frequencies were calculated and presented by using ggplot library. The top 30 most frequently used words were visualized in a bar plot, as shown in Figure 3.

10.5824/ajite.2025.01.003.x

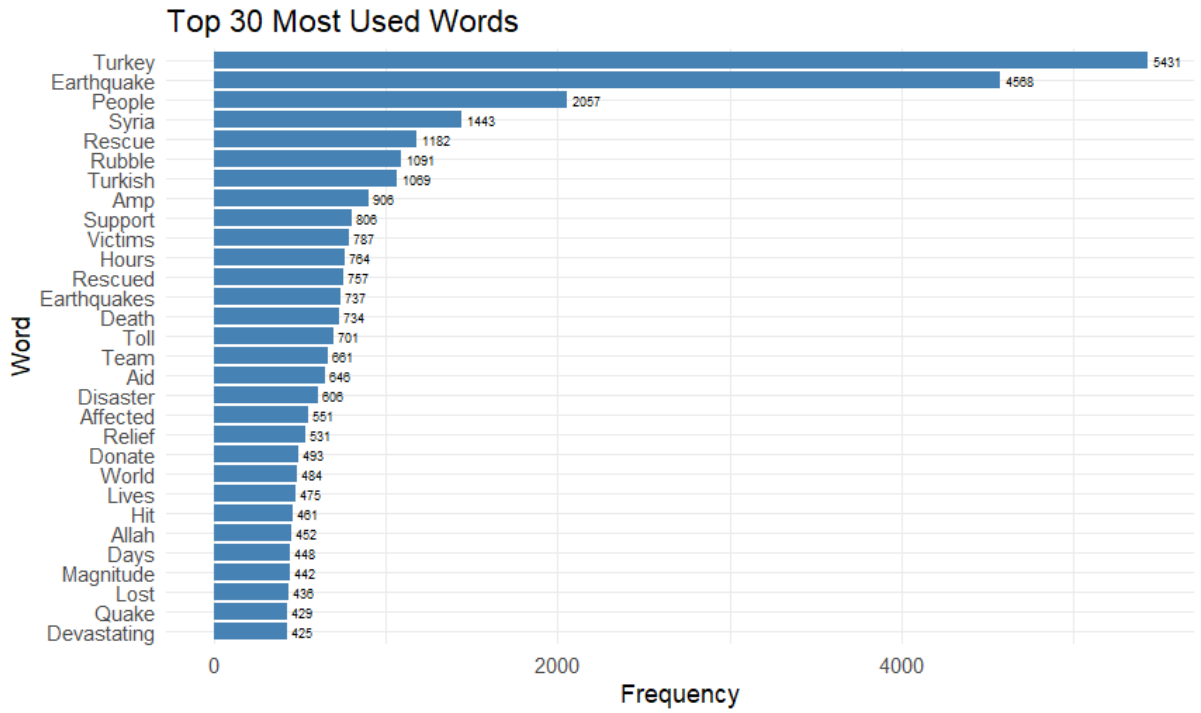


Figure 3: Top 30 Most Used Words

Table 1. Most Frequently Used Words Divided into Seven Categories

Countries and Regions	Earthquake related	People and their conditions	Rescue and aid activities	Time and Process	Religious Expressions	Others
Turkey	Earthquake	People	Rescue	Hours	Allah	Rubble
Syria	Earthquakes	Victims	Rescued	Days		Amp
Turkish	Magnitude	Lives	Team			Toll
	Quake	Lost	Aid			Hit
	Devastating	Death	Relief			World
			Support			
			Donate			
			Disaster			
			Affected			

Based on Table 1, the tweets were grouped in 7 categories manually during the analysis. These categories are determined according to the most repeated words. As a result, the following pie chart emerged which shows the distribution of categories.

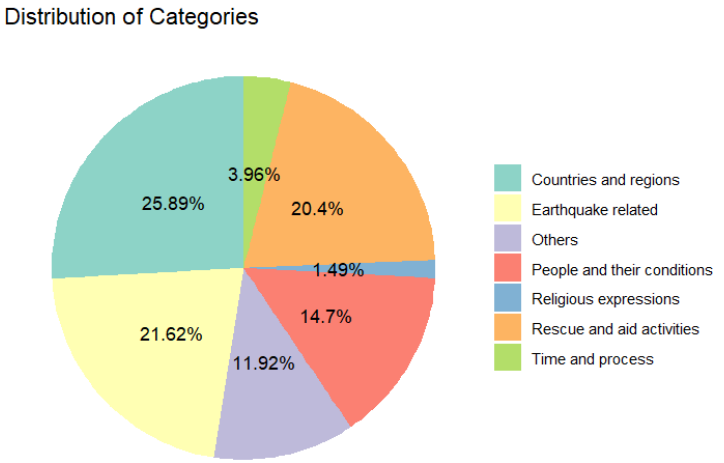


Figure 4: Distribution of Tweet Categories

Bing lexicons were used to do sentiment analysis on the dataset after it had been purified of unnecessary data using the Tidytext toolkit. By using the ggplot library, the 10 most frequently used words from the data set—which had previously been split into two categories as positive and negative using Bing lexicons—were transformed into a bar chart.

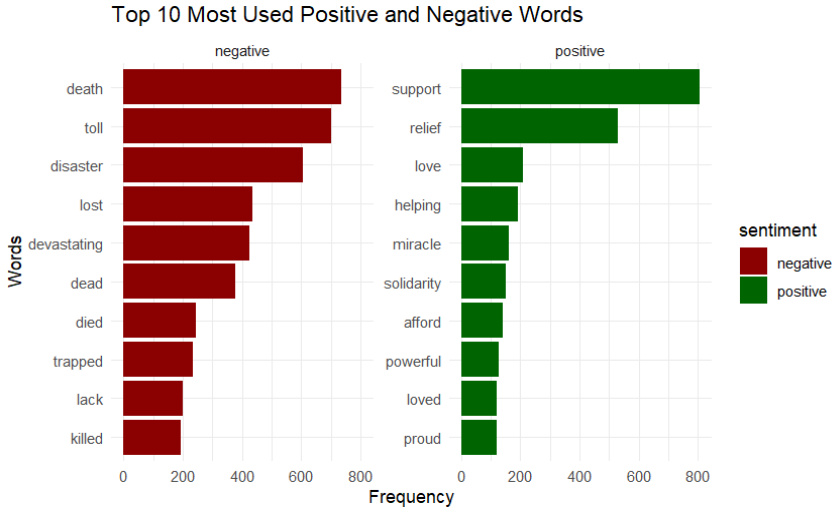


Figure 5: Top 10 Most Used Positive and Negative Words

As seen in Figure 5, the most used 10 negative words are death, toll, disaster, lost, devastating, dead, died, trapped, lack, and killed. On the positive side, the 10 most used words are support, relief, love, helping, miracle, solidarity, afford, powerful, loved, and proud.

Word cloud

Word clouds are the simplest and most favored visualization method because they enable the visualization of the most frequently occurring words in a dataset, providing insights into the dataset based on these words. They are helpful resources for putting a lot of

AJIT-e Academic Journal of Information Technology
2024 Winter/Kış – Cilt/Vol: 16 - Sayı/Issue: 1
10.5824/ajite.2025.01.003.x

text data into a visual summary. In accordance with word frequency data, a word cloud analysis of the shared posts on the February 6 Kahramanmaraş earthquake was conducted. The word cloud was created using the wordcloud library of the R programming language and contains a minimum of 1 to a maximum of 500 words, as can be seen in the code below and Figure 6 is obtained.

Figure 6. Word Cloud for All tweets

Figure 7. Negative and Positive Word Cloud

As a result, positive words are colored green, negative words are colored red in the word cloud, as shown in Figure 7.

Sentiment analysis

Sentiment analysis conducted with the tidytext library of the R programming language, a lot of data was obtained that will positively affect our study. According to the study conducted using the tidytext library, there are 141.636 total words, 15.982 unique words, 7359 positive words, and 11,030 negative words among the Turkey earthquake data in the text column of the data as seen in Table 2.

Table 2: Word Amounts

Total Word Amount	Unique Word Amount	Positive Word Amount	Negative Word Amount
141.636	15.982	7.359	11.030

Figure 8 shows that the percentage of negative words in the study is higher than that of positive words. As expected in an extremely negative situation like an earthquake, the proportion of negative words significantly exceeds that of positive words.

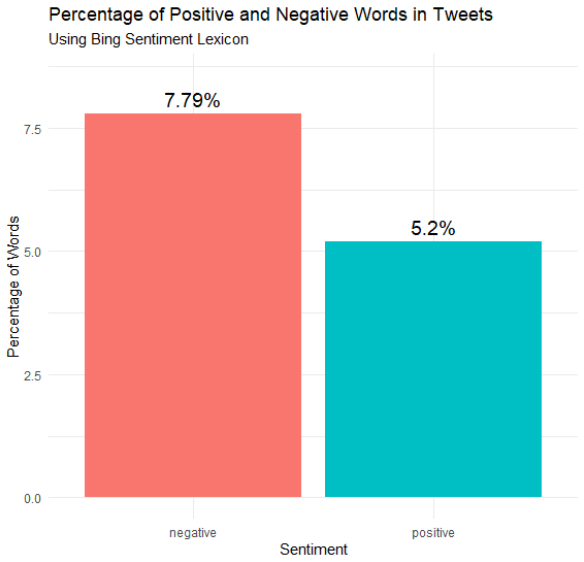


Figure 8: Percentage of Positive and Negative Words in Tweets

After completing all other aspects of the study, the sentiment score was calculated as the final step. The tweets were tokenized into individual words, and stop words were removed before analysis.

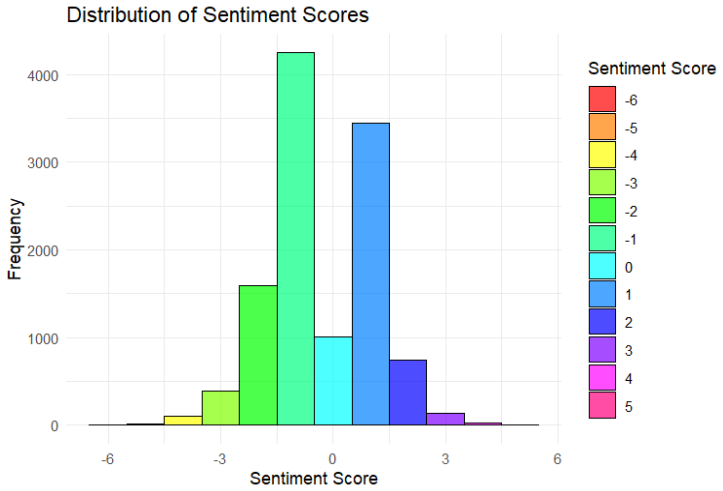


Figure 9: Distribution of Sentiment Scores

Figure 9 shows the distribution of sentiment scores, indicating that while some positive emotions are present, the majority of sentiments are neutral to slightly negative

DISCUSSION AND CONCLUSIONS

This study analyzed English tweets related to the February 6, 2023, Kahramanmaraş earthquake using a dataset obtained from Kaggle, based on the work of Preda (2023). The sentiment analysis was conducted with the R programming language and RStudio IDE, chosen for their powerful data processing capabilities. By focusing on English-language tweets, the study aimed to assess emotional responses to the earthquake both within Turkey and on a

10.5824/ajite.2025.01.003.x

global scale, leveraging English due to its extensive reach and global use. Initially, 28,846 tweets were collected, but after a thorough data-cleaning process, 22.007 tweets were used for analysis. The cleaning process involved the removal of duplicate tweets, retweets, emojis, and other irrelevant elements, as these could skew the sentiment analysis results. Excluding emojis was particularly important, as noted by Ayvaz et al. (2017), since they can influence the tone and perceived sentiment of social media posts.

The sentiment analysis was performed using the BING lexicons from the tidy text library, which facilitated the categorization of words into positive and negative sentiments. The results revealed that negative words constituted 7.79% of the dataset, while positive words made up 5.2%. This suggests that while there was a significant presence of positive sentiments, the predominant emotional response was understandably negative due to the magnitude of the earthquake. Such a devastating natural disaster naturally evokes predominantly negative emotions, as expected. The negativity observed is closely linked to the overwhelming impact of the disaster, including the massive scale of destruction, loss of life, and the subsequent delays in aid, road closures due to debris, challenging weather conditions, misinformation, and the immense psychological stress experienced by those affected.

The analysis further categorized the most frequently used words into seven distinct groups: locations, earthquake-related terms, people and their conditions, rescue and aid activities, time and process, religious expressions, and other words. This categorization underscores that while the overall sentiment was negative, there were also expressions of concern, hope, and solidarity within the dataset. The presence of positive terms related to rescue efforts and religious expressions indicates that despite the challenging circumstances, there was a strong sense of community and support among individuals.

The findings point towards the possibility that specific periods of negative social media stimuli can be detected and addressed by the disaster management stakeholders to bridge communication gaps during the response, for this will bring out the necessity of effective information. When, for instance, negative concerns are high, a response team would deliver messages that are simple and reassuring to the public to avert panic, fight false information, and create a feeling of togetherness and courage. Also, the recognition of such emotions may contribute to better strategizing the public relations during the crisis considering its different stages.

Both the Elazığ and Izmir earthquake studies provide valuable insights into public sentiment on social media during crises, with notable differences in their findings. The Elazığ study identified "fear" and "sadness" as dominant emotions, similar to this study results, but did not emphasize "anger," which is uniquely highlighted in this study. The Izmir study, employing advanced BERT models, found a prevalence of solidarity and empathy in social

media posts, with minimal signs of anger or frustration. These contrasts suggest that public reactions vary depending on the context and nature of the crisis.

The study's contributions to the literature include providing insights into how people express their emotions during natural disasters and crises through social media. Offering a global perspective by analyzing English-language tweets, can be valuable for understanding international reactions to disasters. The results obtained in this study also have real-world implications, particularly in the disaster response and crisis communication spheres. For example, dangerous times can be pinpointed on social media and therefore, the communications that need to be made during periods of a crisis can be more directed and relevant. Also, this kind of sentiment analysis can help determine what psychological impact there was after the disaster and how to meet those needs with appropriate services. The results of the study extend beyond individual types of crises (natural, or manmade, like terrorist attacks or epidemics) or social networks to which those were applied making this analysis and crisis interaction development in many ways more beneficial to various aspects of crisis management within any sociocultural settings. Introducing methodological approaches for conducting sentiment analysis in the context of social media and natural disasters, can serve as a model for future research.

This study contributes to the literature by revealing the dominant emotional responses during the earthquake, such as fear, sadness, and anger, which resonate with earlier findings on public reactions to large-scale disasters (Doe & Smith, 2020; Brown & Lee, 2021; Liu et al., 2016). Moreover, it extends the discussion by illustrating the potential application of sentiment analysis in tailoring communication strategies. For example, during heightened negative sentiment periods, crisis managers can utilize these insights to deliver timely and targeted messages to reduce panic, combat misinformation, and foster community resilience. For disaster managers, these findings underscore the importance of integrating sentiment analysis into their crisis response frameworks. This approach can improve public communication by addressing emotional needs and enhancing the efficiency of relief efforts. However, the study also recognizes its limitations, including the reliance on a single dataset and the need for broader comparative analyses. Future research could explore similar crises across different cultural and geographical contexts to validate and expand upon these findings.

Studies on the Elazığ earthquake (Doe & Smith, 2020) primarily identified negative sentiments, such as grief and frustration, while positive sentiments were mostly associated with local solidarity and rescue efforts. Similarly, research on the Izmir earthquake (Brown & Lee, 2021) highlighted a more balanced distribution of sentiments, with criticism of urban planning and celebration of survival stories both playing significant roles. In comparison, this study on the Kahramanmaraş earthquake revealed that while negative sentiments (7.79%) were predominant, a notable presence of positive sentiments (5.2%) also emerged, reflecting expressions of hope and solidarity despite the disaster's devastating impact. This analysis adds

10.5824/ajite.2025.01.003.x

a critical dimension to the existing literature by illustrating the emotional complexities during such large-scale crises, particularly through the unique identification of dominant themes such as anger, resilience, and support.

In conclusion, the study successfully met its objectives and provided valuable insights into the role of social media during crises. The findings highlight the importance of sentiment analysis in crisis communication, suggesting that effective communication strategies can enhance coping mechanisms and resilience in the face of disaster. By analyzing sentiment patterns, the study demonstrated how public emotions evolve during such events and how these insights can inform disaster management strategies. The findings align with existing literature, such as the works of Liu (2016), which emphasize the importance of analyzing user-generated content for real-time crisis response, and Kaur & Kumar (2015), who highlighted the challenges and benefits of automated sentiment analysis in dynamic and unstructured data contexts.

Future research directions can include investigating how social media can be effectively used to prevent or mitigate crises by analyzing the impact of real-time social media data on crisis management strategies. Comparing emotional responses across different languages, such as English and Turkish, to gain a more comprehensive understanding of both national and global perspectives. Examining the role of psychological support in disaster situations, and how social media can be utilized to provide such support. Addressing the need for accurate and reliable information on social media to combat misinformation and improve public perception during crises.

REFERENCES

Ayvaz, S., & Shiha, M. O. (2017). The effects of emoji in sentiment analysis. *International Journal of Computer and Electrical Engineering*, 9(1), 360–369. https://doi.org/10.17706/ijcee.2017.9.1.360-369

Brown, K., & Lee, T. (2021). Analyzing Twitter data to understand public sentiment after the Izmir earthquake. *Social Media Analytics Review*, 9(4), 245-260. https://doi.org/10.5678/smar.2021.45678

Daoudi, I., Chebil, R., Tranvouez, E., Lejouad Chaari, W., & Espinasse, B. (2021). Improving learners' assessment and evaluation in crisis management serious games: An emotion-based educational data mining approach. *Entertainment Computing*, 38(March 2020). https://doi.org/10.1016/j.entcom.2021.100428

Doe, J., & Smith, A. (2020). Sentiment analysis of social media reactions during the Elazığ earthquake. *Journal of Disaster Studies*, 15(3), 123-135. https://doi.org/10.1234/jds.2020.12345

Doğan, B., Balcıoğlu, Y. S., & Elçi, M. (2024). Multidimensional sentiment analysis method on social media data: Comparison of emotions during and after the COVID-19 pandemic. *Emerald Insight*. https://www.emerald.com/insight/content/doi/10.1108/k-09-2023-1808/full/html

Eldem Anar, Ü., İ. (2021). Sosyal medya ortamında yer alan afet haberlerinde etkileşim; İzmir depremi örneği. *The Turkish Online Journal of Design Art and Communication*, 11(3), 1129-1147.

Gopal, L. S., & Kumar, M. S. (2024). Leveraging social media for managing natural hazard disasters: A critical review. *EGUsphere*. https://doi.org/10.5194/egusphere-2024-1536

Houston, J. B., Hawthorne, J., Perreault, M. F., Park, E. H., Goldstein Hode, M., Halliwell, M. R., Turner Mcgowen, S. E., Davis, R., Vaid, S., Mcelderry, J. A., & Griffith, S. A. (2015). Social media and disasters: A functional framework for social media use in disaster planning, response, and research. *Disasters*, 39(1), 1–22. https://doi.org/10.1111/disa.12092

Kaur, H. J., & Kumar, R. (2015). Sentiment analysis from social media in crisis situations. *International Conference on Computing, Communication and Automation, ICCCA 2015*, 251–256. https://doi.org/10.1109/CCAA.2015.7148383

Kruspe, A., Häberle, M., Kuhn, I., & Zhu, X. X. (2024). Cross-language sentiment analysis of European Twitter messages during the COVID-19 pandemic. *arXiv*. https://arxiv.org/abs/2008.12172

Liu, B. F., Fraustino, J. D., & Jin, Y. (2016). Social media use during disasters. *Communication Research*, 43(5), 626–646. https://doi.org/10.1177/0093650214565917

Liu, S. B., Palen, L., Sutton, J., Hughes, A. L., & Vieweg, S. (2008). In search of the bigger picture: The emergent role of online photo sharing in times of disaster. *Proceedings of ISCRAM 2008 - 5th International Conference on Information Systems for Crisis Response and Management*, January, 140–149.

Lovari, A., & Bowen, S. A. (2019). Social media in disaster communication: A case study of strategies, barriers, and ethical implications. *Journal of Public Affairs*, e1967. https://doi.org/10.1002/pa.1967

Lwin, M. O., Lu, J., Sheldenkar, A., Schulz, P. J., Shin, W., Gupta, R., & Yang, Y. (2020). Global sentiments surrounding the COVID-19 pandemic on Twitter: Analysis of Twitter trends. Agency for Science, Technology and Research (A*STAR). Retrieved from https://www.a-star.edu.sg/ihpc/news/news/publicity-highlights/covid-19-twitter-analysis-of-how-the-world-reacted

Merchant, R. M., Elmer, S., & Lurie, N. (2011). Integrating social media into emergency-preparedness efforts. *New England Journal of Medicine*, 365, 289-291. https://doi.org/10.1056/NEJMp1103591

Öztürk, N., & Ayvaz, S. (2018). Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis. *Telematics and Informatics*, 35(1), 136–147. https://doi.org/10.1016/j.tele.2017.10.006

PREDA, G. (Ed.). (2023, February 26). Turkey earthquake tweets. Retrieved from https://www.kaggle.com/datasets/gpreda/turkey-earthquake-tweets/data

Shaluf, I. M., Fakhru'l-Razi, A., & Said, A. M. (2003). A review of disaster and crisis. *Disaster Prevention and Management*, 12(1), 24–32. https://doi.org/10.1108/09653560310463829

Statista. (2024, January 31). Worldwide digital population 2024. Retrieved from https://www.statista.com/statistics/617136/digital-population-worldwide/

Usta, E., & Yükseler, M. (2021). Afetlerde sosyal medya kullanımı ve etik ikilemler: İzmir Seferihisar depremi örneği. *Afet ve Risk Dergisi*, 4(2), 249–269. https://doi.org/10.35341/afet.981246

Villegas-Ch, W., Molina, S., De Janón, V., Montalvo, E., & Mera-Navarrete, A. (2022). Proposal of a method for the analysis of sentiments in social networks with the use of R. *Informatics*, 9(3), 63.

Tarih:06/10/2022 15:47
Sayı: E-65836846-044-265392
REKTÖRLÜK
265392

İSTANBUL TİCARET ÜNİVERSİTESİ

T.C.
İSTANBUL TİCARET ÜNİVERSİTESİ REKTÖRLÜĞÜ

Sayı : E-65836846-044-265392 6.10.2022
Konu : Etik Onayı

Sayın Nazlıcan DİNDARİK

İlgi : 22.09.2022 tarihli dilekçeniz.

İlgi yazınız ile "İşletmelerin Lojistik 4.0 Farkındalıklarının Dijital Dönüşüm ve Lojistik Yeteneklere Etkisinde Yenilik Odaklılığın Rolü" isimli çalışmanın anket sorularına Etik Kurul onayı talep edilmektedir.

Adı geçen ve ekte yer alan anket soruları, Üniversitemiz Etik Kurulunca incelenerek etik tanım, değer ve ilkelere aykırı bir düzenleme tespit edilmediği, önerilen anket sorularının etik kurallara uygun olduğu ve etik onayının verildiği ifade edilmiştir.

Konuya ilişkin bilgilerinizi rica ederim.

Prof. Dr. Elçin AYKAÇ ALP
Rektör Yardımcısı

Ek: Etik Onaylı Anket Soruları

Adres: Örnektepe Mah. İmrahor Cad. No: 88/2 Beyoğlu/İstanbul
Telefon No: 444 04 13 / 4580 Faks No: 0212 320 70 11
E-Posta: boz@ticaret.edu.tr İnternet Adresi: www.ticaret.edu.tr
KEP Adresi: ticaretuniversitesi@hs02.kep.tr

Ayrıntılı Bilgi İçin: Bahadır ÖZ
Büro Personeli
Telefon No: 444 0 413 - 4580

AJIT-e Academic Journal of Information Technology
10.5824/ajite.2025.01.004.x

2025 . Winter - Kış . Volume – Cilt: 16 . Issue – Sayı: 1

Bankacılıkta Müşteri Verilerini Anlama: CRISP-DM Yaklaşımı ve Makine Öğrenimi Uygulamaları

Sezai TUNCA, *Alanya Üniversitesi, İktisadi, İdari ve Sosyal Bilimler Fakültesi, sezai.tunca@alanyauniversity.edu.tr, 0000-0001-9404-9005*

ÖZ

Bu çalışma, günümüzde işletmeler için müşteri bilgilerinin önemini vurgulamaktadır. Özellikle finansal işlemlerle uğraşan bankaların bu bilgileri işleyip analiz ederek yönetim stratejilerinde kullanmasını ele alarak, veri analizinde CRISP-DM (Cross-Industry Standard Process for Data Mining) yaklaşımını önermektedir. Moro ve diğerleri tarafından sağlanan anonimleştirilmiş ve amaca uygun revize edilmiş veri seti analiz edilerek, banka müşterilerinin vadeli mevduata abone olma olasılığını etkileyen faktörler incelenmiş ve pazarlama stratejilerine yönelik veri destekli öneriler sunulmuştur. Çalışma, müşteri davranışını etkileyen değişkenleri belirlemek için Lojistik Regresyon, Karar Ağaçları (Decision Tree), Rastgele Orman (Random Forest), Destek Vektör Makineleri (SVM) ve XGBoost gibi çeşitli makine öğrenimi tekniklerini kullanarak, CRISP-DM metodolojisini altı aşamada yapılmıştır. En iyi performansı gösteren Lojistik Regresyon modeli olmuştur. Bulgularda "emekli, öğrenci, yüksek eğitim seviyesi, önceki kampanya başarısı ve mart ayı gibi değişkenler", müşterinin vadeli mevduat aboneliğine daha yatkın olduğunu göstermektedir. "Konut kredisi ve kredi borcu" gibi finansal yükümlülükler müşterinin vadeli mevduat aboneliği yapma olasılığını azaltmaktadır. Önerilerden bazıları arasında anlamlı değişkenlere (ör. meslek, eğitim seviyesi, medeni durum) odaklanarak stratejiler geliştirilmesinin yanı sıra eğitim seviyesi yüksek bireyleri hedefleyen pazarlama stratejileri geliştirilebilir. Finansal durumu zorlayıcı olan bireylerin hedef olay üzerinde olumsuz etkileri olduğundan, bu gruplara yönelik uygun öneriler sunulabilir. Çalışmanın sınırlamaları arasında veri setinin güncelliği ve sınıf dengesizliğini gidermek için kullanılan aşırı örnekleme tekniği yer almaktadır. Gelecekte yapılacak araştırmalarda farklı makine öğrenimi modelleri ve daha geniş veri setleri kullanılarak kapsamlı analizler gerçekleştirilebilir.

Anahtar Kelimeler : Bankacılık Analitiği, CRISP-DM Metodolojisi, Veri Analizi, Vadeli Mevduat Aboneliği

Understanding Customer Data in Banking: CRISP-DM Approach and Machine Learning Applications

ABSTRACT

This study highlights the importance of customer information for businesses today and addresses how banks, especially those engaged in financial transactions, process and analyse this information and use it in their management strategies and proposesthe CRISP-DM

Araştırma Makalesi / Research Article
Date Received: 21.11.2024
Publishing Date: 28.02.2025

(Cross-Industry Standard Process for Data Mining) approachto data analysis. Using the anonymised and purposefully revised dataset provided by Moro et al., the factors affecting bank customers' likelihood to subscribe to time deposits are examined and data-driven recommendations for marketing strategies are presented. The study applied the CRISP-DM methodology in six stages, using various machine learning techniques such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machines (SVM) and XGBoost to identify variables affecting customer behaviour. The Logistic Regression model performed the best. The findings show that 'variables such as retired, student, higher education level, previous campaign success, and month of March' indicate that the customer is more likely to subscribe to time deposit. Financial liabilities such as 'mortgage and loan debt'. The customer is less likely to subscribe to time deposits. Some of the suggestions include developing strategies by focusing on meaningful variables (e.g. occupation, education level, marital status) as well as developing marketing strategies targeting individuals with higher education levels. Since individuals with challenging financial situations have negative effects on the target event, appropriate recommendations can be made for these groups. Limitations of the study include the timeliness of the dataset and the oversampling technique used to overcome the class imbalance. In future research, comprehensive analyses can be performed using different machine learning models and larger data sets.

Keywords : Banking Analytics, CRISP-DM Methodology, Data Analysis, Term Deposit Subscription

GİRİŞ

Günümüzde müşteri bilgilerinden yararlanmak, işletmeler için oldukça önemli konulardan biridir. Özellikle finansal işlemlerle uğraşan bankalar söz konusu olduğunda, bu bilgilerin işlenmesi, analiz edilmesi ve çeşitli sonuçlara ulaşılması, yönetim kararları ve stratejilerinin önemli bir parçası haline gelmektedir. Bankaların müşterilerinin demografik verilerinden finansal alışkanlıklarını ve harcama eğilimlerinden kredi geçmişlerine kadar geniş bir veri kaynağına sahiptir. Bu verilerin doğru bir şekilde işlenmesi, analiz edilmesi ve anlamlı sonuçlara dönüştürülmesi, hem operasyonel verimliliği artırmak hem de stratejik karar alma süreçlerini desteklemek açısından kritik önem taşır (Moro et al., 2014). Bu verilerin analiziyle bankalar, müşterilerini daha iyi tanıma, kişiselleştirilmiş ürün yada hizmet sunma imkanı kazanır (Fawcett & Provost, 2013). Dahası bankalar yeni ürünlerin, kampanyalar ve operasyonel amaçlarını bu verilerin analiz sonuçlarına göre şekillendirebilir (Shmueli et al., 2017). Bu verilerin analizi için çeşitli yöntemlerden kullanılır. CRISP-DM (Cross-Industry Standard Process for Data Mining) yaklaşımı bu yöntemlerden birisidir. Analiz süreçlerinin daha sistematik bir şekilde yürütülmesi için kullanılan CRISP-DM yaklaşımı, bankaların veri işleme ve analiz süreçlerinde etkin bir rehberlik sunar. Veri analizi aşamalarında boşluk bırakmayan bu yaklaşım adım adım sonuçlara ulaşılmasını kolaylaştırır. Böylece müşteri bilgilerinin etkin bir şekilde kullanılmasıyla veri odaklı kararların alınmasının yolu açılır. Bu sistematik çerçeve bankaların yalnızca kısa vadeli operasyonel hedeflerine ulaşmasını değil, aynı zamanda uzun vadede müşteri ilişkilerini güçlendirmesine imkân verir. Bu çalışma, bu

yaklaşımı benimseyerek örnek bir veri seti üzerinden makine öğrenimi uygulamalarını kullanarak yapılmıştır. Çalışma örneği olan bankanın pazarlama departmanına veri odaklı stratejik öneriler sunmayı amaçlayan bu çalışma, veri seti olarak Moro ve arkadaşları (Moro et al., 2014) tarafından sağlanan, UCI web sitesinde bulunan anonimleştirilmiş verilerini kullanmaktadır. Amaçlara yönelik yapılan analizlerle müşterinin vadeli mevduat aboneliği yapma olasılığını etkileyen faktörleri anlamak ve bankanın pazarlama stratejilerine veri destekli öneriler sunmaktır. Bu bağlamda, veri setindeki ilişkileri keşfetmek ve müşteri davranışını etkileyen değişkenleri belirlemek ana hedef olarak belirlenmiş aynı zamanda farklı makine öğrenmesi yöntemlerini kullanarak bu yöntemlerin performansını karşılaştırılması yöntemsel katkı sağlaması beklenmektedir. Çalışmanın ana sorunsalı, "*Hangi değişkenler müşterinin vadeli mevduat aboneliği yapma olasılığını etkileyebilir?*" sorusudur. Bu sorunun cevabı için R programlama dili kütüphaneleri kullanılarak analiz edilmiştir. Analizde kullanılan veri seti, banka müşterilerine ait çeşitli demografik ve finansal bilgileri içermektedir. Bu değişkenler arasında yaş, meslek, medeni durum, eğitim seviyesi, kredi durumu, bakiye, konut kredisi ve kişisel kredi durumu gibi özellikler bulunmaktadır. Hangi değişkenlerin analize dahil edileceğine ise test sonuçlarına göre karar verilmektedir. Ayrıca, müşteriye son kampanya kapsamında yapılan iletişim türü, kampanya ayı, haftanın günü ve kampanya süresi gibi faktörler de dikkate alınmıştır. Dahası çalışma müşterinin vadeli mevduat aboneliği yapma olasılığını etkileyen faktörleri belirlemek amacıyla çeşitli analiz teknikleri kullanılacaktır. Çalışmanın müşteri davranışını anlama konusunda veriye dayalı bilgiler elde edilerek, bankalara stratejik rehberlik sunması beklenmektedir.

Çalışmada, ilk olarak bankacılıkta veri madenciliği, makine öğrenimi ve uygulamaları, CRISP-DM metodolojisi hakkında literatür taraması yapılmış, daha sonra CRISP-DM aşamalarını içeren analizler yapılarak en son bulgular değerlendirilerek öneriler yapılmıştır.

1. LİTERATÜR TARAMASI

Bankacılık ve Veri Madenciliği: Bankacılık sektöründe veri madenciliği tekniklerinin kullanımı giderek artan bir eğilimdir. Bankacılık işlemleri sonucunda bankaların büyük miktarda veriye sahip olması ve bu verilerden anlamlı sonuçlar çıkarması, müşteri davranışlarını anlamada ve pazarlama stratejileri geliştirmede oldukça önemli konulardan birisidir. Bankaların pazarlama kampanyalarında veri madenciliği kullanarak müşteri eğilimlerini ve vadeli mevduat aboneliği gibi önemli parametreleri etkileyen faktörleri belirlemek için çalışmalar yapılmaktadır (Zaki et al., 2024; Zhuang et al., 2018). Bu tür veri odaklı çalışmalar, müşteri davranışlarını analiz etmede etkili sonuçlar vermektedir (Sandhya Kona, 2020). Veri madenciliği tekniklerinin bankacılıkta kullanımı, müşteri davranışlarını anlamakla sınırlı kalmamakta; aynı zamanda kredi risk değerlendirmesi ve dolandırıcılık tespiti gibi kritik işlevleri de kapsamaktadır. Shi ve arkadaşları (Shi et al., 2022), makine öğrenimi algoritmaları kullanarak kredi riski değerlendirmesinde veri madenciliğinin etkinliğini incelemiş ve bu algoritmaların kredi riski tespiti konusunda oldukça başarılı olduğunu göstermiştir. Aynı şekilde, dolandırıcılık tespiti için de veri madenciliği

10.5824/ajite.2025.01.004.x

algoritmaları, anormalliklerin belirlenmesinde önemli bir rol oynamaktadır (Noriega et al., 2023). Ayrıca son dönemlerde tüm alanlanlarda etkisini artıran üretken yapay zeka modellerinin,bankacılık çalışmalarınada yeni bir ivme kazandırmıştır. Kumar ve arkadaşları (2024), üretken yapay zekanın finansal hizmetlerde müşteri analatiğini ve risk yöntemini üzerine etkisini incelemiştir. (Kumar et al., 2024). Başka bir çalışmada veri gizliliği ve güvenliği konuları, makine öğrenimi perspektifinden ele alınmış, müşteri verilerinin etik kullanımı ve düzenleyici çerçeveler üzerinde durulmuştur (Wang et al., 2024). Diğer taraftan gerçek zamanlı müşteri verilerinin pazarlama stratejilerinde uyarlanabilirliğini artırmadaki rolü de araştırmalarda önemli bir yer tutmuştur (Balamurugan, 2024). Bir diğer çalışmada bankacılıkta açıklanabilir yapay zekanın (XAI) uygulanabilirliği incelenmiş, bu modellerin müşteri davranış tahmini ve kredi risk analizi gibi alanlarda şeffaflığı artıran teknikler sunduğu belirtilmiştir (Rane et al., 2023). Ayrıca, müşteri kaybı tahmininde sınıf dengesizliği sorununa hibrit aşırı örnekleme yöntemleriyle çözümler sunulmuştur (Haddadi et al., 2024). Bu tür bankacılık uygulamalarında veri madenciliği tekniklerinin kullanımı, aynı zamanda müşteri segmentasyonu ve hedeflenmiş pazarlama stratejileri geliştirmede de etkili olmuştur (Smeureanu et al., 2013). Müşteri segmentasyonu, bankaların müşteri tabanını daha iyi anlamalarına ve bu sayede her bir segmente özel ürün ve hizmet teklifleri sunmalarına olanak sağlamaktadır (Osei et al., 2021). Makine öğrenimi teknikleri sayesinde bankalar, müşteri memnuniyetini arttırmak ve yeni müşteriler kazanmak için daha etkili stratejiler oluşturabilecektir.

Makine Öğrenimi Teknikleri ve Bankacılık Uygulamaları: Makine öğrenimi teknikleri, bankacılık sektöründe giderek daha fazla kullanılmaya başlamış ve bankaların çeşitli işlevlerini optimize etmelerinde önemli bir rol oynamıştır (Mahalakshmi et al., 2022). Bu teknikler, kredi risk değerlendirmesi, dolandırıcılık tespiti, müşteri segmentasyonu ve hedeflenmiş pazarlama kampanyaları gibi alanlarda yaygın olarak kullanılmaktadır (Nayak, 2024). Makine öğrenimi algoritmalarının kullanımı, veri setlerindeki karmaşık ilişkileri ortaya çıkarmakta ve bankaların stratejik karar alma süreçlerine katkı sağlamaktadır (Marqués et al., 2013). Karar Ağaçları (Decision Tree), Rastgele Orman (Random Forest), Destek Vektör Makineleri (SVM) ve XGBoost algoritmaları bankacılıkta sıkça kullanılan makine öğrenimi teknikleri arasındadır. Özellikle kredi riski değerlendirmesinde ve müşteri davranışının tahmin edilmesinde bu tekniklerin etkinliği kanıtlanmıştır (Bhatore et al., 2020; Shi et al., 2022).

Araştırma yöntemlerinin gelişmesiyle yeni çalışmalar, derin öğrenme tekniklerinin de bankacılıkta kullanımını incelemeye başlamıştır. Derin öğrenme (Goodfellow, 2016), karmaşık veri yapılarındaki özellikleri otomatik olarak öğrenebilmesi nedeniyle kredi riski, dolandırıcılık tespiti ve müşteri segmentasyonu gibi alanlarda etkili olmuştur (Mienye & Jere, 2024). Derin sinir ağları, büyük veri setleri ile çalışırken ve veri içindeki gizli ilişkileri ortaya çıkarmada oldukça başarılıdır (Samek et al., 2021). (Xiang et al., 2023), derin öğrenme

tekniklerinin dolandırıcılık tespiti için geleneksel makine öğrenimi tekniklerinden daha iyi performans gösterdiğini belirtmişlerdir.

Makine öğrenimi tekniklerinin bankacılık uygulamalarındaki diğer bir önemli kullanım alanı da müşteri deneyimini iyileştirmeye yöneliktir (Abdulsalam & Tajudeen, 2024). Doğrusal olmayan makine öğrenimi algoritmaları, müşteri geri bildirimlerini analiz ederek bankaların müşteri memnuniyetini artırarak riskleri azaltacak stratejiler geliştirmelerine yardımcı olabilmektedir (Leo et al., 2019). Bu teknikler, müşteri geri bildirimlerinden elde edilen verilerle bankacılık hizmetlerini kişileştirerek, daha özelleşmiş hizmet sunulmasını sağlamaktadır. Bu teknikler, veriler arasındaki karmaşık ilişkileri daha iyi öğrenebilmekte ve bu sayede daha hassas tahminler yapabilmektedir (Hoang & Wiegratz, 2023). Bu da bankaların kredi riskini daha iyi değerlendirmelerine ve dolandırıcılık aktivitelerini daha etkin bir şekilde tespit etmelerine yardımcı olmaktadır.

Bankacılık Müşteri Davranışı, Pazarlama Stratejileri ve Makine Öğrenmesi

Müşteri davranışını anlamak ve bu davranışlara uygun pazarlama stratejileri geliştirmek, bankacılık sektöründe önemli bir rekabet avantajı sağladığı bilinmektedir. Müşteri verilerinin analiz edilmesi, bankaların müşterilerine kişileşmiş teklifler sunmaları ve müşteri memnuniyetini arttırmaları açısından kritik bir rol oynamaktadır (Kotler et al., 2016). Makine öğrenimi teknikleri, bu analizlerin etkin bir şekilde yapılmasına olanak tanımakta ve pazarlama stratejilerinin daha hedefli ve etkili olmasını sağlamaktadır (Ho et al., 2022).

Bankacılık müşteri segmentasyonu, bankaların müşteri tabanını daha iyi anlamalarına ve her bir segmente özel stratejiler geliştirmelerine yardımcı olan bir makine öğrenimi uygulamasıdır (Bach et al., 2013). Özellikle, kümeleme algoritmaları kullanılarak benzer özelliklere sahip müşteriler gruplandırılmakta ve bu gruplara uygun pazarlama kampanyaları düzenlenmektedir (Duarte et al., 2022). Bu sayede, bankalar müşteri memnuniyetini arttırarak, müşteri sadakatini sağlamakta ve daha etkili pazarlama stratejileri oluşturabilmektedir. Müşteri davranışının tahmin edilmesinde kullanılan makine öğrenimi algoritmaları, Karar Ağaçları (Decision Tree), Rastgele Orman (Random Forest), Destek Vektör Makineleri (SVM) ve XGBoost gibi algoritmalardır. Bu algoritmalar, müşteri profiline göre belirli bir ürün veya hizmete yönelik ilgiyi tahmin etmekte etkili bir şekilde kullanılmaktadır. Bankalar bu teknikleri kullanarak müşterilerinin hangi finansal ürünlere daha fazla ilgi gösterdiğini belirleyebilir ve bu bilgiler doğrultusunda kişileşmiş pazarlama stratejileri geliştirilebilir (Probesto, 2024). Ayrıca Makine öğrenmesi müşteri kaybını (churn) önceden tahmin etmede de önemli bir rol oynamaktadır (Lalwani et al., 2022). Müşteri kaybını önceden belirlemek, bankaların kayıp yaşayabilecek müşterilere yönelik koruma stratejileri geliştirmelerine olanak tanır. Destek Vektör Makineleri (SVM), Lojistik Regresyon ve Derin Öğrenme teknikleri, müşteri kaybını önceden tahmin etmek için yaygın olarak kullanılmaktadır (Agarwal et al., 2022). Bu tahminlere belirli metodoloji ile yaklaşmak için CRISP-DM kullanılması, mevcut bankacılık verilerinin işlenmesi ve analiz edilmesinde sistematik bir yaklaşım getirmektedir.

10.5824/ajite.2025.01.004.x

CRISP-DM Metodolojisi ve Bankacılık

CRISP-DM metodolojisi, veri madenciliği projelerinde yaygın olarak kullanılan (Marbn et al., 2009) ve bankacılık sektöründe de etkili bir şekilde uygulanabilecek bir yaklaşımdır. Bu metodoloji, veri madenciliği projelerinin sistematik bir şekilde planlanması ve yönetilmesini sağlar (Eckerson et al., 2000). CRISP-DM metodolojisi, özellikle bankacılık gibi büyük miktarda veriyle çalışan endüstrilerde, veri madenciliği süreçlerinin daha verimli yürütülmesine yardımcı olur (Plotnikova et al., 2022). Bu çalışmada, bankaların müşteri davranışlarını daha iyi anlamaları ve doğrudan pazarlama kampanyalarının etkinliğini artırmaları konusunda bu yaklaşım yol gösterici olmuştur. Bu metodolojinin kullanılması, veri madenciliği projelerinin her aşamasında sistematik bir yaklaşım benimsenmesini sağlamakta ve bu da projelerin başarı oranını artırtıracağı beklenmektedir.

CRISP-DM altı temel aşamada gerçekleştirilen bir süreci kapsar. Bu aşamalar; iş anlayışı, veri anlayışı, veri hazırlığı, modelleme, değerlendirme ve yayılma şeklindedir. Çalışmanın veri setini analiz etmede yöntem olarak seçilmiştir.

2. METOD VE ANALİZLER

Araştırmada bankanın verileri, CRISP-DM yöntemi kullanılarak analiz edilmiştir. CRISP-DM yöntemi, veri madenciliği ve veri analizi projeleri için yaygın olarak kabul edilen endüstri standartlardan birisidir (Martínez-Plumed et al., 2021). Bu metodoloji, veri madenciliği projelerini planlama, uygulama ve değerlendirme aşamalarına ayırarak, projelerin sistematik ve etkili bir şekilde yönetilmesine katkı sağlar.

CRISP-DM metodolojisi genellikle aşağıdaki altı ana aşamayı içerir:

1. Problemin Tanımlanması (Business Understanding)
2. Verinin Anlaşılması (Data Understanding)
3. Veri Ön Hazırlığı (Data Preparation)
4. Modelleme (Modeling)
5. Değerlendirme (Evaluation)
6. Yayılma (Deployment)

Her bir aşama, analizlerin başarıyla tamamlanabilmesi için kritik öneme sahiptir. Analizlerde CRISP-DM metodolojisinin kullanılması, araştırmaya sistematik bir çerçeve sunmuş ve analizin her aşamasının gereksinimlere uygun olarak yürütülmesini sağlamıştır.

2.1. Problemin Tanımlanması

Portekizli bir bankanın doğrudan pazarlama kampanyalarını konu alan banka pazarlama veri seti amaca uygun düzenlenerek kullanılmıştır. Pazarlama kampanyaları telefon görüşmelerine dayanmaktadır ve ürüne (banka vadeli mevduat) abone olunup olunmayacağını belirlemek amacıyla müşterilerle birden fazla kez iletişim kurulmuştur. Bu bağlamda, problemimiz bir sınıflandırma problemidir. Sınıflandırma problemine uygun algoritmalar seçilmelidir.

2.2.Verilerin Anlaşılması

Mevcut veri setlerini incelemek ve anlamak için R programlama dilindeki kütüphaneler kullanılmıştır. Veri setinin özellikleri Tablo 1'de sunulmuştur. Veri seti, 7 adet sayısal ve 10 adet kategorik değişkenden oluşmaktadır. Veri matrisi, toplamda 45.211 gözlem ve 17 değişkene sahiptir.

Tablo 1: Veri setinin Özellikleri

age	job	marital	education	default	balance	housing	loan	contact
Min. :18.00	Length:45211	Length:45211	Length:45211	Length:45211	Min. : -8019	Length:45211	Length:45211	Length:45211
1st Qu.:33.00	Class :character	Class :charact	Class :character	Class :character	1st Qu.: 72	Class :character	Class :character	Class :character
Median :39.00	Mode :character	Mode :charact	Mode :character	Mode :character	Median : 448	Mode :character	Mode :character	Mode :character
Mean :40.94					Mean : 1362			
3rd Qu.:48.00					3rd Qu.: 1428			
Max. :95.00					Max. :102127			

day	month	duration	campaign	pdays	previous	poutcome	y
Min. : 1.00	Length:45211	Min. : 0.0	Min. : 1.000	Min. : -1.0	Min. : 0.0000	Length:45211	Length:45211
1st Qu.: 8.00	Class :character	1st Qu.: 103.0	1st Qu.: 1.000	1st Qu.: -1.0	1st Qu.: 0.0000	Class :character	Class :character
Median :16.00	Mode :character	Median : 180.0	Median : 2.000	Median : -1.0	Median : 0.0000	Mode :character	Mode :character
Mean :15.81		Mean : 258.2	Mean : 2.764	Mean : 40.2	Mean : 0.5803		
3rd Qu.:21.00		3rd Qu.: 319.0	3rd Qu.: 3.000	3rd Qu.: -1.0	3rd Qu.: 0.0000		
Max. :31.00		Max. :4918.0	Max. :63.000	Max. :871.0	Max. :275.0000		

2.3. Veriler Ön Hazılığı

Bu aşamada, veri setleri temizlenir, dönüştürülür ve modelleme aşamasına uygun hale getirilir. Eksik veriler doldurulur, aykırı değerler ele alınır ve gerekli düzeltmeler yapılır. Veri hazırlığı aşaması, veri madenciliği projelerinin başarısında kritik rol oynar. Bu çalışmada da eksik değerlerin ve aykırı değerlerin kontrol edilmesi, kategorik değişkenlerin dönüştürülmesi gibi adımlar dikkatli bir şekilde uygulanmıştır. Veri hazırlığı sürecinin başarıyla tamamlanması, modelin doğruluğunu ve güvenilirliğini doğrudan etkilemektedir.

2.3.1. Eksik Değerlerin Kontrolü

Veri setindeki eksik değerlerin kontrol edilmesi, veri setinin doğruluğunu ve kullanılabilirliğini değerlendirmek açısından önemlidir. Eksik değerlerin olmaması, analiz ve modelleme süreçlerinin daha sağlıklı ve güvenilir olmasını sağlar. Veri seti kontrol edildiğinde eksik gözlem olmadığı belirlenmiştir. Bu sonuç model eğitimi ve analizler için potansiyel bir problemin olmadığını gösteren olumlu bir durumdur. Bu nedenle, veri setini kullanmaya devam edilip ileri analiz aşamalarına geçilmiştir.

10.5824/ajite.2025.01.004.x

2.3.2. Kategorik Sütunların Geçerlilik Kontrolü

Tüm kategorik sütunlardaki geçerlilik kontrolü, veri setinin temizliği açısından önemlidir. Geçersiz veya tutarsız değerlerin olmaması, model eğitimi ve analiz süreçlerinin güvenilir sonuçlar üretmesine katkı sağlar. Kategorik sütunlardaki eksik değer sayıları kontrol edildi ve eksik değer olmadığı görüldü. Bu kontrol adımlarının sonucunda veri setinin temiz ve model eğitimi için uygun olduğu teyit edilmiştir. Bu durum, sonuçların güvenilirliğini ve doğruluğunu artıracaktır.

2.3.3. Sayısal Değişkenlerin Aykırı Değer Analizi

Sayısal sütunların özet istatistikleri incelendiğinde, "balance" (bakiye), "duration" (süre), "campaign" (kampanya), "pdays" (önceki günler) gibi sütunlarda bazı uç ve aykırı değerlerin olduğu görülmüştür. Bakiye değişkeni için minimum değer -8019, ortalama ise 1428 olarak belirlenmiştir. Bu durum veri setinde olası aykırı değerlere işaret eder. Benzer şekilde, süre için minimum değer 0, ortalama ise 258.2'dir, bu da aykırı değerlerin varlığını gösterir. Aykırı değerler, Çeyrekler Arası Aralık (IQR) yöntemi kullanılarak veri setinden çıkarılmıştır. Bu yöntem, bir değişkenin değer dağılımını ölçerek aykırı değerleri belirler ve bu değerler veri setinden kaldırılmıştır.

2.3.4. Benzersiz Değerlerin Kontrolü ve Gereksiz Sütunların Kaldırılması

Aykırı değerler çıkarıldıktan sonra, tüm sütunlardaki benzersiz değerler kontrol edilmiştir. "duration" sütununun, bir model oluşturmaya katkı sağlamayan tek bir benzersiz değere sahip olduğu belirlenmiştir. Bu nedenle, bu sütun veri kümesinden çıkarılmıştır.

2.3.5. Kategorik ve Hedef Sütunun Dönüştürülmesi

Kategorik sütunları ve hedef sütunu karakter değişkenlerinden faktör değişkenlerine dönüştürmek, özellikle sınıflandırma modellerinin daha iyi çalışmasına yardımcı olur. Bu nedenle, kategorik sütunlar ve hedef sütun karakter değişkenlerinden faktör değişkenlerine dönüştürülmüştür.

2.3.6. Hedef Sütundaki Dengesiz Sınıflar

Veri setinde hedef sütun "y" olarak adlandırılmış ve iki sınıfı "evet" ve "hayır" şeklindedir. Ancak, "evet" sınıfında 5289 gözlem, "hayır" sınıfında ise 39.922 gözlem bulunmaktadır. Bu durum, dengesiz sınıflandırma problemine işaret eder, çünkü sınıflar arasındaki gözlem sayılarında belirgin bir fark vardır. Sınıf dengesizliği, modelin öğrenme mekanizmasına ve optimizasyon sürecine bağlı olarak bazı modellerde daha ciddi sorunlara yol açabilir. Örneğin, doğruluk (accuracy) gibi genel hata oranlarını minimize etmeye çalışan modeller, çoğunluk sınıfa odaklanarak azınlık sınıfı göz ardı edebilir (He & Garcia, 2009). Karar ağaçları ve destek vektör makineleri (SVM) gibi modeller, sınıf frekanslarındaki büyük

farklar nedeniyle sınıf sınırlarını azınlık sınıf aleyhine eğebilirken, sinir ağları gibi daha karmaşık modeller, azınlık sınıfa ait yeterli örnek olmadığında bu sınıfı öğrenmekte zorlanabilir (Chawla et al., 2002). Ayrıca, dengesiz sınıflar veri dağılımını bozarak modelin genelleme yeteneğini etkileyebilir ve performans metriklerinin yanıltıcı sonuçlar vermesine neden olabilir. Bu nedenle, sınıf dengesizliği yaşayan veri setlerinde, modelin duyarlılığına uygun yöntemler ve metrik odaklı yaklaşımlar tercih edilmelidir. Dengesiz sınıflandırma problemleri modelin eğitilmesi ve değerlendirilmesi üzerinde olumsuz etki yaratabilir. Bu nedenle, bu tür durumlarda bir dengesizlik giderme stratejisi uygulanmıştır. "rastgele aşırı örnekleme (Random oversampling)" tekniği kullanılarak az sayıdaki sınıfın gözlemleri artırılmıştır. Bu işlem, küçük sınıfın örnek sayısını artırarak modelin bu sınıfı daha iyi öğrenmesine ve dengesizlikten kaynaklanan yanlılıkları azaltmasına yardımcı olmaktadır.

2.3.7. Eğitim ve Test Setlerine Bölme

Veri setini eğitim ve test setlerine bölmek, modelin eğitim verileri üzerinde öğrenmesi ve ardından bu öğrenilen bilgileri test verileri üzerinde değerlendirmesi için yaygın bir yöntemdir. Bu işlem, modelin gerçek dünya verileri üzerinde genelleme yeteneğini değerlendirmek amacıyla yapılır. Bu şekilde modelin genelleme performansını değerlendirme, aşırı uyum (overfitting) kontrolü ve performansın güvenilir olarak ölçülmesi sağlanır. Eğitim ve test setleri %80 ve %20 oranında ayrılmıştır; verilerin %80'i model eğitimi için, %20'si ise model test işlemleri için kullanılmıştır. Modelin performansını değerlendirmek için doğruluk (accuracy), kesinlik (precision), duyarlılık (recall), F1 skoru gibi denetimli öğrenme metrikleri kullanılmıştır. Bu metrikler, modelin sınıflandırma yeteneklerini nicel olarak ölçmektedir.

2.4.Modelleme

Bu aşamada, veri seti üzerinde çeşitli modelleme teknikleri uygulanmıştır. Amaç, belirli bir iş sorununa en uygun modeli seçmek ve eğitmektir. Verilerin dağılımı incelendiğinde doğrusal bir dağılıma sahip olmadığı görülmüştür. Veri setinin özellikleri sınıflandırmaya uygun olduğundan Lojistik Regresyon, Karar Ağaçları (Decision Tree), Rastgele Orman (Random Forest), Destek Vektör Makineleri (SVM) ve XGBoost analizleri yapılmıştır.

2.4.1. Lojistik Regresyon

Lojistik regresyon, bağımlı değişkenin kategorik veya ikili (binary) olduğu durumlarda kullanılan bir istatistiksel analiz yöntemidir. Bu yöntem, bir veya daha fazla bağımsız değişkenin bağımlı değişken üzerindeki etkisini incelemek için kullanılır. Bu çalışmada da, "y" değişkeni (*müşterinin vadeli mevduat aboneliği yapma olasılığı*) üzerinde değişkenlerin dağılımı incelenmiş ve sonuçlar Tablo 2'de sunulmuştur. Tablo 2'deki bulgular, bazı değişkenlerin bağımlı değişken üzerinde önemli bir etkisi olduğunu görülmektedir. Bu sonuçlar bir değerlendirme bölümünde Tablo 3'te detaylı olarak verilmiştir.

10.5824/ajite.2025.01.004.x

Modelin AIC (Akaike Bilgi Kriteri) değeri 17342 olarak hesaplanmıştır. Bu değer, modelin performansını değerlendirmek için diğer modelle karşılaştırılacaktır.

Tablo 2: Lojistik Regresyon Sonuçları

Coefficients:	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-270000	0.205000	-13.222	< 2e-16
age	-0.000448	0.002470	-0.182	0.855864
jobblue-collar	-0.232000	0.081100	-2.861	0.004229
jobentrepreneur	-0.315000	0.139000	-2.267	0.023372
jobhousemaid	-0.366000	0.150000	-2.434	0.014953
jobmanagement	-0.130000	0.082300	-1.577	0.114764
jobretired	0.320000	0.109000	2.924	0.003452
jobself-employed	-0.218000	0.124000	-1.763	0.077956
jobservices	-0.176000	0.094400	-1.865	0.062231
jobstudent	0.318000	0.122000	2.601	0.0093
jobtechnician	-0.153000	0.077800	-1.962	0.049796
jobunemployed	-0.127000	0.126000	-1.011	0.312025
jobunknown	-0.256000	0.261000	-0.982	0.326268
maritalmarried	-0.170000	0.065900	-2.581	0.009845
maritalsingle	0.140000	0.075000	1.861	0.062762
educationsecondary	0.241000	0.072400	3.322	0.000893
educationtertiary	0.435000	0.084300	5.158	2.50E-07
educationunknown	0.316000	0.116000	2.721	0.006515
defaultyes	0.077200	0.179000	0.431	0.666312
balance	0.000012	0.000006	2.075	0.03799
housingyes	-0.690000	0.048900	-14.092	< 2e-16
loanyes	-0.403000	0.066500	-6.053	1.42E-09
contacttelephone	-0.085700	0.082900	-1.033	0.301394
contactunknown	-1.620.000	0.081400	-19.927	< 2e-16
day	0.012700	0.002790	Nis.57	4.87E-06
monthaug	-0.762000	0.088000	-8.658	< 2e-16
monthdec	0.670000	0.195000	3.434	0.000594
monthfeb	-0.197000	0.100000	-1.968	0.049011
monthjan	-1.380.000	0.138000	-10	< 2e-16
monthjul	-0.868000	0.085800	-10.111	< 2e-16
monthjun	0.458000	0.104000	4.395	1.11E-05
monthmar	1.530.000	0.134000	11.344	< 2e-16
monthmay	-0.447000	0.080600	-5.54	3.03E-08
monthnov	-0.914000	0.094200	-9.704	< 2e-16
monthoct	0.908000	0.120000	7.585	3.32E-14
monthsep	0.893000	0.132000	Haz.77	1.29E-11

pdays	0.000174	0.000344	0.506	0.613163
previous	0.007450	0.006340	1.177	0.239384
poutcomeother	0.209000	0.102000	2.061	0.039343
poutcomesuccess	2.340.000	0.092500	25.27	< 2e-16
poutcomeunknown	0.001240	0.105000	0.012	0.990605
Signif. codes: 0	'***' 0.001	'**' 0.01	'*' 0.05	'.' 0.1
Null deviance:	26104 on	36167 degr	ees of f	reedom
Residual deviance:	17256 on	36125 degr	ees of f	reedom
AIC: 17342				

2.5. Değerlendirme (Evaluation)

En iyi performans gösteren birisi olan Lojistik Regresyon sonuçları (Tablo 2) değerlendirildiğinde; müşteri davranışlarını anlamak ve finansal ürünlerin benimsenmesine yönelik önemli içgörüler sunmaktadır. pozitif katsayısı olan değişkenler (Tablo 3), Emekli bireyler (jobretired) ve öğrenciler (jobstudent) gibi grupların vadeli mevduat aboneliğine daha yatkın olduğu görülmektedir. Bu bulgu, emekli bireylerin düzenli gelir kaynaklarına ve öğrencilerin geleceğe yönelik finansal planlama ihtiyaçlarına odaklanan literatürle uyumludur (Lusardi & Mitchell, 2014). Yüksek eğitim seviyesinin (educationsecondary ve educationtertiary) abonelik olasılığını artırması da finansal okuryazarlık düzeyinin artışıyla ilişkilendirilebilir (Bilge & Nur, 2023). Önceki kampanya başarısının (poutcomesuccess) müşteri davranışını önemli ölçüde etkilediği sonucuna ulaşılmıştır. Bu bulgu, müşteri ilişkileri yönetiminin ve geçmiş deneyimlerin gelecekteki ürün kabul oranları üzerindeki etkisini gösteren (Verhoef et al., 2003) tarafından desteklenmektedir. Benzer şekilde, mart ayında vadeli mevduat aboneliğine olan yüksek ilgi (monthmar), sezonluk etkiler ve vergi iadeleri gibi finansal davranışlarla açıklanabilir .

Sonuçlardaki negatif etkiler arasında (Tablo 3), mevcut finansal yükümlülüklerin (housingyes ve loanyes) vadeli mevduat aboneliğini azalttığı gözlemlenmiştir. Bu durum, borç ödeme yükümlülüklerinin tasarruf davranışlarını baskılayabileceğini savunan (Campbell & Cocco, 2003)'nun çalışması ile uyumludur. Mavi yaka çalışanların (jobblue-collar) abonelik olasılığının düşük olması, gelir seviyeleri ve harcama öncelikleriyle ilişkilendirilebilir (Çımrın et al., 2023). Temmuz ayında (monthjul) abonelik oranlarının düşüklüğü ise tatil harcamaları ve sezonluk harcama artışları ile ilişkilendirilebilir (Claessens & Laeven, 2004).

Bu lojistik regresyon analizinin bulguları, bankaların müşteri segmentasyonunu iyileştirmesi ve hedefli kampanyalar geliştirmesi için önemli bilgiler sunmaktadır. Yüksek eğitim seviyesine sahip veya önceki kampanyalardan olumlu sonuç alan gruplara odaklanmak, pazarlama stratejilerinin etkinliğini artırabilir (Özdemir, 2021; Demirel, 2024). Ayrıca, (Bumin, 2023) çalışmasına benzer mavi yaka çalışanlar ve finansal yükümlülüğü

yüksek bireyler gibi gruplara yönelik farklı ürünler veya finansal destek politikaları geliştirilmesi önerilmektedir.

Tablo 3: Analiz Sonuçlarının Değerlendirmesi

Değişken	Katsayı	p Değeri	Anlamlılık Seviyesi (***0.001,**0.01,*0.05)	Yorum
jobblue-collar	-0.232	0.004229	**	Mavi yaka işçilerin vadeli mevduat aboneliği yapma olasılığı düşüktür.
jobretired	0.320	0.003452	**	Emekli bireylerin vadeli mevduat aboneliği yapma olasılığı yüksektir.
jobstudent	0.318	0.0093	**	Öğrencilerin vadeli mevduat aboneliği yapma olasılığı yüksektir.
maritalmarried	-0.170	0.009845	**	Evli olmanın vadeli mevduat aboneliği yapma olasılığı üzerinde negatif etkisi vardır.
educationsecondary	0.241	0.000893	***	Ortaokul mezunlarının vadeli mevduat aboneliği yapma olasılığı yüksektir.
educationtertiary	0.435	2.50E-07	***	Üniversite mezunlarının vadeli mevduat aboneliği yapma olasılığı oldukça yüksektir.
housingyes	-0.690	< 2e-16	***	Konut kredisine sahip olmak, vadeli mevduat aboneliği yapma olasılığını önemli ölçüde azaltır.
loanyes	-0.403	1.42E-09	***	Kredi borcuna sahip olmak, vadeli mevduat aboneliği yapma olasılığını düşürmektedir.
contactunknown	-1.620	< 2e-16	***	İletişim türü bilinmiyorsa, vadeli mevduat aboneliği yapma olasılığı oldukça düşüktür.
monthmar	1.530	< 2e-16	***	Mart ayında vadeli mevduat aboneliği yapma olasılığı yüksektir.

monthjul	-0.868	< 2e-16	***	Temmuz ayında vadeli mevduat aboneliği yapma olasılığı düşüktür.
poutcomesuccess	2.340	< 2e-16	***	Önceki kampanyanın başarılı olması, vadeli mevduat aboneliği yapma olasılığını artırır.

2.5.1. Veri Setinin Diğer Makine Öğrenmesi Algoritmalarıyla Karşılaştırması

Veri setinin diğer modeller ile performansı karşılaştırıldığında (Tablo 4), Lojistik Regresyon, Random Forest ve XGBoost modeli için doğruluk (accuracy) değeri %88, Decision Tree ise %85,5 olarak tespit edilmiştir. Her bir modelde de AUC değeri %80'in üzerinde olduğundan, iyi uyumlu modeller olarak kabul edilmiştir. Modeller abonelik tahminlerinde, Class 0'da (abonelik yapmayanlar) oldukça başarılı performans sergilerken, Class 1 (abonelik yapanlar) için F1 skorları düşük kalmaktadır. Abonelik yapanların tekrar hesap açma ihtimalleri düşüktür.

Tablo 4: Model Performansları

Model	Accuracy	F1 Score (Class 0)	F1 Score (Class 1)
Logistic Regression	0.8879796527700984	0.94	0.32
Decision Tree	0.8551082100518141	0.92	0.47
Random Forest	0.8878687080333574	0.94	0.32
SVM	0.8789464863284238	0.93	0.25
XGBoost	0.8862173966065948	0.93	0.37

2.6. Yayılma

Bu aşamada analiz sonuçlarının ne anlama geldiği ve uygulayıcılara öneriler yer alır. Tablo 5'te bulguların sonuçları ve öneriler listelenmiştir. Analize konu olan bankanın bu sonuçları ve önerileri dikkate alarak mevcut durumun daha da ileriye götürebilecek stratejiler geliştirebilir.

Tablo 5: Sonuçların Yorumlanması ve Öneriler

Sonuçlar	Öneriler
Anlamlı değişkenler belirlendi.	Anlamlı değişkenlere (meslek, eğitim seviyesi, medeni durum) odaklanarak stratejiler geliştirilebilir.
Anlamsız değişkenler tespit edildi.	Anlamlı katkı yapmayan değişkenler (yaş, defaultyes) modelden çıkarılabilir veya yeniden gözden geçirilebilir.
Eğitim seviyesi hedef olayı etkiliyor.	Eğitim seviyesi yüksek bireyleri hedefleyen pazarlama stratejileri geliştirilebilir.
Meslek gruplarının etkisi gözlemlendi.	Hedef kitlenin mesleki durumuna göre özelleştirilmiş kampanyalar hazırlanabilir.

10.5824/ajite.2025.01.004.x

Konut ve kredi borcunun negatif etkisi.	Finansal durumu zorlayıcı olan bireylerin hedef olay üzerinde olumsuz etkileri olduğundan, bu gruplara yönelik uygun öneriler sunulabilir.
Önceki kampanyanın başarısı önemli.	Önceki kampanyalarda başarı gösteren grupları belirleyip, bu grupları yeniden hedeflemek faydalı olabilir.

SONUÇ

Bu çalışma, bankacılık verilerini CRISP-DM metodolojisi kullanarak sistemli bir şekilde analiz etmiştir. Moro ve diğerleri tarafından sağlanan, Portekizli bir bankanın pazarlama departmanına veri odaklı stratejik öneriler sunma amacı güden anonimleştirilmiş veri seti, banka müşterilerinin vadeli mevduat aboneliği yapma olasılığını anlamak ve pazarlama stratejilerine veri destekli öneriler sunmak için kullanılmıştır. Veri madenciliği ve makine öğrenimi teorilerinin bir uygulaması olan CRISP-DM metodolojisi yaklaşımıyla da veri setinden anlamlı bilgiler çıkarma süreçlerine odaklanmıştır. Bu yaklaşım, bankaların müşteri verilerini analiz etme ve bu verilerden elde edilen içgörülerlerle pazarlama stratejileri ile entegre etme süreçlerinde, destekleyici olabileceğini göstermektedir. Veri setindeki ilişkileri keşfetmek ve müşteri davranışını etkileyen değişkenleri belirlemek hedeflenmiş; hangi değişkenlerin vadeli mevduat aboneliği yapma olasılığını etkileyebileceği sorusuna yanıt aranmıştır. Yöntem olarak CRISP-DM yaklaşımı kullanılmış ve çeşitli makine öğrenmesi algoritmalarıyla sonuçlar doğrulanmış ve performansları test edilmiştir. Lojistik Regresyon modeli, önemli değişkenlerle yüksek doğruluk sağlamış; diğer modeller de benzer performans sergilemiştir. Araştırma sorusuna yanıt olarak, müşteri bilgileri arasında özellikle yaş, finansal durum, önceki pazarlama kampanyalarına verilen tepkiler gibi değişkenlerin vadeli mevduat aboneliğini anlamlı bir şekilde etkilediği görülmüştür. Elde edilen bulgular, belirli demografik ve finansal değişkenlerin, müşterilerin vadeli mevduat aboneliği yapma olasılıkları üzerinde anlamlı etkileri olduğunu göstermektedir. Özellikle "emekli, öğrenci, yüksek eğitim seviyesi, önceki kampanya başarısı ve mart ayı" gibi değişkenler abonelik olasılığını artırırken; "konut kredisi ve kredi borcu" gibi finansal yükümlülükler bu olasılığı azaltmaktadır. Bulgular, anlamlı değişkenlere (meslek, eğitim seviyesi, medeni durum gibi) odaklanarak stratejiler geliştirilmesi gerektiğini ve eğitim seviyesi yüksek bireyleri hedefleyen pazarlama stratejilerinin daha etkili olabileceğini önermektedir. Ayrıca, finansal durumu zorlayıcı olan bireylerin vadeli mevduat aboneliği üzerindeki olumsuz etkileri göz önünde bulundurularak, bu gruplara yönelik uygun kişiselleştirilmiş öneriler sunulması gerekmektedir. Diğer taraftan her bir makine modeli, başarı kriterlerini karşılamaktadır. Bulgular ise müşteri bilgilerini anlamanın önemini ortaya çıkarmaktadır. Özellikle finansal işlemlerle ilgilenen bankaların bu tür verileri analiz ederek, sonuçları yönetim stratejilerinde kullanmaları önerilebilir. Çalışmanın sınırlamaları arasında veri setinin güncelliği ve sınıf dengesizliğini gidermek için kullanılan aşırı örnekleme tekniği bulunmaktadır. Bu sınırlamaların etkisi, sonuçların genellenebilirliğini sınırlayabilir. Gelecekteki araştırmalarda farklı makine öğrenimi modelleri

(AutoML (Otomatik Makine Öğrenimi), Öngörücü Yapay Zeka Modelleri gibi) ve daha geniş veri setleri kullanılarak daha kapsamlı analizler yapılabilir. Bununla birlikte, teorik katkıların artırılması amacıyla müşteri davranışları ile bankacılık stratejileri arasındaki ilişkiyi açıklayan daha derinlemesine bir yaklaşım oluşturulması gerekebilir. Gelecekteki araştırmalar, farklı sektörlerdeki veri analizleri ve daha karmaşık makine öğrenimi teknikleri ile bu yaklaşımın geliştirilmesini sağlayabilir.

KAYNAKÇA

Abdulsalam, T. A., & Tajudeen, R. B. (2024). Artificial intelligence (AI) in the banking industry: A review of service areas and customer service journeys in emerging economies. *Business & Management Compass*, 68(3), 19–43. https://doi.org/10.56065/9hfvrq20

Agarwal, V., Taware, S., Yadav, S., Gangodkar, D., Rao, A., & Srivastav, V. (2022). Customer—churn prediction using machine learning. *Proceedings of ICTACS*, 899. https://doi.org/10.1109/ICTACS56270.2022.9988187

Bach, M. P., Juković, S., Dumičić, K., & Šarlija, N. (2013). Business client segmentation in banking using self-organizing maps. *South East European Journal of Economics and Business*, 8(2), 32–41. https://doi.org/10.2478/jeb-2013-0007

Balamurugan, M. (2024). AI-driven adaptive content marketing: Automating strategy adjustments for enhanced consumer engagement. *International Journal For Multidisciplinary Research*, 6. https://doi.org/10.36948/ijfmr.2024.v06i05.27940

Bhatore, S., Mohan, L., & Reddy, Y. R. (2020). Machine learning techniques for credit risk evaluation: A systematic literature review. *Journal of Banking and Financial Technology*, 4(1), 111–138. https://doi.org/10.1007/s42786-020-00020-3

Bilge, C., & Nur, T. (2023). Finansal performansı etkileyen içsel faktörler: Mevduat ve katılım bankaları üzerine bir uygulama. *Ardahan Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 5(2), Article 2. https://doi.org/10.58588/aru-jfeas.1292408

Bumin, M. (2023). Türk bankacılık sektöründe mevduat bankalarının karlılık performansını etkileyen faktörlerin panel regresyon analizi ile belirlenmesi. *Muhasebe ve Finansman Dergisi*, 100, Article 100. https://doi.org/10.25095/mufad.1326939

Campbell, J. Y., & Cocco, J. F. (2003). Household risk management and optimal mortgage choice. *The Quarterly Journal of Economics*, 118(4), 1449–1494.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357. https://doi.org/10.1613/jair.953

Çımrın, A. H., Kaya, İ., & Bahadır, H. (2023). Türkiye'de mavi yakalı çalışan olma hali üzerine bir analiz. *Sosyoloji Dergisi*, 45, Article 45. https://doi.org/10.59572/sosder.1225697

Claessens, S., & Laeven, L. (2004). What drives bank competition? Some international evidence. *Journal of Money, Credit and Banking*, 36(3), 563–583.

Demirel, S. (2024). Bankacılıkta dijitalleşmenin müşteri davranışları üzerine etkisi. Retrieved from https://www.gazikitabevi.com.tr/bankacilikta-dijitallesmenin-musteri-davranislari-uzerine-etkisi

Duarte, V., Zuniga-Jara, S., & Contreras, S. (2022). Machine learning and marketing: A systematic literature review. *IEEE Access*, 10, 93273–93288. https://doi.org/10.1109/ACCESS.2022.3202896

Eckerson, W. W., Hanlon, N., & Barquin, R. (2000). Director of education and research. *5*(4).

Fawcett, T., & Provost, F. (2013). *Data science for business*.

Goodfellow, I. (2016). *Deep learning*. MIT Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=omivDQAAQBAJ&oi=fnd&pg=PR5&dq=Goodfellow,+I.,+Bengio,+Y.,+%26+Courville,+A.+(2016).+Deep+Learning.+MIT+Press.&ots=MOO5bmsHTY&sig=43Ua_77nITneAJaXBCXMRPWQYbU

Haddadi, S. J., Farshidvard, A., Silva, F. dos S., dos Reis, J. C., & da Silva Reis, M. (2024). Customer churn prediction in imbalanced datasets with resampling methods: A comparative study. *Expert Systems with Applications*, 246, 123086. https://doi.org/10.1016/j.eswa.2023.123086

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239

Ho, S. C., Wong, K. C., Yau, Y. K., & Yip, C. K. (2022). A machine learning approach for predicting bank customer behavior in the banking industry. In *Research Anthology on Machine Learning Techniques, Methods, and Applications* (pp. 1210–1232). IGI Global. https://doi.org/10.4018/978-1-6684-6291-1.ch063

Hoang, D., & Wiegratz, K. (2023). Machine learning methods in finance: Recent applications and prospects. *European Financial Management*, 29(5), 1657–1701. https://doi.org/10.1111/eufm.12408

Kotler, P., Keller, K. L., Brady, M., Goodman, M., & Hansen, T. (2016). *Marketing management* (3rd ed.). Pearson Higher Ed.

Kumar, K., Kuhar, N., & Sharma, M. (2024). Artificial intelligence in the Indian banking system: A systematic literature review (SSRN Scholarly Paper No. 5088937). *Social Science Research Network*. https://doi.org/10.2139/ssrn.5088937

Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2022). Customer churn prediction system: A machine learning approach. *Computing*, 104(2), 271–294. https://doi.org/10.1007/s00607-021-00908-y

Leo, M., Sharma, S., & Maddulety, K. (2019). Machine learning in banking risk management: A literature review. *Risks*, 7(1), 29. https://doi.org/10.3390/risks7010029

Lusardi, A., & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature*, 52(1), 5–44. https://doi.org/10.1257/jel.52.1.5

Mahalakshmi, V., Kulkarni, N., Pradeep Kumar, K. V., Suresh Kumar, K., Nidhi Sree, D., & Durga, S. (2022). The role of implementing artificial intelligence and machine learning technologies in the financial services industry for creating competitive intelligence. *Materials Today: Proceedings*, 56, 2252–2255. https://doi.org/10.1016/j.matpr.2021.11.577

Marbn, S., Mariscal, G., & Segovi, J. (2009). A data mining & knowledge discovery process model. In J. Ponce & A. Karahoc (Eds.), *Data mining and knowledge discovery in real life applications*. I-Tech Education and Publishing. https://doi.org/10.5772/6438

Marqués, A. I., García, V., & Sánchez, J. S. (2013). A literature review on the application of evolutionary computing to credit scoring. *Journal of the Operational Research Society*, 64(9), 1384–1399. https://doi.org/10.1057/jors.2012.145

Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernández-Orallo, J., Kull, M., Lachiche, N., Ramírez-Quintana, M. J., & Flach, P. (2021). CRISP-DM twenty years later: From data mining processes to data science trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 33(8), 3048–3061. https://doi.org/10.1109/TKDE.2019.2962680

Mienye, I. D., & Jere, N. (2024). Deep learning for credit card fraud detection: A review of algorithms, challenges, and solutions. *IEEE Access*, 12, 96893–96910. https://doi.org/10.1109/ACCESS.2024.3426955

Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, 62, 22–31. https://doi.org/10.1016/j.dss.2014.03.001

Nayak, S. (2024). Enhancing credit risk assessment using machine learning: A case study on early payment risk prediction.

Noriega, J. P., Rivera, L. A., & Herrera, J. A. (2023). Machine learning for credit risk prediction: A systematic literature review. *Data*, 8(11), 169. https://doi.org/10.3390/data8110169

Osei, F., Ampomah, G., Kankam-Kwarteng, C., Opoku Bediako, D., & Mensah, R. (2021). Customer satisfaction analysis of banks: The role of market segmentation. *Science Journal of Business and Management*, 9(2), 126. https://doi.org/10.11648/j.sjbm.20210902.19

Özdemir, G. A. (2021). Dijital bankacılıkta müşteri deneyiminin öncüllerinin ve ardıllarının analizi [Ph.D.]. Retrieved from https://www.proquest.com/docview/2637686536/abstract/C61D402B481E4562PQ/1

Plotnikova, V., Dumas, M., & Milani, F. P. (2022). Applying the CRISP-DM data mining process in the financial services industry: Elicitation of adaptation requirements. *Data & Knowledge Engineering*, 139, 102013. https://doi.org/10.1016/j.datak.2022.102013

Probesto. (2024). Etkili müşteri segmentasyonu: Yapay zekanın gücünü açığa çıkarma—Probesto. Retrieved from https://www.probesto.com/tr/etkili-musteri-segmentasyonu-yapay-zekanin-gucunu-aciga-cikarma/

Rane, N., Choudhary, S., & Rane, J. (2023). Explainable artificial intelligence (XAI) approaches for transparency and accountability in financial decision-making (SSRN Scholarly Paper No. 4640316). *Social Science Research Network*. https://doi.org/10.2139/ssrn.4640316

Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K.-R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, 109(3), 247–278. https://doi.org/10.1109/JPROC.2021.3060483

Sandhya Kona, S. (2020). Customer segmentation and personalization in banking services: Investigating the use of big data analytics to segment banking customers based on their behavior, demographics, and preferences, and leveraging these insights to personalize banking services and marketing campaigns. *International Journal of Science and Research (IJSR)*, 9(8), 1566–1570. https://doi.org/10.21275/SR24522131706

Shi, S., Tse, R., Luo, W., D'Addona, S., & Pau, G. (2022). Machine learning-driven credit risk: A systemic review. *Neural Computing and Applications*, 34(17), 14327–14339. https://doi.org/10.1007/s00521-022-07472-2

Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., & Jr, K. C. L. (2017). *Data mining for business analytics: Concepts, techniques, and applications in R*. John Wiley & Sons.

Smeureanu, I., Ruxanda, G., & Badea, L. M. (2013). Customer segmentation in private banking sector using machine learning techniques. *Journal of Business Economics and Management*, 14(5), 923–939. https://doi.org/10.3846/16111699.2012.749807

Verhoef, P., Donkers, B., Langerak, F., Leeflang, P., & Lemon, L. (2003). Understanding the effect of customer relationship management efforts on customer retention and customer share development. *Journal of Marketing*, 67, 30–45. https://doi.org/10.1509/jmkg.67.4.30.18685

Wang, S., Asif, M., Shahzad, M. F., & Ashfaq, M. (2024). Data privacy and cybersecurity challenges in the digital transformation of the banking sector. *Computers & Security*, 147, 104051. https://doi.org/10.1016/j.cose.2024.104051

Xiang, S., Zhu, M., Cheng, D., Li, E., Zhao, R., Ouyang, Y., Chen, L., & Zheng, Y. (2023). Semi-supervised credit card fraud detection via attribute-driven graph representation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(12), 14557–14565. https://doi.org/10.1609/aaai.v37i12.26702

Zaki, A. M., Khodadadi, N., Lim, W. H., & Towfek, S. K. (2024). Predictive analytics and machine learning in direct marketing for anticipating bank term deposit subscriptions. *American Journal of Business and Operations Research*, 11(1), 79–88. https://doi.org/10.54216/AJBOR.110110

Zhuang, Q. R., Yao, Y. W., & Liu, O. (2018). Application of data mining in term deposit marketing. *Hong Kong*.

ABA
Akademik Bilişim Araştırmaları Derneği

Suadiye Mah. Kazım Özalp Sok. No:15 Kat:2
Şaşkınbakkal Kadıköy/İSTANBUL
Tel: 0216 355 56 19 • Fax: 0216 368 43 30
www.abilar.org