E-ISSN: 3357-2980

JOURNAL OF DATA APPLICATIONS

ISSUE **3** • YEAR **2024**

İSTANBUL UNIVERSITY PRESS

Journal of Data Applications
Issue: 3, 2024
E-ISSN: 2980-3357

Owner
Assoc. Prof. Zeki ÖZEN
İstanbul University, Faculty of Economics,
Department of Management Information Systems, İstanbul, Türkiye

Responsible Manager
Assoc. Prof. Zeki ÖZEN
İstanbul University, Faculty of Economics,
Department of Management Information Systems, İstanbul, Türkiye

Correspondence Address
Prof. Dr. Ümit Yaşar Doğanay Cad.,
İktisat Fakültesi Ek Bina-2 No: 6 Kat:1 Vezneciler / Fatih, İstanbul, Türkiye
E-mail: joda@istanbul.edu.tr
https://iupress.istanbul.edu.tr/en/journal/joda/home

Publisher
İstanbul Üniversitesi Yayınevi / İstanbul University Press
İstanbul Üniversitesi Merkez Kampüsü, 34452 Beyazıt,
Fatih / İstanbul, Türkiye
Telefon / Phone: +90 (212) 440 00 00

Authors bear responsibility for the content of their published.

The publication language of the journal is English.

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

Yayın Türü / *Publication Type*: Yaygın Süreli / *Periodical*

Journal of Data Applications
Issue: 3, 2024
E-ISSN: 2980-3357

EDITORIAL MANAGEMENT BOARD

Editor-in-Chief
Assoc. Prof. Zeki ÖZEN, İstanbul University, Faculty of Economics, Department of Management Information Systems, İstanbul, Türkiye - *zekiozen@Istanbul.edu.tr*

Co-Editors in Chief
Assoc. Prof. Elif KARTAL, İstanbul University, Faculty of Economics, Department of Management Information Systems, İstanbul, Türkiye - *elifk@Istanbul.edu.tr*

Editorial Management Board Member
Assoc. Prof. Gökhan ÖVENÇ, İstanbul University, Faculty of Economics, Department of Management Information Systems, İstanbul, Türkiye - *gokhanovenc@Istanbul.edu.tr*
Assoc. Prof. Emre AKADAL, İstanbul University, Faculty of Economics, Department of Management Information Systems, İstanbul, Türkiye - *emre.akadal@Istanbul.edu.tr*
Assoc. Prof. Elif KARTAL, İstanbul University, Faculty of Economics, Department of Management Information Systems, İstanbul, Türkiye - *elifk@Istanbul.edu.tr*

Ethics Editor
Assoc. Prof. Dr. Sezer Köse BİBER, İstanbul University-Cerrahpaşa, İstanbul, Türkiye

Language Editors
Rachel Elana KRISS, İstanbul University, İstanbul, Türkiye - *rachel.kriss@Istanbul.edu.tr*
Elizabeth Mary EARL, İstanbul University, İstanbul, Türkiye - *elizabeth.earl@Istanbul.edu.tr*

Editorial Assistant
Res. Assist. İdris ALKIŞ, İstanbul University, İstanbul, Türkiye - *idris.alkis@istanbul.edu.tr*

Publicity Manager
Esma ÖZPALTACI, İstanbul Üniversitesi, İstanbul, Türkiye

EDITORIAL BOARD

Prof. Çiğdem Arıcıgil ÇİLAN, İstanbul University, İstanbul, Türkiye - *ccilan@Istanbul.edu.tr*
Prof. Gökhan SİLAHTAROĞLU, İstanbul Medipol University, İstanbul, Türkiye - *gsilahtaroglu@medipol.edu.tr*
Prof. Hasan VERGİL, İstanbul University, İstanbul, Türkiye - *hasan.vergil@Istanbul.edu.tr*
Prof. Mehmet Erdal BALABAN, Sabancı University, İstanbul, Türkiye - *erdal.balaban@sabanciuniv.edu*
Prof. Mehmet Hakan SATMAN, İstanbul University, İstanbul, Türkiye - *mhsatman@Istanbul.edu.tr*
Prof. Mehmet TÜREGÜN, Barry University, Florida, United-States - *mturegun@barry.edu*
Prof. Mitsunori OGIHARA, University of Miami, Florida, United-States - *mogihara@miami.edu*
Prof. Murat GÖK, Yalova University, Yalova, Turkiye - *murat.gok@yalova.edu.tr*
Prof. Selim YAZICI, İstanbul University, İstanbul, Türkiye - *selim@Istanbul.edu.tr*
Prof. Sushil Kumar SHARMA, Texas A&M University-Texarkana, Texas, United-States - *ssharma@tamut.edu*
Assoc. Prof. Beyaz Başak ESKİŞEHİRLİ, İstanbul University, İstanbul, Türkiye - *basakoca@Istanbul.edu.tr*
Assoc. Prof. Maria DRAKAKI, International Hellenic University, Thessaloniki, Greece - *mdrakaki@ihu.gr*
Assoc. Prof. Odelia SCHWARTZ, University of Miami, Florida, United-States - *odelia@cs.miami.edu*
Assist. Prof. Gökçe KARAHAN ADALI, Haliç University, İstanbul, Türkiye - *gokceadali@halic.edu.tr*
Assist. Prof. Saeed TABAR, Ball State University, Muncie, United-States - *stabar@bsu.edu*
Assist. Prof. Shourjo CHAKRAVORTY, İstanbul Teknik University, İstanbul, Türkiye - chakravorty@itu.edu.tr

Journal of Data Applications
Issue: 3, 2024
E-ISSN: 2980-3357

CONTENTS

Araştırma Makaleleri / *Research Articles*

Journal of Data Applications
Issue: 3, 2024
E-ISSN: 2980-3357

İSTANBUL UNIVERSITY PRESS

RESEARCH ARTICLE

How to teach Data Visualisation to Fresh Statisticians: A Case Study in Turkey

Ozancan ÖZDEMİR[1], Ceylan YOZGATLIGİL[2]

ABSTRACT

This article presents the content, objectives, and student opinions regarding the first-year compulsory data processing and visualisation course taught in English, which is the first of its kind in Turkey. After exploring the academic and industrial applications of data visualisation and examining the tools used, the article highlights that the content developed was well-received by students. It was observed that all students developed a positive attitude towards data visualisation by the end of the semester. The students particularly enjoyed the group project, which involved data analysis and reporting using the Python programming language, as it allowed them to apply the data processing and visualisation steps and create a data analysis report.

Keywords: Statistics Education Research, Data Visualisation, Curriculum, Data Science Education

DOI: 10.26650/JODA.1472118

[1]University of Groningen, Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, Groningen, Netherlands
[2]Department of Statistics, Middle East Technical University, Ankara, Türkiye

ORCID: O.Ö. 0000-0002-7850-3885;
C.Y. 0000-0003-3492-0510

Corresponding author:
Ozancan ÖZDEMİR,
University of Groningen, Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, Groningen, Netherlands
E-mail: ozancan@metu.edu.tr

Submitted: 22.04.2024
Revision Requested: 26.06.2024
Last Revision Received: 07.07.2024
Accepted: 30.07.2024
Published Online: 24.01.2025

Citation: Özdemir, O., & Yozgatlıgil, C. (2024). How to teach Data Visualisation to Fresh Statisticians: A Case Study in Turkey. *Journal of Data Applications, 3*, 1-16. https://doi.org/10.26650/JODA.1472118

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International License

Introduction

Tufte (2001) defines data visualisation, which has multiple definitions in the literature, as "the graphic display of quantitative information". The modern form of data visualisation, which began with cave paintings in ancient times, dates back to the 1700s. William Playfair's visualisation of England's balance of trade in his book "Commercial and Political Atlas" published in 1786 is perhaps what has earned him the title of "the father of statistical presentation" today (Figure 1). The bar plot, line plot, and later introduced circle and pie charts by Playfair are still used today. Similarly, Florence Nightingale's created graphs, which are still used, also hold an important place in the history of data visualisation.

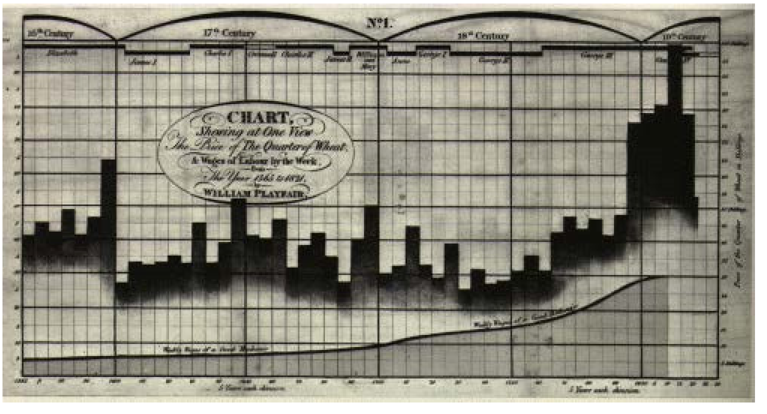


Figure 1. *Time Series Plot Created by W. Playfair*

Friendly et al. (2008) stated that the increase in the number of collected and organised data through the establishment of statistical offices in Europe starting from the 1850s, along with the development of statistical theories proposed by Gauss and Laplace to obtain meaningful results from data, marked the golden age of data visualisation. However, this momentum faded in the early 1900s and gave way to a period known as the "modern dark age" that lasted until the 1950s.

According to Friendly et al. (2008), there are three developments that concluded this approximately 50-year period and ushered in the rebirth of data visualisation:

- John W. Tukey's article "The Future of Data Analysis," published in 1962.
- Jacques Bertin's book "Semiologie Graphique," published in 1967 (Bertin, 1967).
- The emergence of FORTRAN, the first high-level language for computing that enabled the processing of statistical data in a computer environment.

Following these three developments, the increasing amount of data that needs to be interpreted and the emergence of software and programming languages suitable for data visualisation, along with technological advancements, have not only led to the use of data visualisation in various fields such as journalism, academia, and design but also transformed it into an interdisciplinary domain. Particularly during the COVID-19 period, the importance of data visualisation, which has become ubiquitous in our daily lives, has been underscored once again, resembling a hand sanitiser.

Currently, in universities, particularly in Turkey, data visualisation is taught as a subtopic in undergraduate statistics programmes as a tool for exploring data not explaining it, but it does not provide details about the storytelling aspect, which is the true power and allure of data visualisation. Questions with relevance in both academia and industry, such as how to use colour and text effectively in an ideal visual, where components like axes and legends should be positioned, and which strategies can better convey a message to the audience, remain unanswered in the existing curricula. There are not many studies on the Data Visualisation Course curriculum design for university students. Urness (2016) examined the difficulties associated with instructing a foundational course in computer graphics and provided suggestions for integrating a module focused on data visualisation. In place of a course centred solely on programming, the course introduced an alternative curriculum that encompasses diverse subjects, encompassing a module dedicated to data visualisation using Tableau, D3, processing, and Python libraries such as plotlib. The course only focuses on visualisation, not on data preprocessing, cleaning, and tidying. Bandi and Fellah (2017) presented their new data visualisation course for computer science and data science majors, where the learning outcomes of the course were aligned for those of the tech industry. They designed the course with the collaboration of IT and data services professionals from the industry so it focuses on the industry needs only such as for business track performance and make strategic decisions and visualise customer data to understand customer behaviour and preferences, for healthcare visualise patient data to identify trends, patterns, and outliers to find a better treatment, or for the manufacturing visualise production data to identify bottlenecks, improve efficiency, and reduce costs. The course emphasises on both the technical and soft skill needs of the industry. Zentner et al. (2019) explored data visualisation literature, identifying themes shaping effective data dashboard development. Through a mixed methods approach with education professionals, it found that readability was tied to functionality, titles, colour, and layout—reinforcing literature themes and providing a framework for better data visualisation in education. Asamoah (2022) presented a curriculum for teaching basic and advanced data visualisation. The survey results show a preference for hands-on learning over theory, with participants rating the curriculum components and structure highly. The study explored key components for effectively teaching a data visualisation course, however, in the course, only

Tableau was used, not on the skills in data management and programming such as Python or SQL.

Students have traditionally learned data visualisation by drawing graphs, but now interactive and dynamic visualisation tools are increasingly used. Forber et al. (2014) discussed four such tools used in New Zealand classrooms: GenStat for Teaching and Learning Schools and Undergraduate (GTL), Auckland University's iNZight and VIT for bootstrapping and randomisation, and CAST e-books, all of which are publicly available and used internationally. Queiroz et. al. (2017) emphasises that interpreting data is a complex process involving cognitive, technical, contextual, and affective aspects. However, the affective aspects are under-discussed in the literature. They examined an empirical study on the influence of affective expression during data interpretation by final-year statistics and pedagogy undergraduates, revealing that affective expressions were the most frequent category used, despite differing academic backgrounds. Hudiburgh and Garbinsky (2020) created a semester-long group project for introductory statistics courses to enhance students' skills in developing and interpreting data visualisations. This project fosters hands-on learning, mathematical reasoning, and collaboration, and this article details its structure and assessment. Wilke (2019) argues that visualisation is an essential tool for data interpretation. He also provides several practical tips for creating effective data visualisations. For example, he recommends using simple and clear designs, avoiding clutter, and using colour and shading effectively.

A few examples of research on teaching visualisation in statistics education can be given. Garfield et al. (2000) conducted a study in which they compared the performance of two groups of students on a statistical test. One group received instruction in data visualisation, while the other group did not. The students who received instruction in data visualisation performed significantly better on the test. Gould and Gould (2005) developed a curriculum for teaching data visualisation to elementary school students. The curriculum includes lessons on various data visualisation techniques, such as bar charts, line charts, and histograms. The authors found that the curriculum was effective in improving students' understanding of data visualisation and their ability to interpret data. Nolan and Perrett (2016) emphasise the importance of incorporating statistical graphics into undergraduate statistics curricula through various assignments, such as deconstructing and reconstructing plots, copying masterful graphs, and creating interactive visualisations. It discusses the goals and details of each assignment, broader concepts in graphics, and the need for increased focus on statistical graphics at all educational levels. Providing students with detailed comments and a completed matrix of competencies fosters a discussion between the instructor and student that focuses more on the process and content than on the points. The survey of Firat et al. (2022) focuses comprehensively on interactive visualisation literacy, examining and categorising prior research on user understanding and discovering visual patterns. It provides an overview of the evaluation techniques, categorises research into unique groups, and identifies both mature and unexplored areas for future study, serving as a valuable resource for researchers at all levels.

In light of these studies, it can be concluded that visualisation can help students to better understand statistics, be more engaged in statistics lessons, develop critical thinking and problem-solving skills, and communicate their statistical findings more effectively.

This article aims to address the aforementioned issues and serve as an example for other statistics departments while keeping up with the requirements of the era. It will focus on the content of the STAT 112 course, titled "Data Processing and Visualisation," which was introduced as a compulsory course in the curriculum of first-year students at the Middle East Technical University Department of Statistics for the 2022-23 Fall semester, taught in English. Additionally, the Statistics Department at Eskişehir Technical University has opened a data visualisation course for its students. Although this course is not mandatory, it is taught in Turkish and has less intensity and a different subject matter compared to the aforementioned course. The authors aim to create a guide that reflects the content of this course, which is offered for the first time to first-year statistics students in English in Turkey, and the feedback received from students who have taken this course. This guide will serve as a reference and assistance for all departments, particularly those that have not yet offered a data visualisation course, not only in Turkey but also in other regions. Also, since there is a scarcity of shared academic experiences regarding data visualisation courses in the literature, the authors aim to contribute to closing this gap with this article.

The Course: Content, Target, Objective, Progress, and Materials

Target

The target group for the data processing and visualisation course, as stated in the introductory section, is first-year statistics students who have just started their undergraduate studies. Apart from their personal interests and studies, these students do not have solid programming and technical backgrounds. This course has been added to the curriculum of these students as a compulsory course. In the 2022-23 Fall semester, 88 students took this course.

Data visualisation is first introduced using Tableau, an easy-to-use software. Students learn the basics of Python in a parallel compulsory course, Introduction to Computer Programming. In the second half of the semester, students apply their Python knowledge to data handling and visualisation using Python libraries. Datacamp classrooms, which are free for educators and their students, are also used to help students catch up with any computational challenges.

Content and Objectives

The course content was designed based on criteria such as the most commonly used software in data visualisation, software used in the business world, and the students' theoretical and practical average knowledge level. The focus of the course is on fundamental principles and best practises for data manipulation and visualisation because most of the real data is not ready for the use of visualisation directly so we comprise both data manipulation and preparation and then create an efficient visual from these data. The course starts with basic concepts in statistics such as parameters, statistics, data types and appropriate summary statistics and visualisation types for univariate and multivariate analysis such as bar charts, histograms, scatter plots, treemap, and bubble charts. It is divided into two parts:

The first part focuses on basic concepts such as data types, data manipulation, and querying. The second part focuses on data visualisation, starting with exploratory data analysis using various statistical plots. The data manipulation and visualisation methods are demonstrated using Tableau, one of the top 10 data visualisation tools (Simplilearn, 2023), as well as the Flourish and Python packages, specifically plotlib and seaborne. Students will create their own data visualisations and learn to use open-source data visualisation tools.

The course begins with a brief introduction to data visualisation, covering its development and the Gestalt Principles, and the characteristics of effective visualisation techniques. In the second week, the focus shifts to data description and types, along with appropriate visualisation tools for each type, such as bar plots, box plots, violin plots, histograms, and scatter plots. Discussions include topics like colour separation, value representation by colour, and colour palettes in data visualisation. The third week introduces students to Tableau Public, followed by an exploration of basic data manipulation techniques in the fourth week, including selecting, sorting, filtering, summarising, and combining datasets using Tableau. In the fifth week, the course explains the steps for creating visualisations and preparing dashboards with Tableau, while the sixth week covers interactive plotting using Flourish. The seventh week teaches basic data manipulation using Python, with libraries like Numpy and Pandas, and continues into the eighth week with grouping and pivoting datasets using Python. In the ninth week, data quality was addressed, focusing on aspects such as validity, accuracy, completeness, consistency, and uniformity. The tenth week covers data cleaning and tidying, including techniques like recoding, character manipulation, converting dates, string normalisation, string matching, detecting and localising errors, deductive correction, and tidying messy data. During the eleventh week, students learn to deal with common problems like missing or inconsistent values in datasets using Python packages. Week 12 focuses on visualising data with basic plots and interpreting 1D distributions using Python libraries such as Matplotlib and Seaborn. The following week explores the visualisation and interpretation of multivariate data using Python libraries. Finally, the course concludes in the fourteenth week with practical exercises based on real-life examples to reinforce the concepts learned throughout the course.

We aim to achieve comprehensive objectives to provide students with a robust foundation in data visualisation. Students will learn about different data types and perform basic data operations using Tableau and Python, including data transformations such as aggregation and filtering, indexing, slicing, and subsetting in pandas DataFrames. The course offers practical experience in building and evaluating visualisation systems, as well as designing and creating data visualisations. Students will conduct exploratory data analysis using visualisation techniques and utilise the principles of the Grammar of Graphics to convert data into visual figures using seaborne and plotlib libraries. They will also identify potential challenges and pitfalls when working with data in Tableau and Python. The course emphasises the creation of basic and visually appealing diagrams using Tableau and Python, and the organisation of visual presentations of data for effective communication. Additionally, students will prepare dashboards and develop storytelling techniques to impress specific audiences. The design and evaluation of colour palettes for visualisation based on the principles of perception are also covered. Finally, students will identify opportunities for the application of data visualisation in various domains, ensuring that they can apply their skills in various domains.

The class was conducted through in-class lessons, recitation hours for exercises with real-life examples, two projects (one using Tableau to do data cleaning/tidying, EDA, and preparing an impressive dashboard by themselves using the same dataset so there were 88 different EDA and dashboard design, and the other using Python to do cleaning and do EDA with excellent visualisations with a group of 5 students), and one in-class midterm exam.

We prefer Tableau because it is a user-friendly data visualisation tool that does not require any prior programming knowledge. This makes it a good choice for students who are just starting to learn about data visualisation. Moreover, Python is a versatile programming language that can be used for various data tasks, including data cleaning, preprocessing, and visualisation. It is also a popular language for data science and machine learning. Tableau and Python can be used together to create powerful and informative data visualisations. For example, students can use Tableau to create interactive dashboards and Python to create custom visualisations. These are the most important reasons to decide on using these tools in the course.

The grammar of graphics is a set of principles for creating effective data visualisations. It provides a framework for thinking about the different elements of a visualisation, such as data, marks, channels, and aesthetics. Hence, the specific capabilities of the computing platform may affect the way that the grammar of the graphics is implemented. Some computing platforms may have limitations on data encoding, interaction, and output formats, such as the use of certain colour palettes or mark types, interaction techniques, or export to certain file formats.

Students have almost a month to complete their first project, which is an individual report on the EDA and Tableau dashboards. For their second project, students worked in groups to complete data cleaning and tidying, create EDA, and answer research questions using descriptive statistics, tables, or visualisations in a month. In their reports and visualisations, students are expected to follow the Gestalt principles and grammar of the graphics principles.

Progress and Materials

To encourage students to build their own data science portfolios and showcase their future projects, a GitHub account for the course was created, which includes recitation notes, various data visualisation resources, and introductory examples of pandas and NumPy for Python (https://github.com/MetuStat112). Additionally, an associated GitHub-hosted website was created (https://metustat112.github.io/) to serve as a repository for course materials.

After introducing the basics of data types, relevant visualisation types, and their interpretation in both class and recitation hours, students were given real-life datasets to perform data preprocessing and cleaning using Tableau Public and SQL commands. They were then tasked with creating visualisations based on these cleaned datasets and learning the fundamentals of exploratory data analysis (EDA).

Over a span of seven weeks, the course covered topics such as data usage in visualisations, colour usage, and Gestalt Principles, as well as practical applications using Tableau. For the project, students were provided with the New York City Airbnb Open Data from Kaggle. They were asked to clean the dataset, address issues such as user-based errors, lowercase/uppercase inconsistencies, and missing values, and then create a dashboard using Tableau to highlight a topic of their choice. The dashboards published on Tableau Public were compiled and shared on the course's GitHub-hosted website https://metustat112.github.io/airbnb_dashboard.html. One of the dashboards created by a student who took the course is shown in Figure 2.

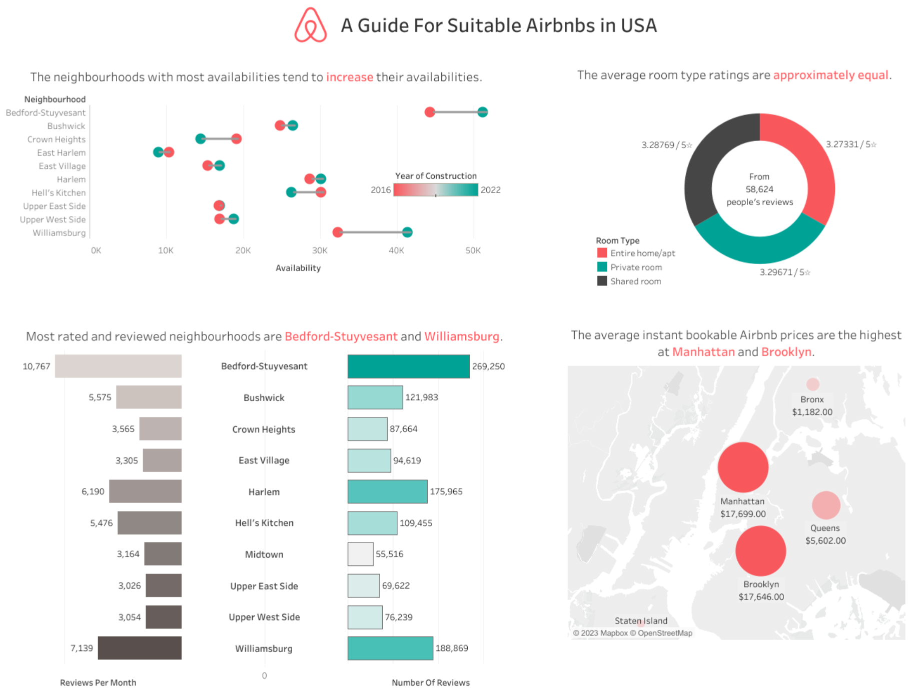


Figure 2. *A Dashboard created by a student on Tableau*

Starting from the second half of the semester (Week 7), the course transitioned to the Python programming language. After introducing commonly used packages and applications for data manipulation, such as NumPy and pandas, students were given a dirty dataset to clean. In addition, basic statistical calculations such as mean, median, mode, standard deviation, and correlation were demonstrated in Python. Visualisation techniques using the plotlib and seaborne packages were explored, along with strategies to improve the visual quality of the graphs produced with these packages.

In the final class, a data visualization expert delivered a presentation to expose students to real-life examples of data visualisation and storytelling.

During the semester, DataCamp for Classrooms was provided to students, and several courses or chapters were assigned as homework assignments related to the topics covered in the class.

At the end of the semester, a group projects conducted to assess and measure students' knowledge and skills in data processing, visualisation, and basic statistics. Students were given a 16-step checklist to follow for data cleaning and tidying. Then, students were divided into groups of 4-5 individuals and provided with synthetic and dirty datasets based on fictional scenarios such as housing prices in Ankara, salary of non-existing NBA players, and exam scores of the non-existing students from nonexistent high school. We aim to help students apply all the techniques that we learned, and it is difficult and time consuming to find the real datasets compromising all the problems separately. This is the main reason for our choice regarding the synthetic data. In these datasets, both categorical and continuous variables are incorporated. These datasets include randomly introduced missing observations and outlier values. Students were required to clean the data, fix the problems, answer 6-8 research questions that they created using visualisations and statistics, and write a 10-page report. Furthermore, to assess the fairness of the group work, all group members were required to write a reflection report. In this report, they answered the following questions:

- What were your overall feelings about this project?
- Did this project help you understand the data preprocessing and visualisation steps in data analysis any better?
- How did your group work together? How many times did you meet to discuss the project? Were the meetings online or face-to-face? Usually, how long do the meetings take?
- What was your role in the group?
- Were there any group members who did not pull their weight? Any group members who tried bossing the group around?

In their reflective reports, students commonly expressed a sense of relief and found the concept of a final project, instead of a final exam, to be beneficial. They highlighted that the project facilitated their understanding of the importance of data organisation, cleaning, and subsequent visualisation—the essential stages of a real-life data visualisation process. Additionally, they discussed the benefits and challenges of teamwork, elaborating on the tasks they undertook both as a team and individually within their groups.

There were very few students who did not join the project meetings and did not give any effort to the project, so they failed the class. The grade distributions of the students are given in the following table (Table 1).

Table 1. *The grade distribution of students*

AA	BA	BB	CB	CC	DC	DD	FD	FF	NA
28	11	14	10	1	6	9	5	1	3

Approximately 60% of the students achieved a grade of BB (3.0 out of 4.0) or higher at the end of the semester.

Student Insights

At the end of the semester, a questionnaire was given to the students, and feedback was collected from some of the students who took the course for the first time, aiming to gain a general understanding of the effectiveness of the course. During this feedback process, students' identities were kept confidential to prevent bias in the results.

The following graph illustrates the overall attitude of students towards the course. Students were asked to rate the course on a scale of 1 to 10, and it was observed that the average score given by students was 8.3 out of 10. This indicates that the teaching above-mentioned approach had a positive impact on the students (Figure 3).

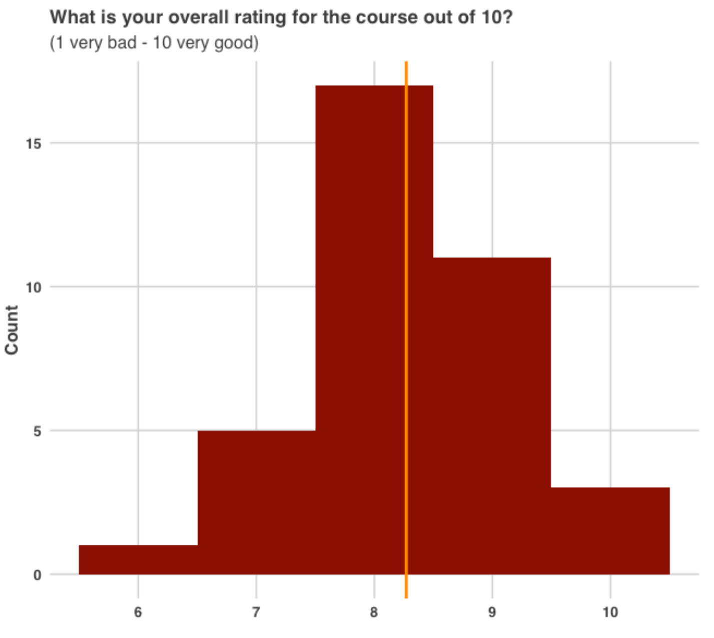


Figure 3. *Overall Rating of the Course, according to students*

Another result that supports this positive impact is shown in the graph below, which depicts the difference between students' perceptions of data visualisation before and after taking the course. While almost half of the students did not have a positive perception of data visualisation before taking the course, it is evident that this situation changed by the end of the semester (Figure 4).

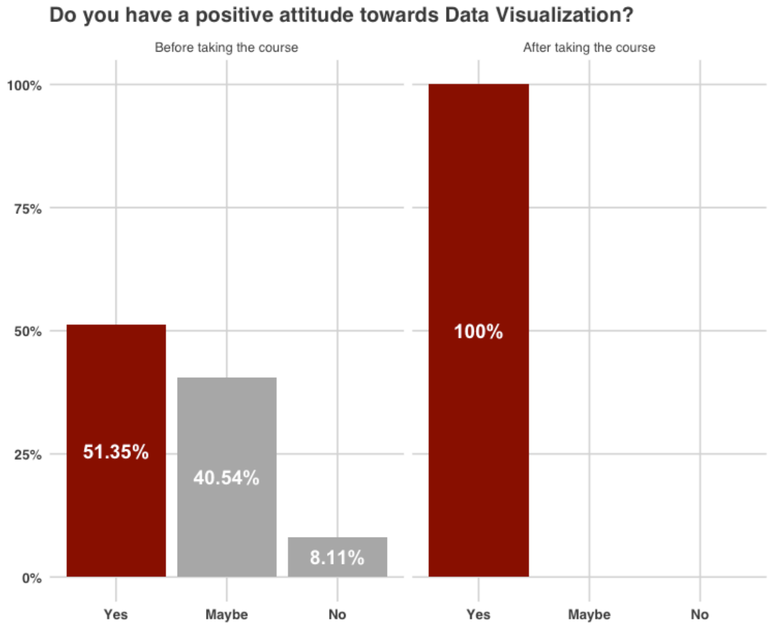


Figure 4. *The Attitude of the Students on Data Visualisation before the course and after the course*

The fact that students mentioned finding the course helpful is another piece of evidence supporting this positive impact (Figure 5).

Figure 5. *The words that students have used to describe the course*

Approximately 60% of the students who took the course expressed satisfaction with both the Tableau and Python components, while 32% specifically mentioned being satisfied with Tableau, and 10% expressed satisfaction with Python. However, when students were asked which project they found beneficial, the majority preferred the project involving Python, which included coding and reporting (Figure 6).

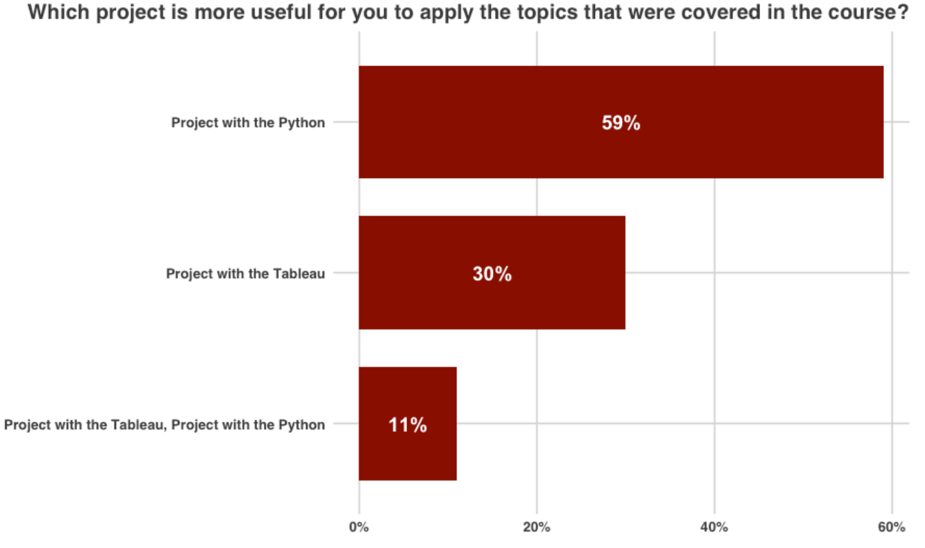


Figure 6. *The breakdown of students' preferences for their favourite part of the course*

When students were asked questions on a 5-point Likert scale to assess their overall thoughts on the course, it was generally observed that they had positive opinions (Figure 7). However, it was also noted that there was a group of students who did not consider the course academically challenging.

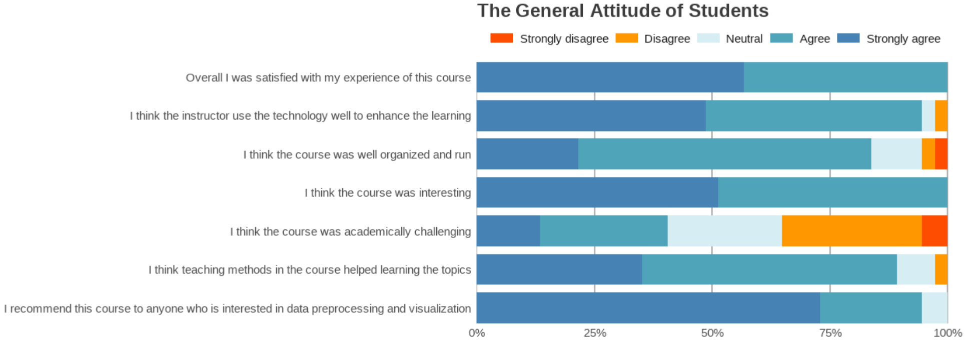


Figure 7. *The distribution of the students by the part of the course that they liked most.*

Conclusion

We provided guidance to educators aspiring to design a data cleansing and visualisation course with this article. Our journey through this lecture design has underscored the importance of data integrity, emphasising that the quality of analysis and visualisation hinges upon the accuracy and reliability of the underlying data.

Throughout this course, students were immersed in the dynamic interplay between raw data and meaningful insights, mastering techniques that unveil concealed patterns and narratives. This skill set not only hinges on data integrity but also encompasses the artistry of impactful visualisation. By mastering techniques ranging from data cleaning and transformation to advanced visualisation tools, students have gained a multifaceted toolkit that empowers them to unravel complexity and present it in a digestible and engaging manner.

The importance of data integrity and visualisation is well-documented in the literature. According to Kandel et al. (2011), data wrangling is crucial for exploratory data analysis and achieving accurate results. Similarly, Heer, et. al. (2010) emphasise that effective data visualisation enables users to understand and communicate complex data relationships.

As the digital landscape continues to expand, the ability to manipulate and visualise data with precision and creativity has become an indispensable skill across disciplines. This lecture design serves as a launch for future explorations, equipping students with the acumen to navigate the dynamic world of data-driven decision-making and storytelling. This lecture design not only imparts technical expertise but also nurtures qualities such as attention to detail, critical thinking, and the ability to convey complex ideas lucidly.

In a world where data drives decision-making, these skills stand as pillars of competence. The ability to navigate complex datasets, distil meaningful information, and present it visually is a coveted asset across disciplines. Companies and industries recognise the value of employees who can extract actionable knowledge from data, and academia benefits from researchers who can convey complex concepts in accessible formats.

Chen et. al. (2017) noted that the ability to present data visually is not only about technical skills but also about storytelling and communication. Few (2006) argues that effective data visualisation can lead to better decision-making and insights.

Students depart with more than technical prowess—they depart as storytellers who can navigate the intricacies of data and present it with impact. Armed with Tableau, Python, and a profound understanding of data's visual language, they are poised to excel, shaping the landscape of both academia and industry with their data-driven insights.

At the end of the course, based on the conducted surveys, all students were satisfied with the course. The majority of them also recognised the efficient use of technology to improve learning levels. Moreover, most of the students believed that the curriculum explained above was well-organised and interesting and could be recommended to anyone willing to learn data visualisation. However, it is seen that they did not agree with the idea that the course context was academically challenging, which might be the point that should be improved for the following semesters. Another interesting result is that more than half of the students found the project with Python more useful than the one with Tableau. This may imply that students would prefer to start the data visualisation process from scratch rather than play with the prepared data and go beyond the limitations of drag-and-drop software like Tableau. Based on this feedback, the course content may be modified to maintain the balance between Python and Tableau.

The literature also supports the the preference for Python over Tableau among students. Python's extensive libraries and flexibility allow for deeper data manipulation and complex visualisations which are often required in advanced data science tasks. This is in contrast to Tableau's user-friendly interface, which, while excellent for creating quick visualisations, can be limited in terms of customisation and complex data processing capabilities (MindBowser, 2023; DataCamp, 2023). This distinction highlights the necessity for educators to balance both tools in their curriculum to cater to diverse learning needs and professional applications.

Ethics Committee Approval: We declared that this study does not require ethics committee approval.

Peer-review: Externally peer-reviewed.

Author Contributions: Conception/Design of Study- O.Ö., C.Y.; Data Acquisition- O.Ö., C.Y.; Data Analysis/Interpretation- O.Ö., C.Y.; Drafting Manuscript- O.Ö., C.Y.; Critical Revision of Manuscript- O.Ö., C.Y.; Final Approval and Accountability- O.Ö., C.Y.; Technical or Material Support- O.Ö., C.Y.; Supervision- O.Ö., C.Y.

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: The authors declared that this study has received no financial support.

References

Asamoah, D. (2022). Improving Data Visualization Skills: A Curriculum Design. International Journal of Education and Development using Information and Communication Technology (IJEDICT), 18(1), 213–235.

Bach, B., Keck, M., Rajabiyazdi, F., Losev, T., Meirelles, I., Dykes, J., Laramee, R. S., AlKadi, M., Stoiber, C., Huron, S., Perin, C., Morais, L., Aigner, W., Kosminsky, D., Boucher, M., Knudsen, S., Manataki, A., Aerts, J., Hinrichs, U., Roberts, J. C., & Carpendale, S. (2023). Challenges and Opportunities in Data Visualization Education: A Call to Action. *IEEE Transactions on Visualization and Computer Graphics*. https://doi.org/10.1109/tvcg.2023.3327378

Bandi, A., & Fellah, A. (2017). Crafting a data visualization course for the tech industry. *Journal of Computing Sciences in Colleges*, 33(2), 46–56.

Chen, C. H., Härdle, W. K., & Unwin, A. (2017). *Handbook of Data Visualization*. Springer Science & Business Media.

DataCamp. (2023). *Visualizing Data with Python and Tableau Tutorial*. Retrieved from www.datacamp.com.

Eberhard, K. (2023). The Effects of Visualization on Judgment and Decision-Making: A Systematic Literature Review. *Manag Rev Q*, 73, 167–214. https://doi.org/10.1007/s11301-021-00235-8

Few, S. (2006). *Information Dashboard Design: The Effective Visual Communication of Data*. O'Reilly Media, Inc.

Friendly, M., Chen, C., Härdle, W. K., & Unwin, A. (2008). *A Brief History of Data Visualization*. https://doi.org/10.1007/978-3-540-33037-0_2

Firat, E. E., Joshi, A., & Laramee, R. S. (2022). Interactive visualization literacy: The state-of-the-art. *Information Visualization*, 21(3), 285-310. https://doi.org/10.1177/14738716221081831

Forbes S, Chapman J., Harraway J, Stirling D., Wild C. (2014). Use of Data Visualisation in the Teaching of Statistics: A New Zealand Perspective. *Special Issue: A Global View of Statistics Education Research*, 13(2), 187-201. https://doi.org/10.52041/serj.v13i2.290

Garfield, J. B., delMas, R., Chance, B., & Ooms, A. (2000). *Activities for teaching statistics: Grade level K-12*. Addison-Wesley Longman.

Gould, R., & Gould, S. (2005). *Visualizing data: Elementary and middle school strategies*. Heinemann, Portsmouth, NH.

Heer, J., & Shneiderman, B. (2012). Interactive Dynamics for Visual Analysis: A Taxonomy of Tools that Support the Fluent and Flexible Use of Visualizations. *Comm. of the ACM*, 1–26.

Heer, J. & Bostock, M. & Ogievetsky, V. (2010). A tour through the Visualization zoo. *Commun, ACM*, 53, 59-67. https://doi.org/10.1145/1743546.1743567.

Hudiburgh, L. M., & Garbinsky, D. (2020). Data Visualization: Bringing Data to Life in an Introductory Statistics Course. *Journal of Statistics Education*, 28(3), 262–279. https://doi.org/10.1080/10691898.2020.1796399

Kandel, S., Paepcke, A., Hellerstein, J. M., & Heer, J. (2011). *Wrangler: Interactive visual specification of data transformation scripts*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA, 3363–3372. https://doi.org/10.1145/1978942.1979444

Playfair, W. (1821). *A letter on our agricultural distresses, their causes and remedies*. Longman, Hurst, Rees, Orme, Brown, and Green, London.

MindBowser. (2023). *Tableau vs Python - Which One Is Better For Data Science?*. Retrieved from www.mindbowser.com.

Nolan, D., & Perrett, J. (2016). Teaching and Learning Data Visualization: Ideas and Assignments. *The American Statistician*, 70(3), 260–269. https://doi.org/10.1080/00031305.2015.1123651

Shneiderman, B. (1996). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. Proc. IEEE Symp. *Visual Languages*, 336–343.

Simplilearn. (2023). *The State of Data Visualization Tools in 2023*. Simplilearn Blog. https://www.simplilearn.com/data-visualization-tools-article

Tufte, E. (2001). *The visual display of quantitative information* (2nd ed.). Graphics Press, Cheshire, Conn. http://www.amazon.com/Visual-Display-Quantitative-Information/dp/0961392142

Urness, T. (2016). Incorporating data visualization in a course on computer graphics. *Journal of Computing Science in Colleges*, 31(5), 147–154.

Wilke, C. O. (2019). *Fundamentals of data visualization*. O'Reilly Media.

Zentner, A., Covit, R., & Guevarra, D. (2019). *Exploring Effective Data Visualization Strategies in Higher Education*.

Queiroz T., Monteiro C., Carvalho L., & François K. (2017) Interpretation of Statistical Data: The Importance of Affective Expressions *Special Issue: Statistical Literacy*, 16(1), 163-180. https://doi.org/10.52041/serj.v16i1.222.

Journal of Data Applications
Issue: 3, 2024
E-ISSN: 2980-3357

RESEARCH ARTICLE

Forecasting Coal Production in India: A Time Series Approach

Avni GANGWAR[1], Diksha RATHOR[1], Praveen Kumar TRIPATHI[1]

DOI: 10.26650/JODA.1557949

ABSTRACT
This article is intended to produce the forecasts for coal production in India through some time series models. This study describes the component-based and correlation-based time series models for its purpose. The separate analyses were performed by applying Naïve, Holt's and ARIMA models on a real data set based on the coal production in India between 1980 and 2022. On the basis of the retrospective predictions and accuracy measure results, an ARIMA (2,2,2) model was selected as a good choice for the data in hand. A particular ARIMA (2,2,2) model was selected by using the AIC and BIC of model selection. For the validity of the finally selected ARIMA (2,2,2) model, a residual diagnostics check has been performed; and the future predictions have been made for the next 5 years. Such an analysis is expected to add some new approaches in the literature of forecasting the energy sources, especially with reference to India.

Keywords: Time Series, Coal Production, Model Selection Criteria, Residual Diagnostic, Prediction

[1]Department of Mathematics and Statistics, Banasthali Vidyapith, Rajasthan, India

ORCID: A.G. 0009-0001-5376-1925; D.R. 0009-0009-5078-365X; P.K.T. 0000-0002-5652-8301

Corresponding author:
Praveen Kumar TRIPATHI,
Department of Mathematics and Statistics, Banasthali Vidyapith, Rajasthan, India
E-mail: pkt2088@gmail.com

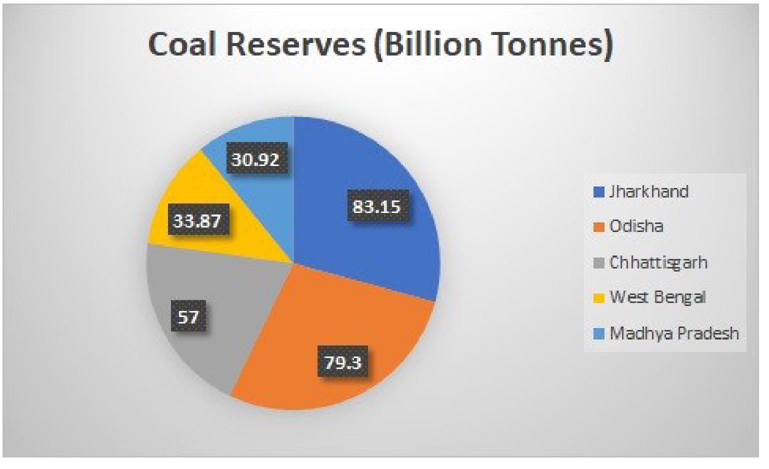
Submitted: 05.10.2024
Revision Requested: 27.11.2024
Last Revision Received: 30.11.2024
Accepted: 09.12.2024
Published Online: 20.12.2024

Citation: Gangwar, A., Rathor, D., & Tripathii P. K.. (2024). Forecasting Coal Production in India: A Time Series Approach. *Journal of Data Applications, 3*, 17-32. https://doi.org/10.26650/JODA.1557949

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International License

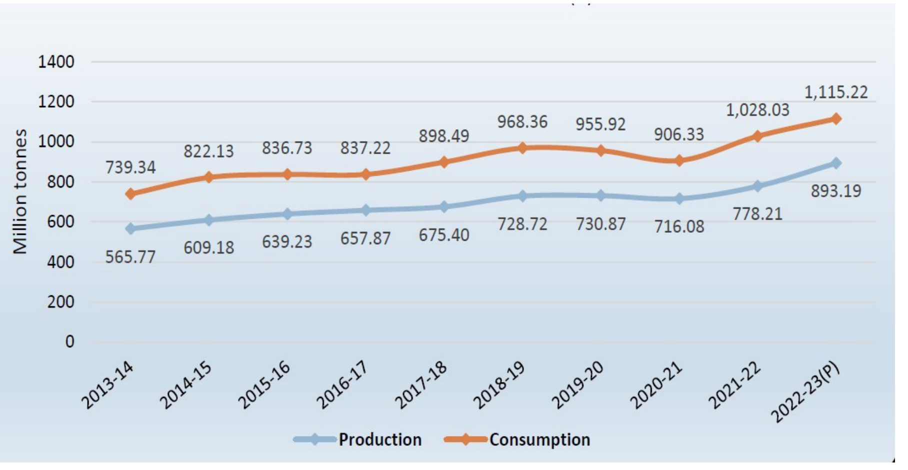
Introduction

Coal mining in India started in 1774, and at present, India is the second largest country, after China, in terms of coal production and consumption. Surpassing the historical figures, India produced 1 billion metric ton (MT) of coal and lignite in March 2024 and touches the 'historical high'. Despite its enormous production, India has to import coking coal to meet its domestic demands. Non-coking coal is imported by the coal-based power plants, captive power plants, sponge iron plants, cement industries and coal traders. Being the most significant and abundant fossil fuel in India, the coal fulfils almost 55% of the energy demand in the country. The major coal-producing states in India are; Jharkhand, Odisha, Chhattisgarh, West Bengal and Madhya Pradesh, with their combined contribution being around 90% in total (see, for example, MOSPI 2024) and the same is demonstrated in Figure 1.



[*Data source: Based on data from MOSPI (2024).]
Figure 1. *Contribution by the major states in India.*

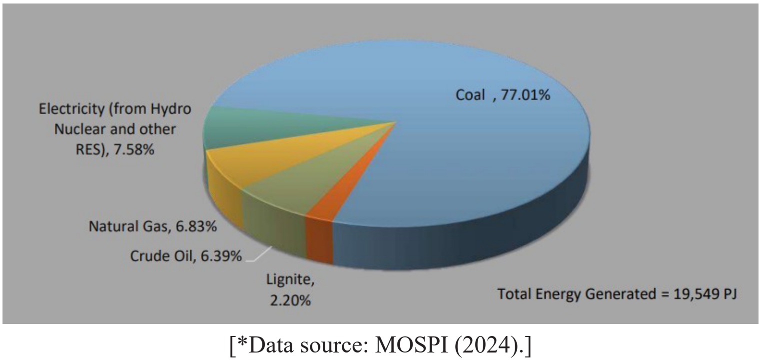
The study says that India's reliance on coal, especially in the power sector is more due to political inclusion than economic and technological reasons (see, for example, Montrone et al. 2021). Both coal production and consumption in India have been increasing over the years, and even more consumption is recorded from 2013-14 to 2022-23 (see, for example, MOSPI 2024). A comparison of the trend plots is shown in Figure 2, which is borrowed from MOSPI (2024). At present, India has sufficient stock of coal with a record production of 1 billion MT.



[*Data source: MOSPI (2024).]

Figure 2. *Trend of Production and Consumption of Coal in India from 2013-2014 to 2022-2023(P).*

Moreover, coal has the largest share in the overall energy generated in India from the different commercial sources. Figure 3 demonstrates that coal is the source of energy and it contributes maximum to the total energy generated in India during 2022-2023 (see, for example, MOSPI, 2024). Due to its vast consumption in different sectors like, steam energy, metallurgy, electricity, raw material for the industries, domestic purpose, etc., the coal is high in demand and, therefore, its production becomes a crucial aspect for various sectors in India. It helps the economic growth in an indirect way through energy generation, industrial developments, job creation, etc. Despite its needs in different sectors, the environmental concerns and health issues are some major points where the government has to think about some cleaner energy sources.



[*Data source: MOSPI (2024).]

Figure 3. *Share of total energy generated (in petajoule) from different commercial sources in India during the financial year 2022-2023(P).*

The above study from the different government organisations motivates us to analyse the coal production in India and to make a reasonable forecast based on historical time series data. With this aim, we have considered some basic time series models with their specific features of statistical model buildings. Loosely speaking, we consider the two component-based time series models, namely, Naive and Holt's linear trend models; and a correlation-based model, namely, the autoregressive integrated moving average (ARIMA) model. The component-based model incorporates the changes in the dataset due to the different components of a time series. For instance, Holt's model captures the inherent trend in the dataset and provides future estimates accordingly. On the other hand, correlation-based models incorporate the fluctuations due to the serial correlations in a data set (see, for example, Hyndman and Athanasopoulos 2018). By considering the ARIMA model, one may ensure that the future estimates are available due to the consideration of the autocorrelation feature of the real data in the past. It is, then, assumed that the correlation characteristics shall remain unchanged in the future. Now, a comparison between the two methods will tell us which feature is more relatable when we go for a future prediction of coal production in India.

The statistical literature on coal production in India is hardly available except, perhaps in the form of data accumulation, survey and report formats on the government websites. Our aim is to analyse the coal production data in India and to choose the most effective model on the basis of forecast accuracy measures. Our study also reveals the fact that which type of model is appropriate for the considered real data between the two described categories of time series models. Some of the recent works on coal data using time series approaches include Li et al. (2019), Makkhan et al. (2020), Chen et al. (2021), Parren˜o (2022), Jai Sankar et al. (2023), Mohanty and Nimaje (2023), among others. In particular, Li et al. (2019) forecasted the coal production in India by 2030 through a combined time series model after some modifications in their linear and non-linear setups. Makkhan et al. (2020) have analysed the black carbon in the regions of coal mines in India through the classical tools of statistics, such as; correlation coefficient, rank and Kandall's tau correlations and ARIMA models. Chen et al. (2021) used the descriptive and graphical tools for a statistical analysis of 'Chongqing coal mine' accidents somewhere in China. Parren˜o (2022) used the ARIMA model to forecast the coal production and consumption in the Philippines. Mohanty and Nimaje (2023) have applied the neural network and ARIMA models to forecast the fatal coal mine accidents in India, where they found the neural network model to be the best fitted model for the data under consideration. In another study, Jai Sankar et al. (2023) have modelled and analysed the coking coal production in India through the ARIMA model and concluded that coking coal production will be declining in India by, 2031.

Since the time series literature on coal production in India is quite limited, no recent study has unveiled the gross coal production in India using a time series model. This paper attempts to fill this gap and provide a systematic future prediction of coal production in India by choosing the most suitable time series model.

The rest of the paper is organised as follows. Section 2 defines the basic structures of the considered time series models. Section 3 describes the dataset and the necessary steps to proceed with the whole analysis. This particular section completes the whole analyses, for the real data set of coal production in India, under different subsections. The last section concludes the paper, which provides a smooth ending to the whole work.

Some basic time series models

Naive model

It assumes that the future value of a time series will be matched by the last observation of the series, and is defined as follows:

$$\hat{y}_{t+h} = y_t. \qquad \textit{Eq. 1}$$

In Eq. 1, $\hat{y}_{t+h}$ is the h -step ahead forecasted value of the most recent observation y_t at current time t. The forecasts obtained by the naive model are also termed as the 'random walk forecasts' (see, for example, Hyndman and Athanasopoulos (2018)).

Holt's linear trend model

Holt (1957) proposed a linear trend model that allows the forecasting of a time series with trend. Holt's model smoothens the time series data in the two phases; one for the level and the other for the trend. Eq. 2 defines the mathematical form of Holt's model as,

$$\boldsymbol{y_{t+h} = L_t + hT_t}$$

$$\boldsymbol{L_t = \alpha y_t + (1-\alpha)(L_{t-1} + T_{t-1})}$$

$$\boldsymbol{T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}}, \qquad \textit{Eq. 2}$$

where L_t and T_t denote the estimates of the level and trend, at time t, respectively, of the time series data; and $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$ are the smoothing parameters for the level and trend, respectively. The trend of the time series plot is identified by its slope at a time (see, for example, Hyndman and Athanasopoulos 2018).

ARIMA model

This model was proposed by Box and Jenkins (1970) and termed as 'Box-Jenkins Model' in the time series literature. The ARIMA model is defined for the differenced series of data, $x_t = \Delta^d y_t$, and can be written as;

$$x_t = c + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} + e_t, \quad Eq.\ 3$$

where c, is the intercept term, and (p,d,q) denotes the combined order of the ARIMA model consisting of the order of the autoregressive (AR) components, degree of differencing and order of the moving average (MA) components in sequence. Also, in Eq. 3, e_t's are independently and identically distributed (i.i.d.) normal variates each with a common mean and variance of 0 and σ^2 respectively (see, for example, Agarwal et al. 2021).

Data description and related analysis

For our analysis, we have considered real data on the annual coal production in India approximately 1980-81 to 2021-22 (see the table in Appendix). The dataset is borrowed from 'Economic Survey 2022-23' released by the Government of India (see India 2024). A time series plot of the data is shown in Figure 4. The plot exhibits an increasing trend of coal production in India approximately 1980-81 to 2021-22, which is obviously a non-stationary pattern of the time series.

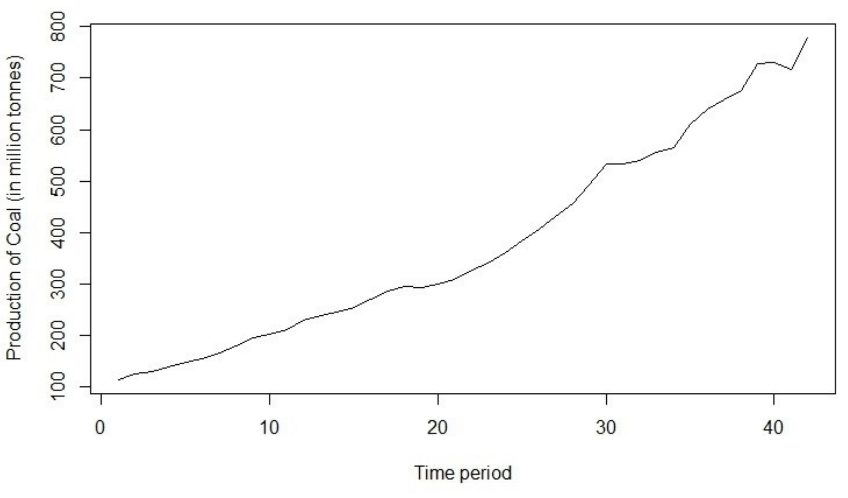


Figure 4. *Time series plot for coal production in India approximately 1980-1981 to 2021-2022.*

We, first, have proceeded with the component-based time series model prediction and classified our data into two parts; the first 36 observations as 'training-sample data' and the remaining 6 as 'test-sample data'. We have analysed the training-sample data by applying the Naive and Holt's models and then predicted for the next 6 observations corresponding to the 'test-sample' data using the R software, and the same is exhibited in Figure 5. As observed, the prediction based on naïve method is constant, and it can be considered as the base level for a further future prediction. On the contrary, Holt's linear trend method exhibits a sharply increasing pattern of the coal production in India. The retrospective predictions based on the two methods can be concluded in two different ways. First, Holt's method provides an over-fitted prediction than the naïve method. Second, none of the two methods is capable of to capture the actual fluctuation of the considered time series. An overall conclusion may be that the component-based time series models, considered here, are not good enough to obtain the future prediction of the coal production in India. We shall, therefore, be looking into some other choices of time series models like ARIMA models in the coming section.

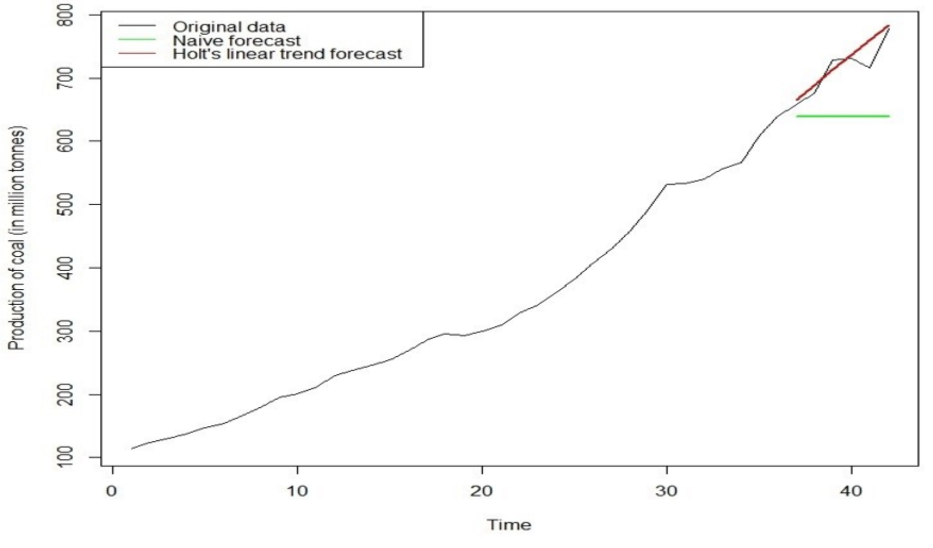


Figure 5. *A retrospective prediction based on naive and Holt's linear trend methods.*

Analysis with the ARIMA model

Analysis using the ARIMA model consisting three stages; model identification, estimation and validation (see, for example, Box et al., 2015). Model identification involves deciding the order of the ARIMA model, that is, realisation of (p, d, q). Model estimation includes the estimation of the AR and MA coefficients in the ARIMA model equation (3); and model validation consists of the authentication of the finally selected ARIMA model for the data in hand. Before we proceed with the three steps, we shall verify whether the time series is stationary or not. As mentioned before, Figure 4 depicts a non-stationary movement over time and, therefore, an appropriate transformation is needed to make it stationary (see, for example, Agarwal et al., 2021). In our case, we made the double differencing of the log-transformed data to obtain a stationary pattern (see Figure 6). We have also performed the 'stationary check' at every stage of transformation for the assurance by using the two most commonly used tests, namely, the augmented Dickey-Fuller (ADF) test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. The ADF test rejects a null hypothesis of the unite root in the data at the 5% level of significance; while the KPSS test does not wish to reject the null hypothesis of stationarity again at the 5% level of significance. For more details on the two tests, one may refer to Tripathi et al. (2017, 2021), among others. The outputs of the two tests are shown in Table 1, which indicates that the log-transformed double-differenced data achieved stationarity.

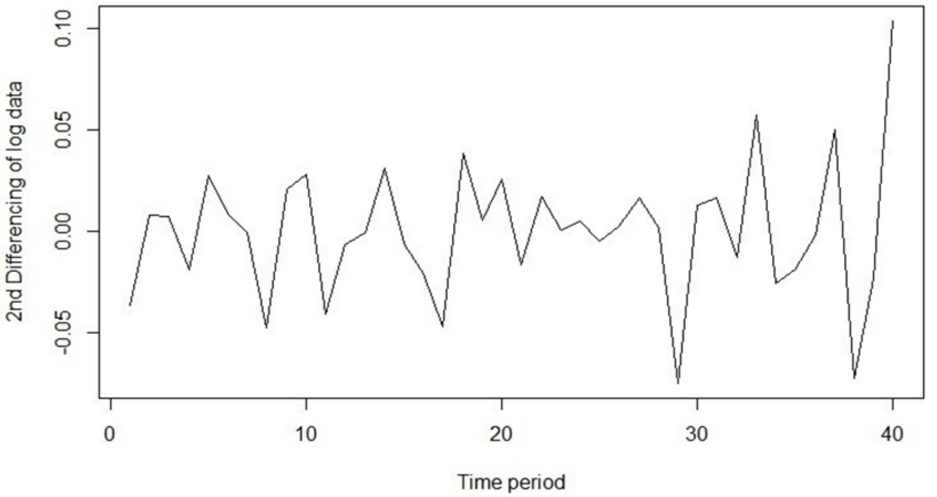


Figure 6. *Time series plot of double-differenced and log-transformed coal production data.*

Table 1. *Outputs of the ADF and KPSS tests for the different forms of time series data.*

Time series	p-value	
	ADF test	**KPSS test**
$z = \log(x)$	0.53	0.01
Δz	0.32	0.02
$\Delta^2 z$	0.01	0.10

To assess the order of the ARIMA model, we plotted the autocorrelation function (ACF) and partial ACF (PACF) values against different lag values in Figure 7, and observed the exact cut-off beyond the significant limits (the dotted lines). One may observe the exact cut-off at lag 2 in each of the two plots of Figure 7. Generally, the PACF cut-off decides the order of the AR process, that is, p and the ACF cut-off decides the order q in the MA process. Since our, further analysis will be based upon stationary data only, which is obtained after double differencing, the order of integration in the ARIMA model will be d = 2. Hence, our selected ARIMA model for the data is ARIMA (2, 2, 2).

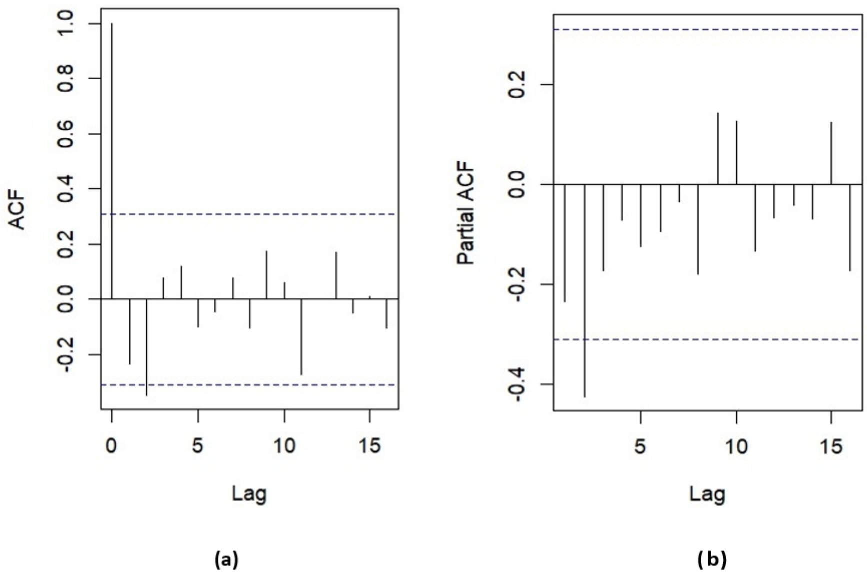


Figure 7. *ACF and PACF plots for the stationary data.*

To be more selective, we may consider some other nearby choices of ARIMA models such as ARIMA (0,2,1), ARIMA (0,2,2), ARIMA (1,2,0), ARIMA (1,2,1), ARIMA (1,2,2), ARIMA (2,2,0) and ARIMA (2,2,1). Such an approach may avoid any misleading conclusion based on a single choice of model assessment (see, for example, Tripathi et al. 2018). A final selection of the ARIMA model is done on the basis of two criteria, namely, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), which are defined as;

$$\mathrm{AIC} = -2log\hat{L} + 2k,$$

$$\mathrm{BIC} = -2\log\hat{L} + k * log\,(T - p), \qquad \textit{Eq. 4}$$

where k and $\hat{L}$ represent the number of parameters and the estimated likelihood of the ARIMA model respectively. Table 2 reports the AIC and BIC values of the considered ARIMA models for the coal production stationary data. As a thumb rule, the model having the least values of AIC (BIC) will be selected. Table 2 shows that the ARIMA (2,2,2) model was selected as the best fitted model for the data in hand.

Table 2. *AIC and BIC values of the considered ARIMA models.*

ARIMA models	AIC value	BIC value
ARIMA (0,2,1)	272.79	282.01
ARIMA (0,2,2)	274.79	281.74
ARIMA (1,2,0)	272.79	282.00
ARIMA (1,2,1)	275.00	281.96
ARIMA (1,2,2)	275.78	284.47
ARIMA (2,2,0)	274.79	281.75
ARIMA (2,2,1)	276.32	285.01
ARIMA (2,2,2)	**270.29**	**281.29**

To estimate the model parameters in ARIMA (2,2,2), we have maximised the likelihood function, given in Eq. 5, by using the non-linear minimisation (*nlm*) function in the R software. Eq. 5 represents the conditional likelihood function, up to proportionality, of the ARIMA (2,2,2) model for the stationary data set $x: x_1, x_2, \ldots, x_T$.

$$f(\underline{x}|\Theta) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{T-2}{2}} \times \exp\left(-\frac{1}{2\sigma^2}\sum_{t=3}^{T}\left(x_t - c - \sum_{i=1}^{2}\phi_i x_{t-i} - \sum_{j=1}^{2}\theta_j e_{t-j}\right)^2\right), \quad \textit{Eq. 5}$$

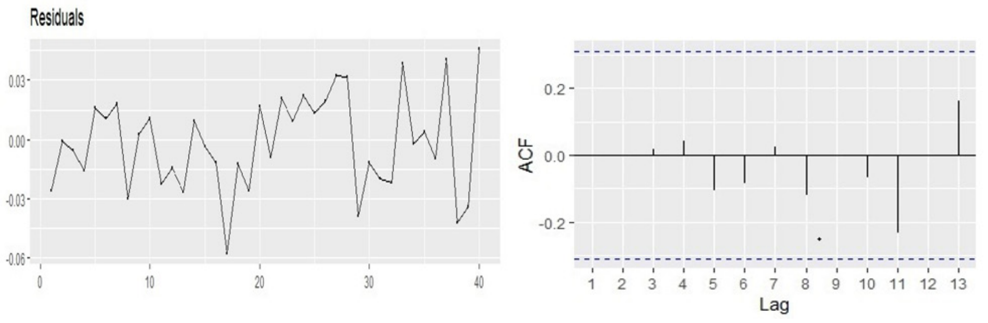
where $\Theta = (\sigma^2, c, \phi_1, \phi_2, \theta_1, \theta_2)$, a set of model parameters. The maximum likelihood estimates (MLE) of the parameters of the selected ARIMA (2,2,2) model are reported in Table 3.

Table 3. *MLE of the ARIMA (2,2,2) model*

Parameters	σ^2	c	ϕ_1	ϕ_2	θ_1	θ_2
MLE	0.80	-0.07	-1.02	-0.70	1.20	0.95

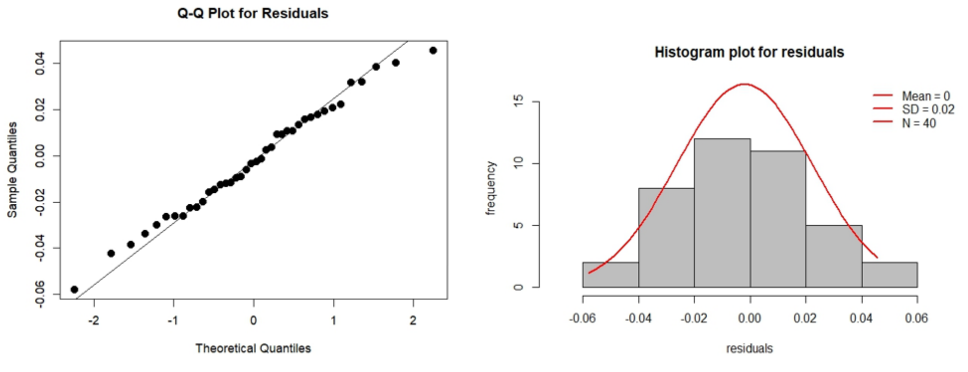
Next, to determine the appropriateness of the ARIMA (2,2,2) model, we have gone through the residual diagnostic checks. A residual is simply the difference between the actual and fitted observations of the data. Ideally, the residuals should be uncorrelated and possess the mean equal to “zero” to get utilised all the information from the data and to avoid any kind of bias in the forecasting respectively (see, for example, Hyndman and Athanasopoulos, 2018). The calculated residuals and their ACF are plotted in Figure 8-a and Figure 8-b, respectively. In the residual plot, no abrupt fluctuation is observed and the ACF plot further indicates that there is no serial correlation between the residuals over the different lags.

Another qualification in the residual diagnostic is that the residuals should follow a normal distribution with a mean of zero and a constant variance. Nothing can be more preferable than a pictorial demonstration of the calculated residuals to demonstrate this feature (see, for example, Tripathi et al., 2022). We, therefore, obtained the Q-Q plot (see Figure 9-a) between the sample and the theoretical quartiles of the residuals, and a histogram-polygon normal curve plot of the calculated residuals (see Figure 9-b). It is observed that the sample quantiles follow the straight line of the theoretical quantiles and further, the plotted histogram of the residuals are well within the normal polygon curve with the mean equal to zero and standard deviation 0.02 (see Figure 9-b). The above diagnosis of the calculated residuals strengthens our assumption that the residuals of the data, based on the ARIMA (2,2,2) model, are not correlated and follow a normal distribution with specific white noise property.



(a) (b)

Figure 8. *Residual and ACF plots of the coal production data based on the ARIMA (2,2,2) model.*



(a) (b)

Figure 9. *Q-Q plot and histogram-polygon plot of residuals based on the ARIMA (2,2,2) model.*

To conclude our work in the favour of most accurate model among the naive, Holt's model and ARIMA (2,2,2) model, we have performed the two 'forecast accuracy measure' tools, namely; root mean square error (RMSE) (see Eq. 6) and mean absolute percentage error (MAPE) (see Eq. 7) which are defined as below:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}, \qquad \textit{Eq. 6}$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100, \qquad \textit{Eq. 7}$$

where $\hat{y}_i$ is the i^{th} observation to be predicted from the 'test sample data' (see, for example, Tripathi and Agarwal 2023). The corresponding predicted values for the original observations can easily be obtained and they are summarised in Table 5 for each of the considered models. Table 4 summarises the results of the forecast accuracy measures for each of the considered models. This clearly indicates that the ARIMA (2,2,2) model outperforms the other two models for the given data with a better accuracy. Therefore, we recommend ARIMA models to predict the coal production in India as these models admit correlation within the data, which causes a major difference in forecasting when compared with the component-based time series models.

Table 4. *Results of the forecast accuracy measures*

Accuracy measures	Naive model	Holt's model	ARIMA (2,2,2) model
RMSE	84.94	20.18	0.02
MAPE	10.27	2.11	0.3

Based on the recommendation of 'forecast accuracy measure', we have made the future prediction of coal production in India by using the ARIMA (2,2,2) model. With the available data of 42 observations, we have predicted for next 5 years coal production in India, which is shown (by red dotted line) in Figure 10. Certainly, the Graphical pattern of the future prediction is enough to conclude that the coal production is going to be increased in the next five years in India.

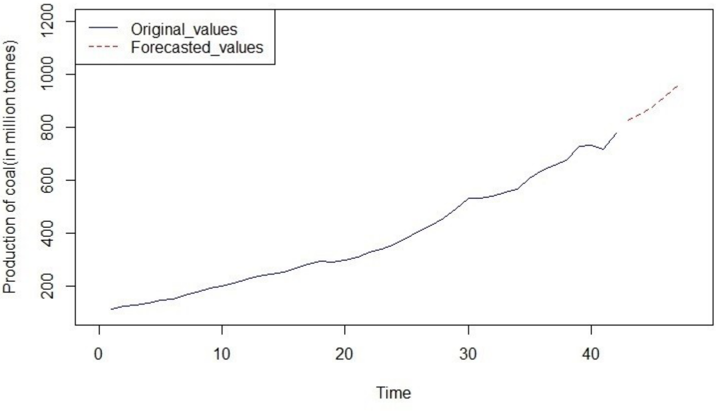


Figure 10. *Forecasts for next 5 years of coal production in India by the ARIMA (2,2,2) model.*

Conclusion and Findings

This paper has successfully analysed the coal production data of India through the time series models. The analyses tackle the component-based and correlation-based time series models for the given data. First, the forecasts were made using naive and Holt linear trend models and, then, the data is being used to analyse the ARIMA model. A suitable ARIMA model is chosen by adopting the two-fold strategy consisting of the Box-Jenkins technique and then on applying the two model selection criteria, namely, AIC and BIC. The two-fold identification of the ARIMA model suggests that ARIMA (2,2,2) is the best candidate model among others. After the successful identification steps, we performed the retrospective forecasts for the competing models (see Table 5). The forecasted values by the ARIMA (2,2,2) model are closer than those by the other component-based models, which can also be verified by their accuracy measures in Table 4. It is, therefore, decided to perform the further analysis with the selected ARIMA (2,2,2) model. After estimating the parameters of the ARIMA (2,2,2) model, the residual diagnostic checks were performed for the validity purpose of the chosen model. Finally, a 'five years' future forecast is being made by using the ARIMA (2,2,2) model graphically. It is observed that the coal in India will be in high demand and its production will increase in the years to come.

Table 5. *Retrospective six-year forecast of coal production in India based on the competing time series models.*

Year	True Value	Forecast value		
		Naive model	Holt's model	ARIMA (2,2,2) model
2016-17	657.8	639.2	664.3	657.6
2017-18	675.8	639.2	688.9	674.8
2018-19	728.7	639.2	713.1	729.1
2019-20	730.8	639.2	736.7	731.0
2020-21	716.9	639.2	759.9	715.8
2021-22	778.2	639.2	782.6	778.5

Our analysis is restricted by the use of time series models and it does not include the other factors that may affect the coal production in India, such as; infrastructure and availability of modern technology, accidental hazards during the production process, natural calamities, etc. One may consider these factors and see their effects on the coal production over the years. Also, a significant analysis can be done to observe the most significant factor and, ultimately, can tune the productivity of energy sources at a desired level whenever necessary. On the theoretical ground, the regression analysis could be another alternative to estimate the coal production with its relevant explanatory variables. Our analysis is limited, but it is not restricted for any future development of its current version. Such an analysis is going to be beneficial for the industry resource planners and government/semi-government entities to a good inventory setup and to design a reasonable path of coal consumption at the ground level.

Availability of data and materials: The data is provided in the manuscript and is open to access.

Ethics Committee Approval: As this study is based entirely on a review of existing literature, rather than primary data collection, ethics committee approval was not required. The analysis and synthesis of previously published research do not involve new interactions with human subjects or require ethical clearance.

Peer-review: Externally peer-reviewed.

Author Contributions: Conception/Design of Study- A.G., D.R.; Data Acquisition- A.G.; Data Analysis/Interpretation- A.G., D.R.; Drafting Manuscript- A.G., D.R.; Critical Revision of Manuscript- P.K.T.; Final Approval and Accountability- P.K.T.; Technical or Material Support- A.G., D.R.; Supervision- P.K.T.

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: The authors declared that this study has received no financial support.

References

Agarwal, M., Tripathi, P. K., & Pareek, S. (2021). Forecasting infant mortality rate of India using ARIMA model: a comparison of Bayesian and classical approaches. *Statistics & Applications*, 19(2), 101–114.

Box, G. & Jenkins, G. (1970). *Time series analysis: forecasting and control*. Holden-Day, San Francisco.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Chen, Y., Xu, J., Tan, K., & Gou, Y. (2021). Statistical analysis and regular research on Chongqing coal mine accidents. *In 2021 International Conference on Intelligent Computing, Automation and Applications (ICAA)*, pages 64–70. IEEE.

Holt, C. C. (1957). Forecasting trends and seasonals by exponentially weighted averages. *ONR Memorandum*, 52, 1957.

Hyndman, R. J. & Athanasopoulos, G. (2018). *Forecasting: principles and practice.* OTexts.

Jai Sankar, T., Angel Agnes Mary, I., & Nalini, K. (2023). Stochastic time series modeling for coking coal production in India. *International Journal of Scientific Development and Research*, 7(10), 443–449.

Li, S., Yang, X., & Li, R. (2019). Forecasting coal consumption in India by 2030: using linear modified linear (MGM-ARIMA) and linear modified nonlinear (BP-ARIMA) combined models. *Sustainability*, 11(3), 695.

Makkhan, S. J. S., Parmar, K. S., Kaushal, S., & Soni, K. (2020). Correlation and time-series analysis of black carbon in the coal mine regions of India: a case study. *Modeling Earth Systems and Environment*, 6, 659–669.

Mohanty, A. & Nimaje, D. S. (2023). Time series forecasting of Indian coal mines fatal accidents. *Journal of Mining Science*, 59(6), 1076–1082.

Montrone, L., Ohlendorf, N., & Chandra, R. (2021). The political economy of coal in India evidence from expert interviews. *Energy for Sustainable Development*, 61, 230–240.

Parren˜o, S. J. E. (2022). Application of time series analysis in forecasting coal production and consumption in the Philippines. *ICTACT Journal on Soft Computing*, 13(1).

Tripathi, P. K. & Agarwal, M. (2023). A bayes analysis of random walk model under different error assumptions. *Annals of Data Science*, pages 1–18.

Tripathi, P. K., Mishra, R. K., & Upadhyay, S. (2018). Bayes and classical prediction of total fertility rate of India using autoregressive integrated moving average model. *Journal of Statistics Applications & Probability*, 7(2), 233–244.

Tripathi, P. K., Ranjan, R., Pant, R., & Upadhyay, S. (2017). An approximate Bayes analysis of ARMA model for Indian GDP growth rate data. *Journal of Statistics and Management Systems*, 20(3), 399–419.

Tripathi, P. K., Sen, R., & Upadhyay, S. (2021). A Bayes algorithm for model compatibility and comparison of ARMA (p, q) models. *Statistics in Transition new series*, 22(2), 95–123.

Tripathi, P. K., Tripathi, A., Harshika, & Meenakshi (2022). Statistical insights on the health of Indian economy: a multivariate analysis. *Journal of Data, Information and Management*, 4(3), 305–327.

MOSPI (2024). Energy Statistics India 2024. https://mospi.gov.in/sites/default/files/publication_reports/Energy_Statistics_2024/Energy_Statistics_India2024.pdf

India (2024). Statistical Appendix. *Economic Survey 2023-24.* https://www.indiabudget.gov.in/economicsurvey/doc/Statistical-Appendix-in-English.pdf

Appendix

Coal production (in million tonnes) in India approximately 1980-81 to 2021-22 (see India 2024).

Year	**Production**	**Year**	**Production**	**Year**	**Production**
1980-81	113.9	1994-95	253.8	2008-09	457.1
1981-82	124.2	1995-96	270.1	2009-10	492.8
1982-83	130.5	1996-97	285.7	2010-11	535.7
1983-84	138.2	1997-98	295.7	2011-12	540.0
1984-85	147.4	1998-99	292.3	2012-13	556.4
1985-86	154.2	1999-00	300.0	2013-14	565.8
1986-87	165.8	2000-01	309.6	2014-15	609.2
1987-88	179.7	2001-02	309.6	2015-16	639.2
1988-89	194.6	2002-03	327.8	2016-17	657.8
1989-90	200.9	2003-04	341.3	2017-18	675.8
1990-91	211.7	2004-05	361.3	2018-19	728.7
1991-92	229.3	2005-06	382.6	2019-20	730.8
1992-93	238.3	2006-07	407.0	2020-21	716.08
1993-94	246.0	2007-08	430.8	2021-22	778.19

Journal of Data Applications
Issue: 3, 2024
E-ISSN: 2980-3357

RESEARCH ARTICLE

Measuring Artificial Intelligence Integration in Higher Education: A Bibliometric Analysis of Quantitative Studies

Hatice ÇİFÇİ[1], Mehmet Altuğ ŞAHİN[2], İbrahim ÇİFÇİ[2], Gürel ÇETİN[2]

ABSTRACT
This study explores the current state of Artificial Intelligence (AI) adoption in higher education, evaluating its scope via bibliometric methods. The research builds upon the knowledge acquired from quantitative studies and establishes guidance for future studies. A total of 24 publications from the combined database of Scupos and Web of Science (WOS) were collected and used as the resource for the bibliometric analysis. The bibliometric analysis using Biblioshiny identified seven indicators, including annual publications, the top 10 contributing countries, the most relevant sources, a thematic map, motor and niche themes, emerging or declining themes, and basic themes. In addition, for the keyword analysis, the authors used the VOSviewer, which identified three clusters: pedagogy, AI tools, and ethics. As a result, the paper provides an improved understanding of AI adoption in education and a framework that includes both students' and educators' perspectives on the measures and quantitative research in AI utilization in education. Such knowledge not only provides significant information on the current state of literature and trends but also implications for educators, administrators, and educational technology (EduTech) suppliers.

Keywords: AI Adoption, Education, Students, Educators, Bibliometric Analysis

DOI: 10.26650/JODA.1536942

[1]İstanbul University, Institute of Social Sciences, Department of Tourism Management, İstanbul, Türkiye
[2]İstanbul University, Faculty of Economics, Department of Tourism Management, İstanbul, Türkiye

ORCID: H.Ç. 0000-0002-9761-226X;
M.A.Ş. 0000-0003-1048-1963;
İ.Ç. 0000-0001-7469-1906;
G.Ç. 0000-0003-3568-6527

Corresponding author:
Hatice ÇİFÇİ,
İstanbul University, Institute of Social Sciences, Department of Tourism Management, İstanbul, Türkiye
E-mail: hatice.cifci@outlook.com

Submitted: 21.08.2024
Revision Requested: 30.09.2024
Last Revision Received: 15.11.2024
Accepted: 23.11.2024
Published Online: 24.01.2025

Citation: Çifçi, H., Şahin, M. A., Çifçi, İ., & Çetin, G. (2024). Measuring Artificial Intelligence Integration in Higher Education: A Bibliometric Analysis of Quantitative Studies. *Journal of Data Applications, 3*, 33-62. https://doi.org/10.26650/JODA.1536942

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International License

Introduction

Artificial intelligence (AI) promises a great potential to transform education, with the advancement of various tools, particularly in the higher education area (Huang & Rust, 2018), including ChatGPT, Canva, Education Copilot, Grammerly, and Quilboot (EDUCAUSE, 2019). However, research indicates that AI adoption in educational institutions remains tentative (McGrath et al., 2023; Perrault & Clark, 2024), despite its substantial potential to revolutionise higher education compared to other technological breakthroughs (Bates et al., 2020). AI can benefit education in various ways, for instance, by enhancing learning analytics systems (Cerratto Pargman & McGrath, 2019), providing accurate and expeditious results, eliminating bureaucracy through algorithmic systems (Burrows et al., 2015), and improving effectiveness and outcomes in education and research (Klutka et al., 2018).

Hence, some authors (e.g., Zawacki-Richter et al., 2019) highlight the importance of deepening the insight into AI's effectiveness in higher educational settings, making this phenomenon interesting to explore (McGrath et al., 2023). However, there is still no consensus on AI integration in higher education (Molenaar, 2022). Various challenges have been suggested, shaping both non-academic and academic discourse mainly through conceptual and qualitative approaches. These include AI illiteracy (Luckin et al., 2022; Laupichler et al., 2022), fear of job losses (Akata et al., 2019), resistance to change and tendency to avoid the risk (Bearman et al., 2023), risk of biases and prejudice within data and regarding learning analytics (Mittelstadt et al., 2016), and limited funding for alternatives to traditional teaching methods (Wheeler, 2019).

A large volume of prior work on AI in higher education has paid more attention to technological implications through systematic reviews (Bearman et al., 2023; Bond et al., 2024; Laupichler et al., 2022), and in some cases through qualitative approaches (Al-Mughairi & Bhaskar, 2024). This leaves room for future investigation to better understand and reveal the factors influencing AI adoption in higher education practises (Buckingham Shum et al., 2019). Although these papers provide useful insights and knowledge on the current state of AI in education, none, to the best of our knowledge, sufficiently address the adoption challenges in foundational research, which questions why AI has not yet become a revolutionary element in education (Dhawan & Batra, 2020). Consequently, based on this perspective, it is relevant to guide AI implementation by analysing quantitative research that provides measurements on the actual adoption of AI in education.

This research addresses the need for further research on the potential of AI in educational environments while underscoring the importance of a close and more nuanced examination of the adoption within the educational practise that pertains to AI learning through quantitative measures and scales. As a strictly bibliometric study, this research, therefore, seeks to offer pertinent guidance by reviewing quantitative studies on the optimal integration of AI into learning environments, ensuring the effective transfer of knowledge from AI as a theoretical concept to its practical application in higher education. As such, this research contributes to the development of more sophisticated AI models for education and provides insights into possible avenues for further research.

Literature Review

The higher education domain is growing at an unprecedented pace and requires theoretical development to advance the existing knowledge of AI from the stakeholders' point of view (Bond et al., 2024; Crompton & Burke, 2023). Since AI applications may not always be compatible with teaching and learning processes and goals, it becomes pertinent to identify educational contexts that can integrate AI in a manner that makes it easy for educators, students, and other stakeholders in education to use it for their intended pedagogical purposes.

Based on these considerations, the AI application in education can be organised in terms of beneficiaries divided into (a) student-centric AI applications and (b) educator-centric AI applications (Baker et al., 2019). However, comparatively, scholars have paid limited or modest efforts in researching this phenomenon, although evidence has indicated that there is much discussion about it (Dhawan & Batra, 2020). When the use and incorporation of AI-enabled technologies in educational contexts are considered, they should not simply be seen as matching with technology or design ideas or otherwise meeting the integration requirements set out by the formal technological frameworks. However, educators' and students' needs should also be considered when these technologies are to be incorporated into educational programmes (Luckin et al., 2016). In this regard, the present research outlines a broad systematic review of quantitative research involving students' and educators' AI adoption to guide future studies and technologies. The paper first explores the quantitative literature involving students and then focuses on research on AI adoption by educators as participants.

For example, Bisdas et al. (2021) discussed the topic from the student perspective and evaluated the attitude of medical and dental students about AI integration in their education, finding that students have positive attitudes towards incorporating AI into their training. Abdelwahab et al. (2023) searched business students' perceptions about AI integration and revealed the importance of curriculum and educational facilities updates for students' integration due to a lack of adequate knowledge of the increasingly AI-integrated work environment. Yuk Chan and Tsi's (2023) study regarding the AI integration of teachers that teachers can integrate AI to enhance teaching without replacing them. Mohd Rahim et al.'s (2022) study revealed that perceived trust is an important predictor of students' adoption of AI.

Chan and Hu's (2023) study revealed the students' main concerns about using AI as accuracy, privacy, and ethics, as well as the potential influences on personal growth, career opportunities, and societal norms. Li (2023) observed that the perceived usefulness (PU) and perceived ease of use (PEU) of AI-based systems had positive impacts on attitudes, behavioural intentions, and practical applications among students.

In contrast, college students' sentiments towards AI-based systems had no substantial influence on their learning motives to reach goals or subjective standards. However, the study of Bilquise et al. (2024) demonstrated that PU, autonomy, and trust did not significantly influence the acceptance of an advising chatbot. Foroughi et al. (2023) identified that factors such as performance expectancy, effort expectancy, hedonic motivation, and perceived learning value exert a significant impact on individuals' intentions to adopt the generative AI ChatGPT. Alhumaid et al. (2023) found that perceived compatibility, trialability, the perceived advantage, and ease of doing influence students' AI adoptions. Strzelecki (2023) found that habit is the most significant predictor of students' behavioural intention to adopt AI. Delcker et al. (2024) study revealed that students' attitudes towards AI are influenced by the perceived benefits of AI technology. Salloum et al. (2024) also revealed that students' willingness to adopt AI chatbots is affected by perceived usefulness, ease of use, and flow experience. Dahri et al. (2024) showed that the increased use of AI tools enhanced student satisfaction and significantly influenced learning outcomes. However, students' engagement and personal innovativeness did not play a significant role in affecting AI tool adoption. Table 1 showcases the summary of these studies.

Despite the importance of educators in the integration of AI-based technologies (Çelik, 2023; Seufert et al., 2021; Wang et al., 2024), there is a shred of limited empirical evidence explaining how educators use AI technologies in the higher education context, highlighting a gap in research on exploring educators' viewpoints on AI-based instruction (Çelik, 2023). Table *2* illustrates a summary of some of the recent quantitative studies related to AI adoption in higher education from educators' perspectives by highlighting the variables, theories, models, and frameworks involved.

Even though some valuable individual studies might not fully capture all the factors influencing the integration of AI for academia; Chatterjee and Bhattacharjee (2020) identified the perspective of stakeholders (i.e., teachers, students, administrative staff) in adopting AI into higher education. Wang et al. (2021) examined teachers' intention to adopt AI tools in their classrooms in higher education settings through the Technology Acceptance Model (TAM) with four additional dimensions, including anxiety, self-efficacy, attitude towards AI, and behavioural intention. An et al. (2023) studied the behavioural intentions of English teachers regarding the use of AI for teaching. Zhang et al. (2023) identified the factors for determining pre-service teachers' intentions to use AI. Wang et al. (2023) conceptualised teachers' AI readiness through four components—cognition, ability, vision, and ethics—and also explored their interrelationships and implications on teachers' professional practise. Çelik (2023) developed an intelligent Technological Pedagogical Content Knowledge (TPACK) framework by extending it to an ethical aspect. In the work of Shwedeh et al. (2024), the moderated model of the perceived relationship regarding the adoption of the AI system trust in data privacy and security, learning personally and professionally, stakeholders' needs, and policy and regulations on educational sustainability were examined using the socio-technical systems theory.

Ning et al. (2024) developed a five-item Al-TPACK scale based on the assumption of the interactional and combined consequences of AI technology, pedagogy, and subject matter in educational contexts. Wang et al. (2024) analysed pre-service teachers' perspectives to integrate AI usage and found that anxiety, social influence, and performance expectancy strongly predicted behavioural intention rather than effort expectancy and facilitating conditions. Lastly, Jain and Raghuram (2024) examined the TAM and TPACK models with an additional dimension of perceived trust towards the adoption of Gen-AI by Indian higher education institution members. Hence, a variety of research provides fruitful insights and knowledge into the existing literature on AI adoption in higher education, encompassing different theories, models, and contexts. It strengthens the current work's rationale by addressing areas needing further exploration in understanding AI-based deployment in academia, with prominent studies shown in Table 2.

Table 1. *Summary of Recent Quantitative Studies on AI Adoption in Higher Education Based on Students' Perspectives.*

Author/s	Database	Objective	Constructs Used	Model/ Theory	Key Findings
Bisdas et al. (2021)	WoS and Scopus	Measuring attitudes towards AI integration during the education phases of medical and dental students.	Attitudes and feelings (8 items)	N/A	The research established that students possessed a basic concept and a favourable attitude regarding the integration of AI into their learning.
Abdelwahab et al. (2023)	WoS	Understanding business students' perception of how higher education institutions prepare them for workplaces with AI integration.	Awareness (2 items) Teaching facilities (1 item) Programme/curricula (3 items) Teaching of AI skills (2 items)	The Quality Indicator Model	The findings indicate that students cannot integrate AI due to higher education institutions (HEIs) insufficient infrastructure and opportunities for AI.

Table 1. *Summary of Recent Quantitative Studies on AI Adoption in Higher Education Based on Students' Perspectives.*

Author/s	**Database**	**Objective**	**Constructs Used**	**Model/ Theory**	**Key Findings**
Chan and Hu (2023)	WoS and Scopus	Measuring the perceptions of university students on generative AI Technologies.	Knowledge of generative AI Technologies (6 items) Willingness to use (8 items) Concerns (4 items)	John Biggs' 3P Model	The students recognised the opportunities in individual instructional facilitation, writing and idea generation instruments, and research and analysis instruments. However, there were certain doubts and questions associated with the topics referring to accuracy, privacy, ethical questions, and the impact on the individual's or society's development, job opportunities, perspectives, and norms.
Delcker et al. (2024)	WoS and Scopus	Exploring the perceptions and expectations of first-year students on AI tools integrating the DigiComp2.2 framework.	Skills (4 items) Knowledge (6 items) Attitudes (5 items)	The Unified Theory of Acceptance and Use of Technology (UTAUT) Model	This shows that first-year students' attitudes towards AI are the main contributors to the intended use of AI tools. Furthermore, the perceived benefits of AI technology are antecedent variables for the perceived suitability of AI robots to substitute humans as cooperation partners.
Salloum et al. (2024)	Scopus	Measuring students' perceptions of adopting AI across various educational institutions.	User satisfaction (3 items) PU (3 items) PEU (3 items) Flow Experience (2 items) Adoption of Chatbots (2 items)	The Technology Acceptance Model (TAM), Flow Theory	All predictors positively affected the students' intention to adopt AI chatbots.

Table 1. *Summary of Recent Quantitative Studies on AI Adoption in Higher Education Based on Students' Perspectives.*

Author/s	**Database**	**Objective**	**Constructs Used**	**Model/ Theory**	**Key Findings**
Mohd Rahim et al. (2022)	WoS and Scopus	Identifying factors that influence the effectiveness of chatbot adoption in the HEI context.	Performance expectancy (5 items) Effort expectancy (5 items) Social influence (5 items) Facilitating conditions (5 items) Hedonic motivation (3 items) Habit (3 items) Interactivity (5 items) Design (5 items) Ethics (4 items) Perceived trust (4 items) Behavioural intention (3 items) Use intention (4 items)	UTAUT2, Information Systems (IS) Theory	Perceived trust was significantly impacted by interactivity, design, and ethics. Moreover, the results revealed that perceived trust, performance expectancy, and habit towards the use of chatbots had a significant impact on behavioural intention.

Table 1. *Summary of Recent Quantitative Studies on AI Adoption in Higher Education Based on Students' Perspectives.*

Author/s	**Database**	**Objective**	**Constructs Used**	**Model/ Theory**	**Key Findings**
Dahri et al. (2024)	WoS and Scopus	Measuring students' intention to adopt AI tools in higher education institutions.	Performance Expectancy (5 items) Facilitating Conditions (5 items) Students' Engagement (4 items) Assessment Effectiveness (4 items) Students' Interaction (5 items) Information Accuracy (5 items) Personal Innovations (6 items) Pedagogical Fit (5 items) AI Tools Use (4 items) Behavioural Intentions (3 items) Student Satisfaction (4 items) Improve students' Academic Performance (4 items)	UTAUT	Performance and effort expectancy, AI tool information accuracy, pedagogical fit, and student interaction played a significant role in the acceptability and usage of AI tools in higher education qualifications.
Foroughi et al. (2023)	WoS and Scopus	Investigating the determinants of the intention to use ChatGPT for educational purposes.	Performance Expectancy (6 items) Effort Expectancy (4 items) Social Influence (3 items) Facilitating Conditions (4 items) Hedonic Motivation (3 items) Learning Value (4 items) Habit (3 items) Personal Innovativeness (3 items) Information Accuracy (3 items) Intention to Use (3 items)	UTAUT2	Performance and effort expectancy, hedonic motivation, and learning value influenced the willingness to use ChatGPT, while personal innovativeness and information accuracy negatively moderated the relationships between ChatGPT use and its determinants.

Table 1. *Summary of Recent Quantitative Studies on AI Adoption in Higher Education Based on Students' Perspectives.*

Author/s	**Database**	**Objective**	**Constructs Used**	**Model/ Theory**	**Key Findings**
Alhumaid et al. (2023)	Scopus	Measuring students' perceptions of using AI for educational purposes in the UAE.	Artificial Intelligence Application Adoption (2 items) Perceived Compatibility (3 items) Trialability (3 items) The relative advantage Ease of Doing Business (3 items) Technology Export (3 items)	Diffusion Theory	The results show that the diffusion theory variables have a greater impact compared to the ease of doing business and technology export variables.
Strzelecki (2023)	WoS and Scopus	Developing a model examining the predictors influencing the adoption and use of ChatGPT among students in higher education.	Performance expectancy (4 items) Effort expectancy (4 items) Social influence (3 items) Facilitating conditions (4 items) Hedonic motivation (3 items) Habit (4 items) Behavioural Intention (3 items) Personal innovativeness (4 items) Use Behaviour (1 item)	UTAUT2 (Extended UTAUT Model)	Habit emerged as the best predictor of behavioural intention, with performance expectancy and hedonic motivation following as the most significant predictors. Behavioural intention, followed by personal innovativeness, stood out as the most dominant determinant of use behaviour.

Table 1. *Summary of Recent Quantitative Studies on AI Adoption in Higher Education Based on Students' Perspectives.*

Author/s	**Database**	**Objective**	**Constructs Used**	**Model/ Theory**	**Key Findings**
Li (2023)	WoS and Scopus	Investigating the factors influencing the college students' engagement with AI-based systems and examining the role of learning motivations.	PEU (5 items) PU (5 items) Attitude (4 items) Learning motivation-Learning interest (5 items) Learning motivation-Achieving goal (4 items) Learning motivation-Subjective norm (6 items) Behavioural intention (4 items) Actual use (4 items)	TAM	PU and the perceived ease of use of AI-based systems had positive effects on students' attitudes, behavioural intentions, and engagement with AI-based systems. However, college students' attitudes towards AI-based systems had no significant influence on their learning motivation related to the achievement of their goals and subjective norms.
Bilquise et al. (2024)	WoS and Scopus	Identifying antecedents of behavioural intention in university students' use of an academic advising chatbot.	PEU (4 items) PU (4 items) Perceived Autonomy (5 items) Perceived Trust (5 items) Anthropomorphism (5 items) Social Influence (4 items) Behavioural Intention to Adopt (3 items)	TAM, UTAUT, The Service Robot Acceptance (sRAM) Model, The Self-Determination Theory (SDT) Model	The analysis of the results obtained shows the influence of functional elements, PEU, and social influence on the behavioural intention to accept the chatbots.

Table 2. *Summary of Recent Quantitative Studies on AI Adoption in Higher Education Based on Educators' Perspectives*

Author/s	**Database**	**Objective**	**Constructs Used**	**Model/ Theory**	**Key Findings**
Wang et al. (2021)	WoS and Scopus	Measuring teachers' intention to adopt AI tools in their classes in higher education settings.	Anxiety (4 items) Self-efficacy (2 items) Attitude towards AI (2 items) PEU (2 items) PU (3 items) Behavioural Intention (4 items)	TAM	Perceived predisposing factors with respect to the adoption of AI-based applications by the teachers included attitudes towards use (ATU), PEU, PU, subjective norms (SE), and actual use (AN). SE had a positive impact on both PEU and ATU, which paved the way for adopting AI. In addition, strengthening SE diminished teachers' resistance (AN) towards adopting AI in teaching.
Yang et al. (2021)	WoS and Scopus	Examining the acceptance of the e-Schoolbag technology by K-12 teachers. The main objective of this study is to determine how teachers' Technological Pedagogical Content Knowledge (TPACK) abilities influence their inclination to use the e-Schoolbag.	Technological knowledge (TK) (3 items) Pedagogical knowledge (PK) (3 items) Content knowledge (CK) (3 items) Pedagogical content knowledge (PCK) (3 items) Technological content knowledge (TCK) (3 items) Technological pedagogy knowledge (TPK) (3 items) Technological pedagogical content knowledge (TPACK) (3 items) PU (6 items) PEU (6 items)	TAM, The Technological Pedagogical and Content Knowledge (TPACK) Model	TPACK significantly enhanced EOU and positively impacted the PU of e-Schoolbag applications, although its impact on PU scored comparatively lower. TK, PK, and CK did not have a direct effect on TPACK, while TPK and TCK directly contributed to TPACK. TK contributed significantly to both TPK and TCK, whereas PK affected TPK and PCK. CK notably influenced PCK.

Table 2. *Summary of Recent Quantitative Studies on AI Adoption in Higher Education Based on Educators' Perspectives*

Author/s	Database	Objective	Constructs Used	Model/ Theory	Key Findings
Wang et al. (2023)	WoS and Scopus	Exploring AI readiness in four dimensions.	Cognition (5 items) Ability (6 items) Vision (3 items) Ethics (4 items)	N/A	It established a connection between teachers' ability in the application of AI and ethicality in education. Technical proficiency, visionary thinking, and ethical awareness were associated with higher levels of AI adoption by teachers. Fear of AI hampered educational innovation, while its adoption enhanced teachers' job satisfaction. The teacher cluster based on AI readiness implied that the level of innovation as well as the level of job satisfaction tended to be high, and this aspect was not affected by the socio-economic status and gender of the teacher.
Shwedeh et al. (2024)	Scopus	Analysing how the effects of AI adoption, trust (measured with data privacy and security), stakeholders' needs, policy, and regulations act as moderators in the perceived relationship for education sustainability.	Educational sustainability (7 items) Trust (8 items) AI adoption (9 items) Policies and regulations (7 items)	Socio-Technical Systems Theory	AI adoption positively affected educational sustainability. Policies and regulations did not affect educational sustainability. Trust positively affected educational sustainability. While policies and regulations moderated AI adoption and educational sustainability, they did not moderate AI adoption and trust.

Table 2. *Summary of Recent Quantitative Studies on AI Adoption in Higher Education Based on Educators' Perspectives*

Author/s	**Database**	**Objective**	**Constructs Used**	**Model/ Theory**	**Key Findings**
Chatterjee and Bhattacharjee (2020)	WoS and Scopus	Exploring how the stakeholders (i.e., teachers, students, administrative staff) adopt AI in higher education settings.	Perceived Risk (4 items) Performance Expectancy (5 items) Effort Expectancy (5 items) Facilitating Conditions (5 items) Attitude (5 items) Behavioural Intention (5 items) Adoption of AI in Higher Education (4 items)	UTAUT	By employing the UTAUT model, the research developed and empirically tested hypotheses to demonstrate the applicability of the model to encourage the use of AI among the stakeholders. Therefore, it suggested that the use of AI in the Indian higher education sector can enhance the governance and decision-making processes.
An et al. (2023)	WoS and Scopus	Exploring the integration of AI to enhance English as a Foreign Language (EFL) teachers' practises by investigating their perceptions, knowledge, and behavioural intentions in a K-12 setting.	Performance Expectancy (4 items) Effort Expectancy (4 items) Facilitating Conditions (4 items) Social Influence (3 items) AIL-TK (3 items) AI-TPK (7 items) AI-TPACK (10 items) Behavioural Intention (4 items)	UTAUT, TPACK	While performance expectancy, social influence, AI language technological knowledge, and AI-TPACK had significant positive predictive power on behavioural intention; effort expectancy, facilitating conditions, and AI-based pedagogical knowledge showed indirect effects on it.

Table 2. *Summary of Recent Quantitative Studies on AI Adoption in Higher Education Based on Educators' Perspectives*

Author/s	Database	Objective	Constructs Used	Model/ Theory	Key Findings
Zhang et al. (2023)	WoS and Scopus	Identifying factors influencing pre-service teachers' behavioural intentions to use AI-based educational applications. It also explores gender differences in the proposed model.	PU (3 items) PEU (4 items) AI self-efficiency (4 items) AI Anxiety (3 items) Perceived enjoyment (3 items) Subjective norms (2 items) Job relevance (3 items) Behavioural intention (2 items)	TAM3	The research revealed that the determinants influencing behavioural intention were based on the TAM3 model in using AI-driven educational applications and highlighted the importance of addressing gender-specific elements in teacher education.
Çelik (2023)	WoS and Scopus	Developing a scale to measure the knowledge of teachers for using AI tools in instructional settings and extending TPACK components to include ethical considerations.	Intelligent–TPACK Scale (5 dimensions with 27 items)	TPACK	It proposed an Intelligent-TPACK framework with an improved scale.
Sun et al. (2024)	WoS and Scopus	Exploring teachers' intention to integrate AI-based Teaching methods based on STEM educators.	TPACK (4 items) PU (4 items) PEU (4 items) Self-efficacy (4 items) Willingness to integrate AI (4 items)	TAM TPACK	A direct influence on WIAI was directed by TPACK, PU, PE, and SE. In addition, TPACK had a direct effect on PE, PU, and SE, while PE and PU influenced SE directly. The mediating roles of PE, PU, and SE were discovered in the relationship between TPACK and the willingness to integrate AI (WIAI).

Table 2. *Summary of Recent Quantitative Studies on AI Adoption in Higher Education Based on Educators' Perspectives*

Author/s	Database	Objective	Constructs Used	Model/ Theory	Key Findings
Ning et al. (2024)	WoS and Scopus	Developing and validating an AI-TPACK measurement tool for teachers and exploring the interrelationships among its components to ensure alignment with theoretical assumptions.	AI-TK (5 items) AI-TCK (6 items) AI-TPK (6 items) AI-TPACK (5 items)	AI-TPACK	The developed framework functions as a comprehensive guide for the extensive evaluation of teachers' AI-TPACK, and a sophisticated grasp of how different AI-TPACK components interact leads to a more profound explanation of the generative mechanisms that underpin teachers' AI-TPACK.
Wang et al. (2024)	WoS and Scopus	Analysing pre-service teachers' perspectives regarding the adoption of generative AI into their Teaching practises.	Performance expectancy (4 items) Effort expectancy (4 items) Social influence (3 items) Facilitating conditions (4 items) GenAI Anxiety (4 items) Technology Self-Efficiency (4 items) GenAI TPACK (4 items)	UTAUT, TPACK	Generative Artificial Intelligence (GenAI) anxiety, social influence, and performance expectancy strongly predicted teachers' behavioural intentions. Effort expectancy and facilitating conditions have no impact on influencing their intentions.
Jain and Raghuram (2024)	WoS and Scopus	Examining the relationships among TAM, TPACK, and trust as predictors and their combined impact on the adoption of Gen-AI among undergraduate and postgraduate students and faculty members.	PEU (3 items) PU (3 items) TPACK (3 items) Trust (3 items)	TAM, TPACK	The study revealed that the relationships between age, gender, and the intention to adopt AI in higher education settings were non-compensatory and nonlinear.

Methodology

The search was conducted using the databases of the Web of Science (WoS) and Scopus, designed to locate articles related to the acceptance and integration of AI in higher education. The search string used for WoS is as follows: ("ARTIFICIAL INTELLIGENCE" OR "AI" OR "GENERATIVE AI" OR "GENERATIVE ARTIFICIAL INTELLIGENCE" (Topic) and "EDUCATION*" OR "HIGHER EDUCATION*" (Topic) and "TECHNOLOGY ACCEPTANCE" OR "ACCEPT*" OR "INTEGRATED*" OR "ADOPT*" OR "PERCEP*" OR "TOOL*" OR "CHATGPT*" OR "CHATBOT*" OR "TAM*" OR "UTAUT*" OR "*TPACK*" (Topic) and "STUDENT*" OR "TEACHER*" (Topic) and "QUESTIONNAIRE*" OR "SURVEY*" (Topic). The search string used for Scopus is as follows: (TITLE-ABS-KEY ("artificial intelligence" OR "ai" OR "generative ai" OR "generative artificial intelligence") AND TITLE-ABS-KEY ("education*" OR "higher education*") AND TITLE-ABS-KEY ("technology acceptance" OR "accept*" OR "integrate*" OR "adopt*" OR "percept*" OR "tool*" OR "ChatGPT*" OR "chatbot*" OR "tam*" OR "utaut*" OR "*tpack*") AND TITLE-ABS-KEY ("student*" OR "teacher*") AND TITLE-ABS-KEY ("questionnaire*" OR "survey*") AND LANGUAGE (English)) AND PUBYEAR > 2019 AND (LIMIT-TO (DOCTYPE , "ch") OR LIMIT-TO (DOCTYPE , "ar")).

The generally preferred approach for bibliometric analysis is to use databases from either WoS or Scopus or to analyse each database separately due to the challenges linked with the integration process (Echchakoui, 2020). However, in this study, a procedure that merges data from the WoS and Scopus databases was adopted to provide a more comprehensive analysis. Figure 1 illustrates the research process flowchart following the PRISMA 2020 guidelines introduced by Page et al. (2021), which were adopted from the health and medical sciences into the field of tourism (Husamoglu et al., 2024). On July 05, 2024, a bibliometric analysis was conducted using data obtained from predefined databases (See Figure 1).

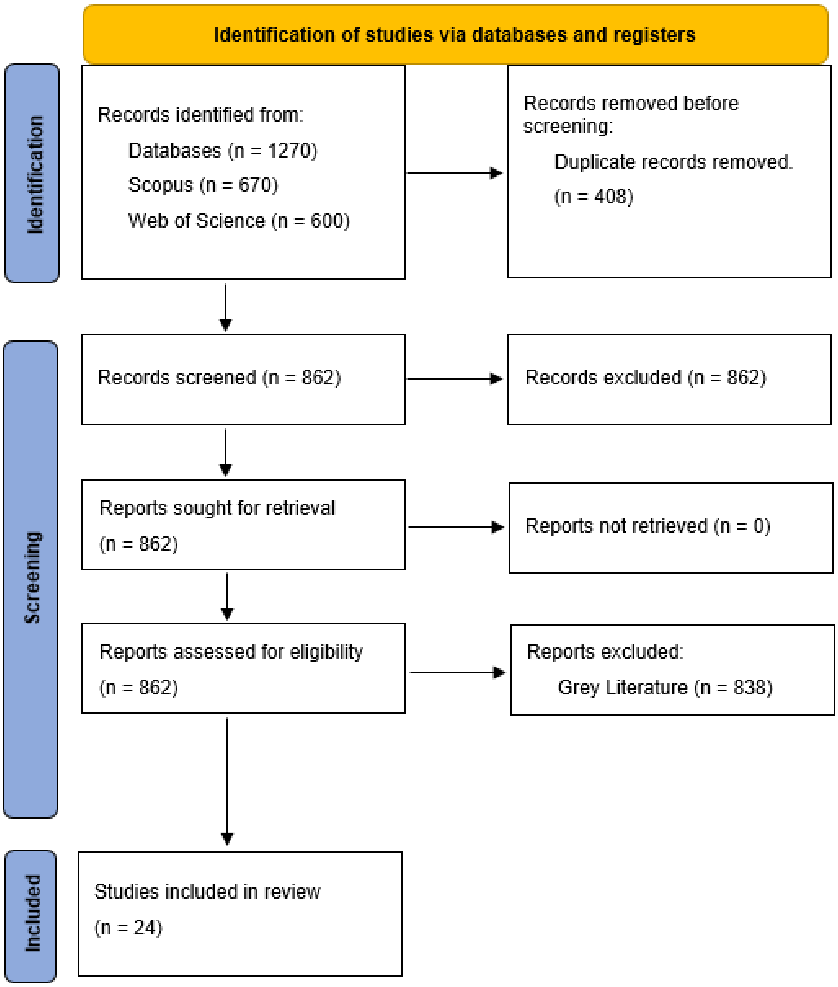


Figure 1. *PRISMA 2020 flowchart.*

A combined 24 publications, 21 from WoS and 24 from Scopus, were identified in this study. The bibliometric analysis does not consider the grey literature. The combined data includes two book chapters and 22 articles published since 2020. Productive authors, journals, countries, the most cited studies, annual publications, and thematic maps were analysed using the biblioshiny package in RStudio (2024.04.2+764), and the keyword analysis of the authors used the VOSviewer (1.6.20) software.

Table 3. *Key information on the combined data.*

Category	Details
DATA OVERVIEW	
Period	2020:2024
Source Types (Journals, Books, etc.)	13
Total Publications	24
Annual Growth Rate %	73,21
Average Document Age	1,04
Average citations per doc	26,04
CONTENT DETAILS	
Author's Keywords (DE)	100
AUTHORS	
Total Authors	123
Single-Authored Documents	3
AUTHORS COLLABORATION	
Single-Authored Documents	3
Average Co-Authors per Document	5,29
International Co-Authorships (%)	25
DOCUMENT TYPES	
Articles	17
Early Access Articles	5
Book Chapter	2

Table 3 shows a detailed summary of the main findings from the publications. The dataset, spanning from 2020 to 2024, comprises 24 documents sourced from 13 different journals and books, with an impressive annual growth rate of 73.21%. Due to the relatively novel nature of AI integration studies in quantitative research, which commenced in 2020, the research universe for this study encompasses works published from 2020 to the present year, 2024, during which the VOSviewer analysis has identified a total of 24 publications based on the specified search criteria. Consequently, the research universe for this study encompasses works published from 2020 to this year, 2024. The average age of the documents is 1.04 years, and each document has garnered an average of 26.04 citations (Sjöstedt et al., 2015, p. 6). In terms of content, the documents feature 100 distinct author keywords. The research involved 123 authors, with only three producing single-authored works. Collaboration is evident, with an average of 5.29 co-authors per document and 25% of the publications involving international co-authorship. The types of documents included 17 articles, five early-access articles, and two book chapters, highlighting a diverse range of scholarly outputs.

Results

Annual Publications

Between 2020 and July 6, 2024, 24 publications were produced. This includes 1 publication in 2020, 3 in 2021, 1 in 2022, and 10 in 2023. As of July 6, 2024, 9 publications have been recorded.

Top 10 Contributing Countries

Bibliometric analysis identified the top contributing countries to quantitative scientific research on AI integration in education based on country scientific production through the corresponding author's affiliation, aligning frequencies with the total article count. Therefore, in bibliometric analysis, the total frequencies of Country Scientific Production may be greater than the total documentation because each author is counted for each affiliation in an article, even if there are co-authors from other countries. On the other hand, the "Corresponding Author's Country" that assigns each article to a single country according to the affiliation of the corresponding author shows comparatively higher frequencies that are closer to the total unique word count. It also determines the Multiple Country Publications (MCP) index to estimate international cooperation through the identification of the articles having authors from different countries.

According to our findings, China led with the highest number of publications (39), demonstrating its intensive focus on advancing research and development. The United Kingdom followed with 10 publications, Malaysia ranked third with 9, Germany contributed 5, India had 4, and Pakistan contributed 3 publications. Several countries, including Australia, Finland, Jordan, Libya, the Netherlands, Poland, Singapore, Switzerland, and the United Arab Emirates, each produced 2 publications, indicating active participation in global scientific endeavours. Lastly, Ecuador, Egypt, Indonesia, Lithuania, Romania, Russia, Saudi Arabia, and Sweden each contributed 1 publication, reflecting a diverse geographic representation in scientific research (Please see Figure 2). The colour in Figure 2 represents the density of publications within specific geographical regions. This analysis underscores the collaborative and international nature of contemporary scientific endeavours, with substantial contributions from countries across various continents, highlighting a widespread commitment to advancing knowledge and addressing global challenges through scientific research.

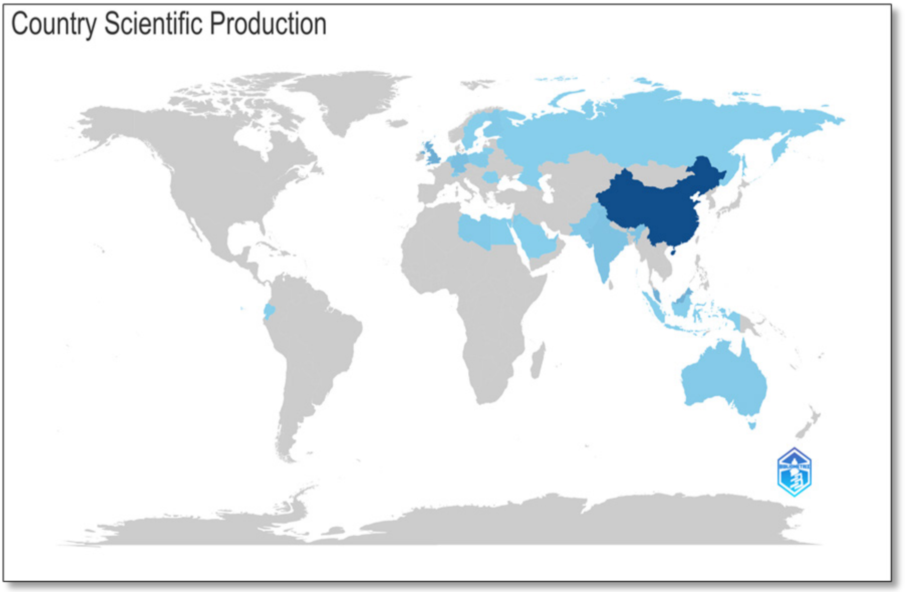


Figure 2. *Top 10 countries.*

Most Relevant Sources

The analysis of the most relevant sources in the dataset highlights the key journals and publications contributing to the research field (Figure 3). The journal "Education and Information Technologies" leads with the highest number of articles, totalling 5 publications. Following this, the "International Journal of Educational Technology in Higher Education" and "Sustainability" each have 3 articles, reflecting their significant roles in disseminating research. Journals such as "Computers in Human Behaviour," "Interactive Learning Environments," and "Studies in Big Data" each contributed 2 articles, showcasing their relevance in the field. Other notable sources with single contributions include "Behavioural Sciences," "British Journal of Educational Technology," "Educational Technology & Society," "Frontiers in Public Health," "Industry and Higher Education," "International Journal of Data and Network Science," and "International Journal of Human-Computer Interaction."

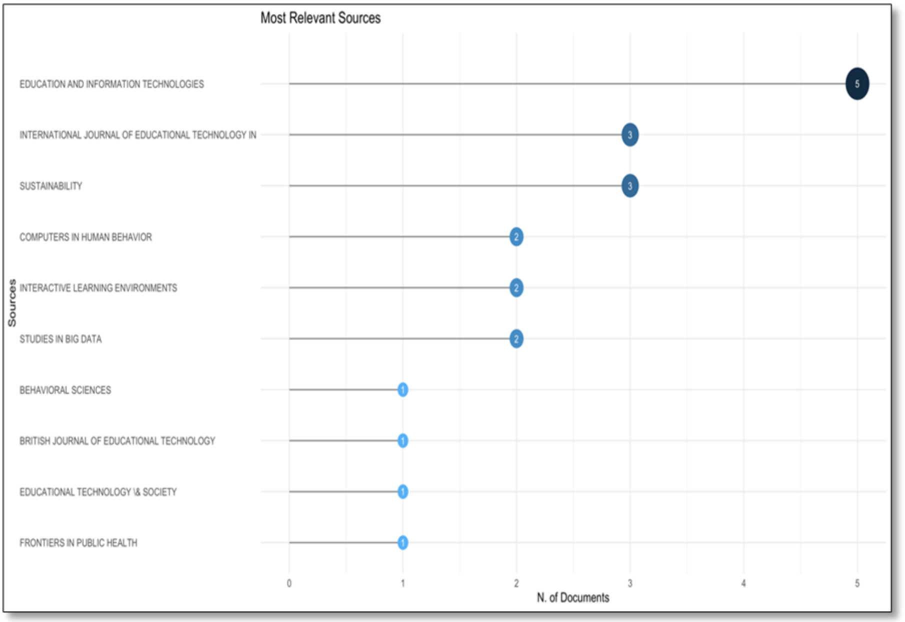


Figure 3. *The most productive sources.*

Thematic Map

The thematic map visualises specific keywords according to their centrality and density, thereby illustrating trends and focal points within the research domain. Figure 4 presents a thematic map with circles representing separate quadrants that outline clustered nodes (Aria & Cuccurullo, 2022; Callon et al., 1991; Cobo et al., 2011; Husamoglu et al., 2024). Configured to set the cluster frequency at eight, the Walktrap clustering algorithm provided a clearer understanding. The map, which uses the Walktrap clustering algorithm, evaluates the graph's structure and identifies clusters of documents characterised by high interaction status (Pons & Latapy, 2005).

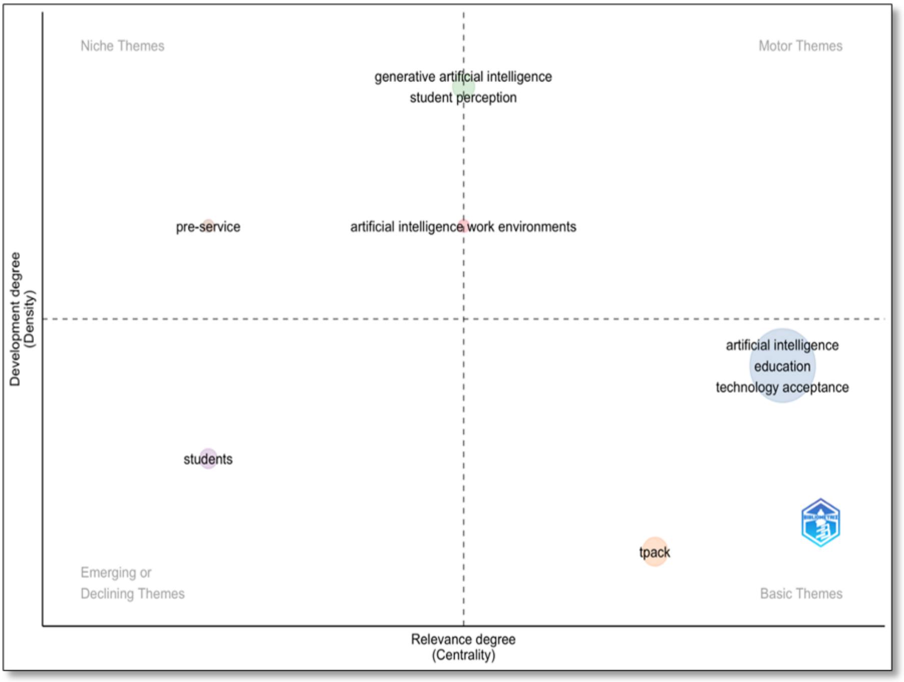


Figure 4. *Thematic map.*

Motor and Niche Themes

The upper right quadrant displays the related motor themes, which exhibit high density, centrality, and development (Aria & Cuccurullo, 2022; Cobo et al., 2011; Husamoglu et al., 2024). The motor themes identified in the thematic map consist of "generative artificial intelligence" and "student perception". The high centrality of these themes indicates that they have strong connections with other research topics and occupy a central position within the research network. The high density indicates that these themes are actively being researched and are heavily discussed in the relevant literature. The "generative artificial intelligence" theme has become a central part of studies examining the creative and generative aspects of AI technologies. This theme shows that AI plays a significant role not only in data analysis and decision-making processes but also in creating new content, supporting creative processes, and driving innovation. The "student perception" theme encompasses studies investigating the effects of educational technologies and pedagogical practises on student perceptions. This theme is crucial for understanding the impact of AI and technological innovations in education on student experiences and learning outcomes.

Niche Themes

Niche themes with high density but low centrality are placed in this quadrant, emphasising their significance in the research field (Aria & Cuccurullo, 2022; Cobo et al., 2011; Husamoglu et al., 2024). "Pre-service" was identified as a niche theme in the thematic map. The high density of this theme indicates that it is actively researched and discussed extensively within its specific area. However, its low centrality means that it has limited interaction with other research topics and is relatively isolated within the research network.

Emerging or Declining Themes

Themes marked by low viscosity and weak centrality are considered either emerging or declining (Aria & Cuccurullo, 2022; Cobo et al., 2011; Husamoglu et al., 2024). The "Students" theme typically involves studies focused on various aspects of student life, experiences, and outcomes in higher educational settings. This could include research on student engagement, learning processes, academic achievement, and social interactions within educational institutions. As an emerging theme, "students" might represent a new area of interest that is beginning to gain attention and could see increased research activity in the future. This could be driven by new educational policies, technological advancements, or societal changes that highlight the importance of understanding student-related issues. Conversely, as a declining theme, "students" might indicate an area where research interest has saturated, possibly due to the maturation of the field, shifts in research priorities, or the resolution of key issues that previously drove research in this area.

Basic Themes

The quadrant showcases basic themes that, despite their low density, have high centrality and are crucial components in the research field (Aria & Cuccurullo, 2022; Cobo et al., 2011; Husamoglu et al., 2024). The identified basic themes in the thematic map are "artificial intelligence," "education," and "technology acceptance." The high centrality of these themes indicates that they are foundational topics that interact with a broad range of other research areas. Their low density suggests that, while they are not the focus of intense, concentrated research activity at present, they remain crucial to the structure and coherence of the research network. This theme encompasses a broad range of studies related to the development and application of AI technologies. As a basic theme, AI serves as a critical underpinning for numerous research areas, including machine learning, data science, and robotics. Its foundational nature ensures that it remains highly relevant across diverse research topics, even if individual studies may not focus on AI alone. The theme of education covers various aspects of teaching, learning, curriculum development, and educational policy. As a basic theme, education is integral to a wide array of research endeavours, influencing studies in fields such as psychology, sociology, and technology. Its central role highlights its importance in shaping research discussions and frameworks across multiple disciplines. This theme explores how individuals and organisations adopt and integrate new technologies. It includes theories and models that explain the factors influencing technology adoption, such as perceived ease of use and usefulness. As a basic theme, technology acceptance is pivotal for understanding the broader implications of technological innovations across different sectors, including healthcare, business, and education.

VOSviewer

VOSviewer visualisation provides an in-depth analysis of the key themes and relationships in AI research within education. The provided visualisation is a VOSviewer map highlighting the interrelationships between various concepts in the field of artificial intelligence and education. The map is divided into three distinct clusters, each representing different thematic connections (Please see Figure 5).

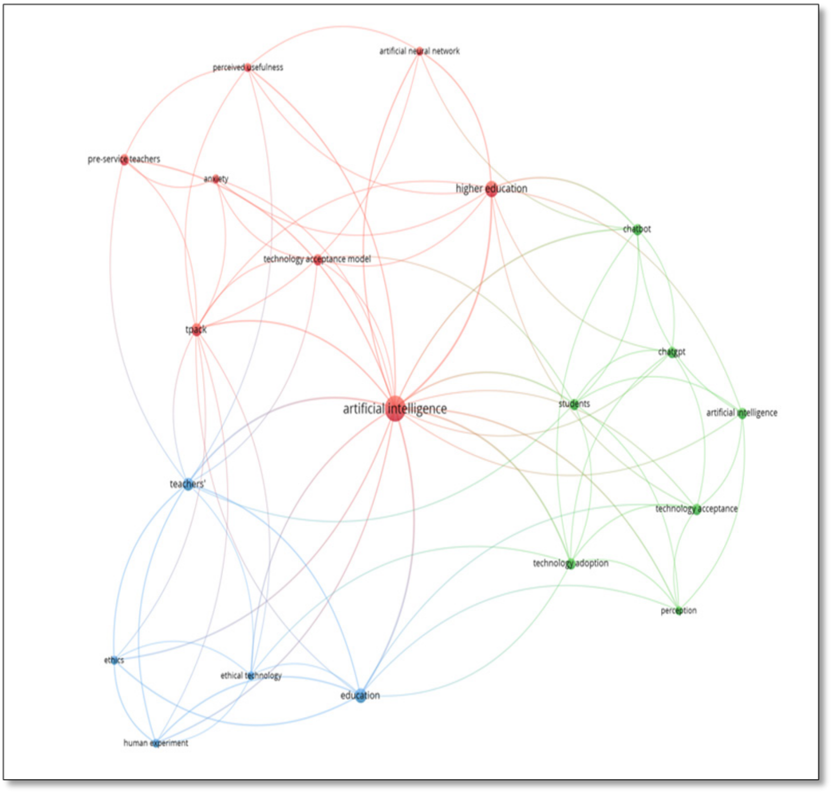


Figure 5. *Authors' keywords cluster analysis.*

Pedagogy: The red cluster centres around “Artificial Intelligence” and includes key terms such as “Anxiety,” “Artificial Neural Network,” “Higher Education,” “Perceived Usefulness,” “Pre-Service Teachers,” “TAM” and “TPACK.” This cluster indicates a focus on the integration of AI within educational settings, particularly in higher education. This study explores how AI and related technologies, such as neural networks, are perceived and adopted by pre-service teachers. The cluster also addresses the psychological aspects, such as anxiety, that might affect the acceptance and usefulness of these technologies. The TPACK framework is highlighted, emphasising pedagogical implications and the need for teachers to integrate technology effectively into their teaching practises.

AI tools: The green cluster also centres around “Artificial Intelligence” but focuses more on specific AI applications such as “Chatbot” and “ChatGPT.” It includes terms such as “Perception,” “Students,” “Technology Acceptance,” and “Technology Adoption.” This cluster illustrates the growing interest in using AI-driven chatbots in educational contexts. This study explores how students perceive these technologies and the factors influencing their acceptance and adoption. The cluster shows a strong emphasis on understanding how these AI tools can enhance the learning experience and the general receptiveness of students towards these innovations.

Ethics: The blue cluster shifts the focus to broader educational and ethical concerns with terms like “Education,” “Ethical Technology,” “Ethics,” “Human Experiment,” and “Teachers.” This cluster underscores the need to address the ethical aspects of implementing AI in education. It suggests a focus on ensuring that AI technologies are used responsibly and ethically, considering the potential implications for human experiments and the broader educational environment. The inclusion of “Teachers” emphasises the role of educators in navigating these ethical challenges and incorporating ethical technology into their teaching.

Overall, the VOSviewer map provides a comprehensive overview of the interconnections between AI and education. It highlights the importance of understanding the psychological, perceptual, and ethical dimensions of AI integration in educational settings. The map showcases a multifaceted approach to AI in education, emphasising the need for effective technology adoption, ethical considerations, and addressing the perceptions and anxieties of both teachers and students.

Conclusion

Theoretical Contribution

This research provides a theoretical background in offering a bibliometric analysis of the current state of the adoption of AI in higher education. It has set a path for understanding the shift in the discourse about AI in higher educational settings by highlighting the publication trends, top contributing countries, thematic clusters, and keywords with bibliometric analysis methods like Biblioshiny and VOSviewer.

The conducted analysis indicates a noticeable and continuing pattern of growth in AI integration research in higher education, particularly from 2020 to 2024. This trend reflects an increased interest in the topic within academia, with contributions distributed across various countries and prominent journals, demonstrating a growing focus on the in-depth exploration of AI's applications and implications in educational contexts.

The thematic map identifies "Generative AI" and "student perception" as motor themes, indicating active research and significant connections with other topics, while "Artificial intelligence," "education," and "technology acceptance" function as basic themes that serve as foundational elements across the research landscape. Additionally, the "students" theme, which explores various aspects of student life in higher education, is emerging, reflecting growing interest, whereas "pre-service" education appears as a niche theme with a specialised focus but limited broader interaction.

The thematic clusters further reveal the focus areas within AI integration in education. The "pedagogy" cluster centres on incorporating AI into educational practises, addressing challenges like anxiety, and adopting frameworks such as TAM and TPACK. The "AI tools" cluster emphasises practical applications, such as chatbots and ChatGPT, examining how students perceive these technologies and the factors influencing their acceptance. Lastly, the "ethics" cluster highlights the need for responsible AI use, focusing on ethical considerations and the role of educators in navigating these challenges.

The thematic map and the identification of the core and emerging themes (e.g., pedagogy, AI tools, and ethics) provide a clear understanding of the key areas of interest and reveal the directions for the future growth of the theory. Altogether, the research highlights the role of AI in higher education institutions while presenting a systematic method for analysing its diffusion based on bibliometrics. The paper also provides and compares various alternative scales to measure AI integration in education from both students' and instructors' perspectives. Therefore, it provides a foundation for rational decision-making and establishes a direction for subsequent studies that will seek to optimise the positive effects of AI and minimise the challenges of AI integration learning environments.

Practical Contribution

This research provides recommendations that would be beneficial to policymakers, educators, education technology suppliers, and other stakeholders in higher education institutions. As the list of topics might give an insight into, it offers a well-defined guideline for decision-making and strategizing by identifying the major themes and issues underlying AI implementation, including learning theories, technologies, and issues of ethics. Such knowledge can also help to design relevant interventions, standards, and actions to improve the integration of AI into teaching-learning processes.

Our analysis of emerging theme signals that educators and students might have concerns when using AI tools primarily because of the perceived lack of defined ethical rules. To address this issue, the development of a comprehensive guideline within educational settings becomes an urgent priority. This guideline should result from a collaborative effort involving relevant stakeholders, covering key topics, eliminating uncertainties regarding ethical issues, and achieving broad acceptance and applicability among its beneficiaries.

As discussed by Bisdas et al. (2021), students already have a positive attitude towards the use of AI tools in their education. However, some studies (e.g., Abdelwahab et al., 2023) posit that students face adoption challenges because the use of AI does not integrate into their curriculum and thereby have concerns about using AI tools just because of unknown standards of ethics (Mohd Rahim et al., 2022; Chan & Hu, 2023). Therefore, it would be wise for responsible authorities to organise training sessions specifically on how to properly employ AI tools. Through this strategy, educators can learn the pedagogical skills necessary to train AI applications for education and research purposes through "train the trainer" programmes and then also provide assistance to their students in gaining the knowledge needed to effectively use AI tools in education and research.

Limitations and Future Studies

This research has several limitations. First, the focus was mainly on sources from Scopus and Web of Science, which may not necessarily cover all the scholarly sources on AI adoption in higher education in the existing literature. The sources for the current research are restricted to Scopus and WoS up to July 22, 2024. Further studies can expand on AI integration into higher education by employing other databases. Moreover, one of the main limitations is that bibliometric analysis does not provide a rich qualitative analysis of the problem, including factors such as the perceptions of educators and students. Future research should build on these findings using a mixed-methods approach to examine participants' perceptions and understanding of AI use in learning environments. Moreover, longitudinal studies could document changes in the topics across the years and evaluate the effects of the advancements in AI technologies on learning and teaching practises and achievements.

Ethics Committee Approval: As this study is based entirely on a review of existing literature, rather than primary data collection, ethics committee approval was not required. The analysis and synthesis of previously published research do not involve new interactions with human subjects or require ethical clearance.

Peer-review: Externally peer-reviewed.

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: This research has been partially supported by the Erasmus+ Programme under the project titled 'Empowering Educators for AI-Integrated Learning' (Project ID: 2023-2-TR01-KA220-HED-000185020).

References

Abdelwahab, H. R., Rauf, A., & Chen, D. (2023). Business students' perceptions of Dutch higher educational institutions in preparing them for artificial intelligence work environments. *Industry and Higher Education*, *37*(1), 22–34. https://doi.org/10.1177/09504222221087614

Alhumaid, K., Al Naqbi, S., Elsori, D., & Al Mansoori, M. (2023). The adoption of artificial intelligence applications in education. *International Journal of Data and Network Science*, *7*(1), 457–466. https://doi.org/10.5267/j.ijdns.2022.8.013

An, X., Chai, C. S., Li, Y., Zhou, Y., Shen, X., Zheng, C., & Chen, M. (2023). Modeling English teachers' behavioral intention to use artificial intelligence in middle schools. *Education and Information Technologies*, *28*(5), 5187–5208. https://doi.org/10.1007/s10639-022-11286-z

Aria, M., & Cuccurullo, C. (2022, March 20). *Science Mapping Analysis with bibliometrix R- package: An example*. https://bibliometrix.org/documents/bibliometrix_Report.html

Bilquise, G., Ibrahim, S., & Salhieh, S. M. (2024). Investigating student acceptance of an academic advising chatbot in higher education institutions. *Education and Information Technologies*, *29*(5), 6357–6382. https://doi.org/10.1007/s10639-023-12076-x

Bisdas, S., Topriceanu, C.-C., Zakrzewska, Z., Irimia, A.-V., Shakallis, L., Subhash, J., Casapu, M.-M., Leon-Rojas, J., Pinto dos Santos, D., Andrews, D. M., Zeicu, C., Bouhuwaish, A. M., Lestari, A. N., Abu-Ismail, L., Sadiq, A. S., Khamees, A., Mohammed, K. M. G., Williams, E., Omran, A. I., … Ebrahim, E. H. (2021). Artificial Intelligence in Medicine: A Multinational Multi-Center Survey on the Medical and Dental Students' Perception. *FRONTIERS IN PUBLIC HEALTH*, *9*. https://doi.org/10.3389/fpubh.2021.795284

Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemsitry. *Scientometrics*, *22*(1), 155–205. https://doi.org/10.1007/BF02019280

Çelik, I. (2023). Towards Intelligent-TPACK: An empirical study on teachers professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior*, *138*. https://doi.org/10.1016/j.chb.2022.107468

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, *20*(1). https://doi.org/10.1186/s41239-023-00411-8

Chatterjee, S., & Bhattacharjee, K. K. (2020). Adoption of artificial intelligence in higher education: A quantitative analysis using structural equation modelling. *Education and Information Technologies*, *25*(5), 3443–3463. https://doi.org/10.1007/s10639-020-10159-7

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics*, *5*(1), 146–166. https://doi.org/10.1016/j.joi.2010.10.002

Dahri, N. A., Yahaya, N., Al-Rahmi, W. M., Vighio, M. S., Alblehai, F., Soomro, R. B., & Shutaleva, A. (2024). Investigating AI-based academic support acceptance and its impact on students' performance in Malaysian and Pakistani higher education institutions. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-12599-x

Delcker, J., Heil, J., Ifenthaler, D., Seufert, S., & Spirgi, L. (2024). First-year students AI-competence as a predictor for intended and de facto use of AI-tools for supporting learning processes in higher education. *International Journal of Educational Technology in Higher Education*, *21*(1). https://doi.org/10.1186/s41239-024-00452-7

Echchakoui, S. (2020). Why and how to merge Scopus and Web of Science during bibliometric analysis: The case of sales force literature from 1912 to 2019. *Journal of Marketing Analytics*, *8*(3), 165–184. https://doi.org/10.1057/s41270-020-00081-9

Foroughi, B., Senali, M. G., Iranmanesh, M., Khanfar, A., Ghobakhloo, M., Annamalai, N., & Naghmeh-Abbaspour, B. (2023). Determinants of Intention to Use ChatGPT for Educational Purposes: Findings from PLS-SEM and fsQCA. *International Journal of Human-Computer Interaction*. https://doi.org/10.1080/10447318.2023.2226495

Husamoglu, B., Akova, O., & Cifci, I. (2024). Regenerative stakeholder framework in tourism. *Tourism Review*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/TR-12-2023-0889

Jain, K. K., & Raghuram, J. N. V. (2024). Gen-AI integration in higher education: Predicting intentions using SEM-ANN approach. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-12506-4

Li, K. (2023). Determinants of College Students' Actual Use of AI-Based Systems: An Extension of the Technology Acceptance Model. *Sustainability*, *15*(6). https://doi.org/10.3390/su15065221

Mohd Rahim, N. I., A. Iahad, N., Yusof, A. F., & A. Al-Sharafi, M. (2022). AI-Based Chatbots Adoption Model for Higher-Education Institutions: A Hybrid PLS-SEM-Neural Network Modelling Approach. *Sustainability (Switzerland)*, *14*(19). https://doi.org/10.3390/su141912726

Ning, Y., Zhang, C., Xu, B., Zhou, Y., & Wijaya, T. T. (2024). Teachers' AI-TPACK: Exploring the Relationship between Knowledge Elements. *Sustainability (Switzerland)*, *16*(3). https://doi.org/10.3390/su16030978

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*. https://doi.org/10.1136/bmj.n71

Pons, P., & Latapy, M. (2005). Computing Communities in Large Networks Using Random Walks. In pInar Yolum, T. Güngör, F. Gürgen, & C. Özturan (Eds.), *Computer and Information Sciences—ISCIS 2005* (pp. 284–293). Springer Berlin Heidelberg.

Salloum, S. A., Almarzouqi, A., Aburayya, A., & Alfaisal, R. (2024). Adoption of Chatbots for University Students. *Studies in Big Data*, *144*, 233–246. https://doi.org/10.1007/978-3-031-52280-2_15

Shwedeh, F., Salloum, S. A., Aburayya, A., Fatin, B., Elbadawi, M. A., Al Ghurabli, Z., & Al Dabbagh, T. (2024). AI Adoption and Educational Sustainability in Higher Education in the UAE. *Studies in Big Data*, *144*, 201–229. https://doi.org/10.1007/978-3-031-52280-2_14

Strzelecki, A. (2023). To use or not to use ChatGPT in higher education? A study of students' acceptance and use of technology. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2023.2209881

Sun, F., Tian, P., Sun, D., Fan, Y., & Yang, Y. (2024). Pre-service teachers' inclination to integrate AI into STEM education: Analysis of influencing factors. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13469

Wang, K., Ruan, Q., Zhang, X., Fu, C., & Duan, B. (2024). Pre-Service Teachers' GenAI Anxiety, Technology Self-Efficacy, and TPACK: Their Structural Relations with Behavioral Intention to Design GenAI-Assisted Teaching. *Behavioral Sciences*, *14*(5). https://doi.org/10.3390/bs14050373

Wang, X., Li, L., Tan, S. C., Yang, L., & Lei, J. (2023). Preparing for AI-enhanced education: Conceptualizing and empirically examining teachers' AI readiness. *Computers in Human Behavior*, *146*. https://doi.org/10.1016/j.chb.2023.107798

Wang, Y., Liu, C., & Tu, Y.-F. (2021). Factors Affecting the Adoption of AI Based Applications in Higher Education: An Analysis of Teachers Perspectives Using Structural Equation Modeling. *Educational Technology and Society*, *24*(3), 116–129. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85110454980&partnerID=40&md5=c59227a37550a07ad8a4695ad30dcc82

Yang, J., Wang, Q., Wang, J., Huang, M., & Ma, Y. (2021). A study of K-12 teachers' TPACK on the technology acceptance of E-schoolbag. *Interactive Learning Environments*, *29*(7), 1062–1075. https://doi.org/10.1080/10494820.2019.1627560

Zhang, C., Schießl, J., Plößl, L., Hofmann, F., & Gläser-Zikuda, M. (2023). Acceptance of artificial intelligence among pre-service teachers: A multigroup analysis. *International Journal of Educational Technology in Higher Education*, *20*(1). https://doi.org/10.1186/s41239-023-00420-7

Journal of Data Applications
Issue: 3, 2024
E-ISSN: 2980-3357

RESEARCH ARTICLE

Detection of Urgent Messages Shared on Twitter during an Earthquake using the Deep Learning Method

Mücahit SÖYLEMEZ[1], Ali ÖZTÜRK[2,3]

ABSTRACT

The ability of Twitter to provide real-time information during disasters is becoming more widely acknowledged, making it an essential forum for people to voice their concerns and ask for help during emergencies. These platforms can speed up the distribution of help, but they are also prone to false information, which might make disaster response more difficult. Using a carefully selected dataset of 10,200 tweets that have been extensively preprocessed and tokenized for reliable training and validation, this study uses deep learning models, such as LSTM, BLSTM, and BLSTMA, to classify tweets during earthquake events into two categories: "under the debris" and "not under the debris." The model performance was further improved via hyperparameter adjustment, which included neuron counts, dropout rates, dimensions, and embedding types. The results of this study showed that while the BLSTMA model had the best accuracy (96.64%) and F1 score (0.9116), conventional machine learning techniques like XGBoost and SVM. However, in other measurements, it was shown that standard machine learning techniques like SVM and XGBoost performed better. Using Bag of Words vectorisation, SVM obtained 95.81% accuracy and an F1 score of 0.9579, whereas XGBoost earned 95.84% accuracy and an F1 score of 0.9584. By demonstrating the usefulness of the BLSTMA model in real-time disaster response and the complementary advantages of conventional approaches in the analysis of complex disaster data, these findings highlight the significance of customising machine learning and deep learning approaches to particular tasks.

Keywords: Deep Learning, Natural Language Processing, Disaster Management, Twitter Analysis, Emergency Message Detection.

DOI: 10.26650/JODA.1606403

[1]İstanbul University, Faculty of Economics, Department of Management Information Systems, İstanbul, Türkiye
[2]KTO Karatay University, Department of Computer Engineering, Konya, Türkiye
[3]Chief Systems Engineers, Havelsan A.Ş., Ankara, Türkiye

ORCID: M.S. 0009-0009-0059-616X; A.Ö. 0000-0002-1797-2039

Corresponding author:
Mücahit SÖYLEMEZ,
İstanbul University, Faculty of Economics, Department of Management Information Systems, İstanbul, Türkiye
E-mail: mucahit.soylemez@istanbul.edu.tr

Submitted: 24.12.2024
Revision Requested: 16.01.2025
Last Revision Received: 21.01.2025
Accepted: 22.01.2025
Published Online: 31.01.2025

Citation: Söylemez, M., & Öztürk, A. (2024). Detection of Urgent Messages Shared on Twitter during an Earthquake using the Deep Learning Method. *Journal of Data Applications, 3*, 63-84. https://doi.org/10.26650/JODA.1606403

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International License

Introduction

Natural disasters are among the foremost events that threaten human life and economies on a global scale. Among these, earthquakes stand out as the most significant natural disaster, leading to the loss of human life. The necessity for countries on active fault lines to establish an effective disaster management system becomes increasingly evident with each passing day. Therefore, leveraging all available technological resources should be a primary objective for these nations. The greater the destruction caused by a disaster; the more critical communication becomes. Considering this, implementing additional measures in disaster management has become an essential requirement for every country. For instance, Hurricane Harvey, which struck the United States in certain years, caused significant hardship for millions of Americans, prompting the U.S. to take additional precautions. Similarly, taking extra measures against earthquakes, which are even more impactful, should be a priority in the disaster management strategies of all countries.

The role of social media platforms in crisis management has gained increasing importance with advancements in technology. Platforms like Twitter, in particular, have become critical tools for communication during disasters, facilitating rapid information flow for both individuals and aid teams. In recent years, during an earthquake in Turkey, survivors trapped under debris resorted to sharing their locations and conditions through social media platforms like Twitter due to the inability to communicate via traditional phone lines or conventional communication channels (Al Jazeera, 2023). These platforms proved instrumental for rescue teams, helping to save many lives (Euronews, 2023). However, the dissemination of false information on such social media platforms could not be entirely prevented.

Earthquakes, as one of the most destructive natural disasters, lead to significant loss of life, widespread damage to infrastructure, and disruption of daily life. In the aftermath of an earthquake, many individuals are trapped under debris, and the timely identification of those in need of rescue becomes critical. In such chaotic situations, traditional communication systems may fail, and the flow of information can be severely hindered, which delays rescue efforts and worsens the crisis.

Social media platforms, particularly Twitter, play a pivotal role in these scenarios by providing individuals with the opportunity to share real-time information about their location, conditions, and urgent needs (Al Jazeera, 2023). While this communication channel holds the potential to significantly aid disaster response teams, it also comes with the challenge of filtering out irrelevant or false information. To address these challenges, this study aims to analyse tweets shared by individuals trapped under debris during earthquakes using a deep learning-based artificial intelligence system. The goal is to accurately determine whether these individuals are truly in need of rescue, thereby enhancing the effectiveness of disaster management efforts by providing reliable information (Euronews, 2023).

Literature Review

Social media platforms have become a crucial source of information for crisis management. In this context, Powers et al. (2023) investigated how social media may help during natural disasters by detecting emergency signals. With an emphasis on tweets from Hurricane Harvey, they investigated many methods for recognising messages from persons in imminent need. Their results showed that certain models, such as XLNet and BERT, outperformed others; CNN's accuracy was 72%, but BERT's was 78%. The paper makes the case that more data might further improve these systems' efficacy while highlighting the importance of social media in disaster response. (Powers et al., 2023)

During natural disasters, Pradip Bhere and colleagues investigated the use of deep learning to differentiate between relevant and irrelevant tweets. To improve tweet classification, they created a system that combined a convolutional neural network (CNN) with Word2Vec feature vectors. They tested the model with different word embeddings (Custom Weight, Google News, Twitter Glove) using a dataset of 10,000 tweets from Kaggle that dealt with disasters. According to the results, the model's accuracy using Google News embeddings was 86%, while its accuracy using their proposed method was 84%. In addition, they evaluated its accuracy using a confusion matrix on tweets from the 2013 Colorado floods. This work demonstrates how deep learning can be used to identify social media data during emergencies and raises the possibility that future performance could be improved with larger datasets. (Bhere et al., 2020).

Kumar et al. compared several machine learning and deep learning techniques for classifying social media tweets regarding catastrophes in order to examine performance under data imbalance. The study found that the deep learning models performed better than the conventional classifiers. For the hurricane dataset, BIGRU got the greatest F1 score (0.87), while for the earthquake dataset, GRU-CNN had the highest F1 score (0.88). These findings demonstrate how well deep learning models categorise tweets concerning catastrophes and how they may improve local disaster response activities. (Kumar et al., 2019).

Behl et al. (2021) investigated a range of machine learning and deep learning models to classify Twitter data during COVID-19 and natural catastrophe occurrences. They examined various models, including multilayer perceptions (MLP-TF and MLP-W), convolutional neural networks (CNN-W and CNN-WF), and logistic regression (LR-TF). Their findings revealed that the LR-TF model achieved the best performance, with 88% accuracy on a dataset of earthquakes in Nepal and Italy and 81% accuracy on the COVID-19 dataset. Both CNN-W and CNN-WF delivered similar performance across the datasets, though CNN-WF's accuracy was slightly lower at 78% on the COVID-19 dataset. The accuracy of MLP-TF decreased from 87% on the combined dataset to 77% on the COVID-19 dataset. In contrast, the MLP-W model performed best on the COVID-19 dataset, achieving an accuracy of 83% (Behl et al., 2021).

Madichetty et al. developed a Stacked Convolutional Neural Network (SCNN) model to identify resource-related tweets during emergencies. The model integrates CNN and KNN classifiers at the base level, with an SVM meta-classifier processing the outputs for the final classification. When tested on datasets from the 2015 Nepal and 2016 Italy earthquakes, the SCNN model outperformed other combinations, achieving the highest accuracy of 77.5% for Nepal and 76.99% for Italy. These results highlight the effective collaboration of CNN, KNN, and SVM in categorizing social media data for disaster management (Madichetty & Sridevi, 2020).

Muhammed Ali Sit and colleagues analysed tweets during Hurricane Irma to explore the use of social media in disaster management. They found that Long Short-Term Memory (LSTM) networks performed best, achieving 74.78% accuracy and 75.14% F1 score in classifying disaster-related tweets. Other methods, such as CNN and logistic regression, performed less well. The study highlights LSTM as the most effective model for analysing social media data during crises.

All these studies have highlighted the potential of social media platforms in crisis management and have provided valuable contributions to the disaster management literature using deep learning, machine learning, and natural language processing techniques.

In our study, tweets from individuals trapped under the debris were successfully identified using natural language processing and deep learning models applied to a unique dataset.

Studies in the literature emphasise the importance of artificial intelligence-based approaches developed for analysing and interpreting social media data during crises. Various research efforts have introduced innovative methods using natural language processing, deep learning, and machine learning techniques to contribute to disaster management processes. Below is a summary of the contributions these studies have made to the literature.

Materials and Methods

Artificial intelligence, with a deep-rooted history spanning from ancient times to the modern era, has evolved into a discipline that finds applications in almost every field of contemporary technology (Mijwel, 2015). This study focuses on deep learning and natural language processing, which are subfields of artificial intelligence, aiming to effectively classify social media data for disaster management. To achieve this goal, the performance of the models developed using deep learning techniques and natural language processing methods was evaluated. We implemented a machine learning pipeline to classify the textual data using various recurrent neural network (RNN) architectures. The preprocessing steps included converting text to lowercase, removing URLs and special characters, tokenization, stopwords removal, and stemming using the TurkishStemmer library. The dataset was split into training and validation sets, and word embeddings were generated using both a pre-trained Word2Vec model and random initialisation. Three RNN architectures—LSTM, bidirectional LSTM (BLSTM), and BLSTM with attention (BLSTMA)—were constructed using TensorFlow and Keras. Each architecture was configured with different hyperparameters, such as the number of units, dropout rates, and learning rates, to evaluate their performance. Training was conducted using class-balanced weights and monitored using callbacks for early stopping and learning rate adjustment. Performance metrics, including accuracy, precision, recall, and F1 score, were calculated to assess the effectiveness of each model configuration. The experimental results are presented in detail in the subsequent sections.

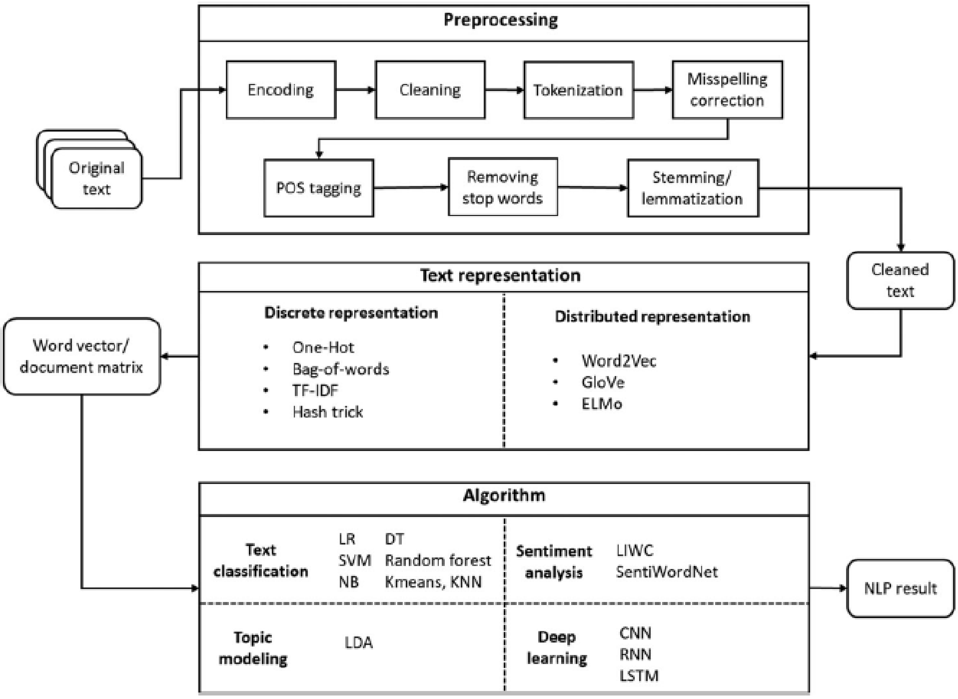


Figure 1. *Natural Language Processing Steps (Kang et al., 2020)*

Natural Language Processing (NLP)

Natural Language Processing (NLP) is an area of artificial intelligence that focuses on enabling machines to understand, interpret and respond to human language in a meaningful way. The core of NLP is the transformation of raw textual data into structured forms suitable for machine learning algorithms. This transformation involves a series of preprocessing steps, text representation techniques, and algorithmic methods tailored to specific applications such as sentiment analysis, text classification, and machine translation (Kang et al., 2020).

Preprocessing in the NLP

A number of preprocessing steps are taken at the beginning of the NLP process to prepare the textual input for the machine learning algorithms. These steps include encoding compatibility, cleaning up HTML elements and redundant components, segmenting text into units (tokenisation), spelling correction, tagging word types (POS tagging), removing stop words, and reducing words to base or root forms (stemming or lemmatisation). These procedures ensure that the raw text is standardised and cleaned, allowing for better performance of downstream NLP models (Kang et al., 2020). Figure 1 shows these steps in detail. The effects of these preprocessing steps are discussed in the Dataset and Experimental Results sections.

Text representation techniques

Once pre-processing is complete, the cleaned text is converted into numerical representations for the machine learning models. This conversion can be performed using separate or distributed representation methods:

- Discrete representations: Techniques such as One-Hot Encoding, Bag-of-Words (BOW), Term Frequency-Inverse Document Frequency (TF-IDF) and the shing trick are used to represent text as vectors based on word occurrence or importance within a document. These methods capture basic textual information but may lose contextual relationships (Kang et al., 2020).

- Distributed representations: Advanced methods such as Word2Vec, GloVe, and FastText generate dense vector representations that preserve the semantic relationships between words. These embeddings are particularly useful for capturing word meaning and context, making them suitable for tasks such as sentiment analysis and machine translation (Kang et al., 2020).

Applications such as text categorisation, machine translation and sentiment analysis often use these representation techniques. Discrete representations are often paired with traditional machine learning algorithms such as logistic regression (LR), support vector machines (SVM), and decision trees (DT), while distributed representations are often integrated with neural networks such as convolutional neural networks (CNN), recurrent neural networks (RNN), and long short-term memory (LSTM) (Kang et al., 2020).

This study was undertaken for evaluating the effectiveness of various NLP techniques in a text classification task. As outlined in Figure 1, the NLP process begins with preprocessing, the aim of which is to clean and standardise the raw text. Subsequently, both discrete and distributed text representation techniques were applied to convert the text into machine-readable formats. Finally, these representations were used to train and evaluate classification models, such as LSTM-based neural networks. The study's objective is to determine the most effective techniques for enhancing text classification accuracy. In summary, this study underscores the significance of preprocessing, investigates diverse text representation methods, and employs sophisticated algorithms to attain optimal performance in NLP tasks. Figure 1 offers a comprehensive visual depiction of the entire NLP pipeline (Kang et al., 2020).

Dataset

For the analysis of messages shared on Twitter within the scope of disaster management, a dataset comprising 10,200 tweets was prepared for the period between February 6 and 8. The data collection process was conducted using the Python programming language and the SntTwitter library. During this process, tweets were collected within a specific focus area by using disaster-related hashtags (e.g., #earthquake, #help). The collected data were recorded with attributes such as date, content, and hashtags used and subsequently manually classified.

The classification process divided the tweets into two categories: "emergency help messages" (1) and "general information sharing" (0). This classification was performed to facilitate the training process of the supervised learning algorithms. Table 1 provides examples from the dataset, offering a framework for understanding how social media data are processed in the context of disaster management.

Table 1. *Twitter Dataset*

Date	**Tweet**	**HashTags**	**Label**
2023-02-06 23:59:56+00:00	SİTE 1 NO:20 HATAY MERKEZ MELİSA YARDIM BEKLİYOR	['afaddeprem','yardım','deprem' 'afad']	1
2023-02-06 29:59:51+00:00	Aksever Mahallesi Meltem Sokak Güler Apartmanı	['hatayyardımbekliyor','hatayafad', 'hatayardım']	1
2023-02-06 23:59:38+00:00	ÖNEMLİ DUYURU YAYALIM hatay hatayyardımbekliyor	['hatay','hatayyardımbekliyor', 'ENKAZALTINDA'	0

To provide a general perspective by analysing the frequency and contexts of words within the dataset, a word cloud analysis was conducted. Figure 2 visualises the most frequently used words, highlighting key themes that are prominent in the context of disaster management. This analysis contributes significantly to the identification of critical themes and patterns in the data.

Figure 2. *Word Cloud in the Dataset*

The dataset predominantly features earthquake-related emergency tweets, where the most frequent words—"deprem" (earthquake), "yardım" (help), "lütfen" (please), and "enkaz" (debris)—indicate a high volume of distress calls and urgent requests for rescue operations. As illustrated in Figure 3, the frequency distribution of these keywords highlights their critical role in disaster-related communication. Additionally, the presence of geographical terms like "Hatay" and references to AFAD (Turkey's Disaster and Emergency Management Authority) suggests a focus on specific affected regions and official response efforts, emphasising the critical need for immediate aid and coordinated rescue missions.

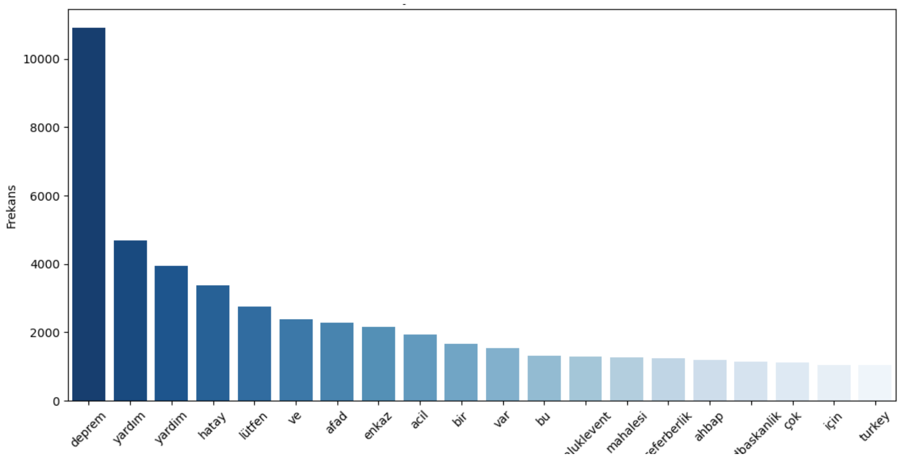


Figure 3. *Frequency analysis of the 20 most frequently mentioned words*

Deep Learning Methods

Deep learning is a branch of machine learning that provides effective solutions to complex problems by extracting meaningful insights from large datasets (Goodfellow, Bengio, & Courville, 2016). This method, particularly prominent in sequential data, natural language processing, and time-series analysis, has enabled groundbreaking advancements in data analysis processes. At the core of deep learning techniques lie artificial neural networks (ANNs), which are inspired by the neural system of the human brain. ANNs are structures capable of solving non-linear problems by learning input and output values and processing them through a specific algorithm to generate results (Goodfellow, Bengio, & Courville, 2016). The development of this field, which began with single-layer perceptions, led to the creation of multilayer perceptions (MLPs) to address the need for understanding non-linear relationships and complex data structures. MLPs, capable of learning more intricate relationships, have found wide applications in classification and regression problems within machine learning (Rosenblatt, 1958; Goodfellow, Bengio, & Courville, 2016). However, the processing of sequential and time-dependent data highlighted the limitations of ANNs, prompting the development of Recurrent Neural Networks (RNNs) and their derivatives.

Recurrent Neural Networks (RNNs) are specialised architectures designed to work with sequential data and are capable of learning from sequential inputs (Goodfellow, Bengio, & Courville, 2016). However, RNNs face challenges such as gradient vanishing when learning long-term dependencies. To address these issues, **Long Short-Term Memory (LSTM)** models were developed. LSTM models employ mechanisms such as forget, input, and output gates, effectively controlling the flow of information and excelling in learning long-term dependencies (Hochreiter & Schmidhuber, 1997). Due to these features, LSTM models are widely used in fields such as text processing, time-series analysis, and natural language processing (Hochreiter & Schmidhuber, 1997; Graves, Mohamed, & Hinton, 2013).

An advanced version of LSTM, **Bidirectional Long Short-Term Memory (BLSTM),** enhances contextual representation by learning both past and future contexts in sequential data (Zhou et al., 2016). BLSTM processes information bidirectionally, achieving robust results in preserving semantic coherence, particularly in text data. The performance of the BLSTM further improves when combined with the **attention mechanism**. The attention mechanism enables the model to focus on critical inputs, prioritising key information, especially in long and complex sequences (Zhou et al., 2016).

Finally, the **Bidirectional Long Short-Term Memory with Attention (BLSTMA)** model, which integrates the attention mechanism into BLSTM, not only remembers information during the learning process but also focuses on the most critical elements, achieving higher accuracy and efficiency (Zhong et al., 2020). This progressive evolution of the LSTM, BLSTM, and BLSTMA models significantly enhances the sequential data processing capabilities of artificial neural networks, paving the way for groundbreaking applications in various fields, particularly natural language processing.

Training Parameters Of Deep Learning Models

In this study, a comprehensive performance comparison of the deep learning algorithms was conducted using metrics such as accuracy, precision, recall, and F1 scores. In addition, various combinations of hyperparameters were tested, including random embedding and FastText-based embedding techniques with representations of 100, 200, and 300 dimensions; neuron counts of 64, 128, and 256 units; and dropout rates of 0.2, 0.3, and 0.4. The results were evaluated to assess the potential of optimising disaster management tasks, such as analysing social media data, using deep learning algorithms (LSTM, Bidirectional LSTM - BLSTM, and Attention-augmented Bidirectional LSTM - BLSTMA).

A total of 154 variations were created by testing these hyperparameter combinations within the framework of the four specified parameter sets. For each variation, the average accuracy values were calculated to enable a comprehensive performance comparison of the models, particularly in their application to disaster management. This analysis identified the optimal configuration for effectively processing and interpreting social media data in crisis scenarios.

Experimental Results

The dataset used in this study contained unnecessary words and was cleaned by applying the data preprocessing steps outlined in Figure 1. The data cleaning steps employed in this context can be summarised as follows:

- Turkish stopword removal: Common but semantically insignificant words such as “ve” (and) and “bir” (a) were removed from the dataset using the NLTK library.

- Removal of URLs and username tags: The URLs and username tags present in the tweets were extracted from the dataset.

- Cleaning of special characters: Emojis and special characters such as “#” and “$” found in the tweet text were removed.

- Standardisation of uppercase and lowercase letters: All text was converted to lowercase to ensure consistency in character usage.

- Elimination of repeated letters: Expressions with unnecessarily repeated letters were corrected.

These preprocessing steps were performed to make the data more meaningful and analysable. Once the preprocessing was completed, the tweets were tokenised by splitting them into word sequences. During the tokenization step, each word was assigned a unique integer value. For a detailed explanation of the data preprocessing and tokenization steps, refer to Table 2.

Table 2. *Tokenization Process*

Step	Result
Raw Text	SİTE 1 NO:20 HATAY MERKEZ MELİSA YARDIM BEKLİYOR afaddeprem yardım deprem afad
Convert to Lowercase	site 1 no:20 hatay merkez melisa yardım bekliyor afaddeprem yardım deprem afad
Remove Punctuation	site 1 no20 hatay merkez melisa yardım bekliyor afaddeprem yardım deprem afad
Tokenize Words	[‘site’,’1’,’no20’,’hatay’,’merkez’,’melisa’,’yardım’,’bekliyor’,’afaddeprem’, ‘yardım’,’deprem’,’afad’]
Assign Unique Index	{‘site’:1,’1’:2,’no20’: 3, ‘hatay’: 4, ‘merkez’: 5, ‘melisa’: 6, ‘yardım’: 7, ‘bekliyor’: 8, ‘afaddeprem’: 9, ‘yardım’: 10, ‘deprem’: 11, ‘afad’: 12}
Convert Words to Indexes	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]
Fixed-Length Sequence	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]

In the deep learning approach, the dataset was divided into training and validation sets, with the training set used for model learning and the validation set for evaluating model performance. The models were trained using the binary_crossentropy loss function and the Adam optimisation algorithm. To prevent overfitting and enhance performance, the early stopping method was applied.

The model was trained for 30 epochs, with the data split into training, validation, and test sets in a 67%-33% ratio. During training, early stopping was applied by monitoring the val_loss metric, ensuring that the best weights were retained to avoid performance degradation. Additionally, data augmentation techniques such as word order shuffling and random word dropping were used to increase the model's generalisation capacity. The model was trained using Word2Vec and random embedding matrices and structured with LSTM, Bidirectional LSTM (BLSTM), and Attention-based BLSTM (BLSTMA) architectures. The models were trained with class weights, continuously monitored with validation data, and the model with the highest accuracy was selected.

During the training process, the classification models learned the context of the tweets, enabling them to accurately classify new messages. The changes in accuracy based on the number of epochs are illustrated in Figures 4, 5, and 6.

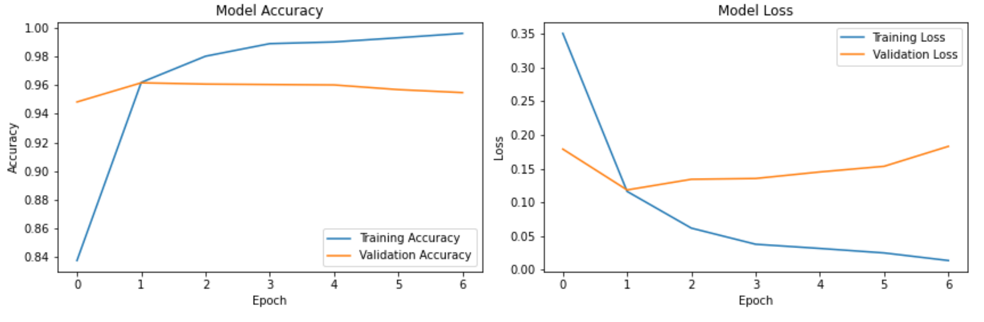


Figure 4. *LSTM Training Process*

As seen in Figure 4, the graphs show the LSTM model's training and validation performance throughout the six epochs. It is clear that by the fourth epoch, the training accuracy has increased consistently to about 98%, while the validation accuracy has stabilised at a somewhat lower level, between 94% and 95%. Additionally, after the second epochs, the training loss steadily dropped from 0.9 to about 0.05, while the validation loss reached a plateau at 0.25. Nevertheless, the difference between these two forms of loss remains, indicating that the LSTM model has trouble generalising, most likely because it is inadequate at capturing long-range dependencies (Hochreiter & Schmidhuber, 1997).

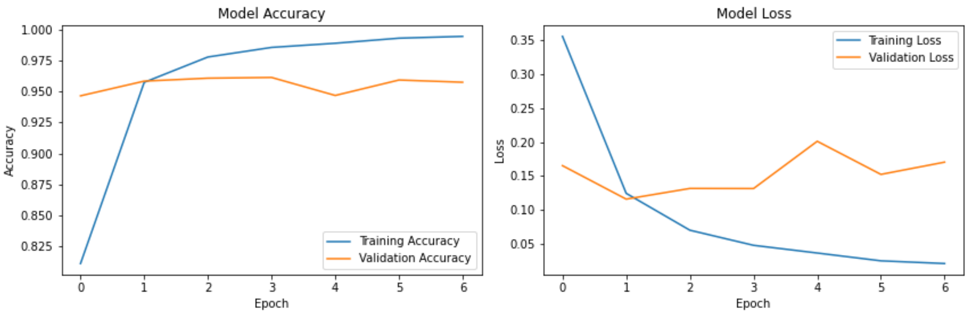


Figure 5. *BLSTM Training Process*

The BLSTM model has 99% training accuracy in the first two epochs, as shown in Figure 5. When compared to the LSTM model, the validation accuracy shows a little improvement, consistently hitting a 95% level. After the second epoch, the validation loss converges at about 0.2. The initial training loss of the BLSTM model is 0.8, and it eventually drops to less than 0.05. In conclusion, the smaller gap between training and validation loss indicates that the BLSTM model's bidirectional nature is better than the unidirectional LSTM model in capturing contextual information (Schuster & Paliwal, 1997).

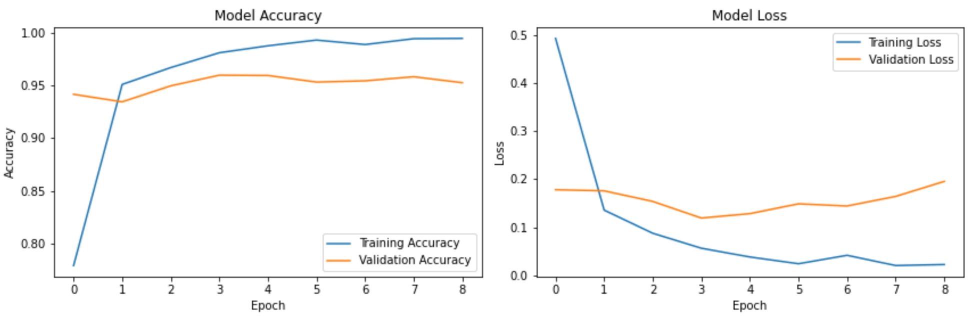


Figure 6. *BLSTMA Training Process*

As demonstrated in Figure 6, the performance of the BLSTM with Attention (BLSTMA) model across eight epochs indicates its effective learning and generalisation capabilities. The training accuracy (blue line) rapidly attains close to 99% by the second epoch and stabilises, while the validation accuracy (orange line) progressively increases, reaching approximately 96%, with a negligible discrepancy between the two. The training loss (BLUE line) displays a sharp decrease from 0.5 to around 0.02 by the eighth epoch, whereas the validation loss (ORANGE line) stabilises at approximately 0.18 (Vaswani et al., 2017). The narrow gap between the training and validation metrics reflects the BLSTMA model's ability to effectively generalise without overfitting. This ability is attributed to the inclusion of attention mechanisms, which enhance the model's focus on significant input features, as discussed in the following section. (Vaswani et al., 2017).

The confusion matrix of the BLSTMA model, which achieved the best results in the deep learning models, is shown in Figure 7. This shows that the model correctly classified 2670 negative samples (true negatives) and 583 positive samples (true positives). There were 60 false positives (negative samples misclassified as positive) and 53 false negatives (positive samples misclassified as negative). These results highlight the model's strong ability to handle imbalanced data with high precision and recall for both classes.

The differences in performance between the BLSTMA and other models likely stem from the attention mechanism, which enables the model to focus on the most relevant features of the input data. This targeted focus improves the model's capacity to capture subtle patterns, especially in complex datasets. Additionally, the use of FastText embeddings contributes to better word representation, capturing semantic and syntactic nuances. In comparison, models without attention mechanisms may struggle to distinguish between similar samples, leading to slightly higher misclassification rates.

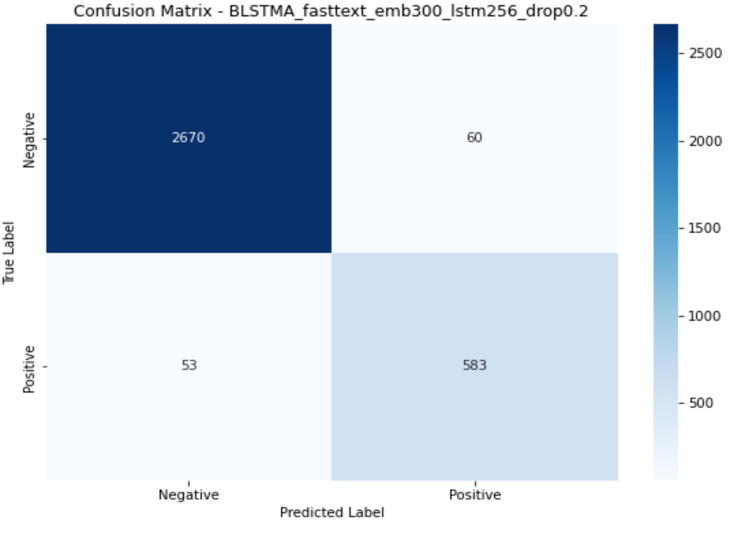


Figure 7. *Confusion Matrix of the BLSTMA Model*

The performance metrics of the deep learning models obtained with the best-performing parameters are presented in Table 3.

Table 3. *Performance comparison of the deep learning models*

Model Type	Embedding Type	Embedding Size	LSTM Unit Count	Dropout Rate	Accuracy	Precision	Recall	F1 Score
LSTM	Random	300	128	0.4	0.964646	0.889894	0.927673	0.908391
BLSTM	Random	300	256	0.4	0.965835	0.895296	0.927673	0.911197
BLSTMA	Fasttext	300	256	0.2	0.966429	0.906687	0.916667	0.911650

In this study, where approximately 162 different parameters were evaluated, the best results are presented in Table 3. To better analyse the performance of the different models, the average accuracy values were calculated based on four selected parameters and compared across the models.

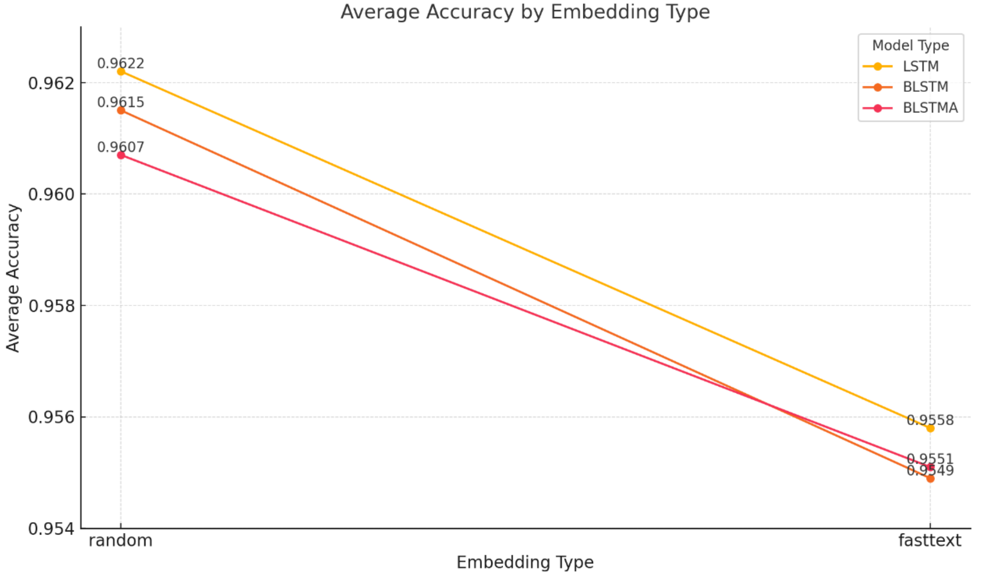


Figure 8. *Average accuracy values for embedding types*

In Figure 8, the BLSTMA model demonstrated the lowest average accuracy with the random embedding type but outperformed the BLSTM model when using the FastText embedding type. This indicates that BLSTMA, while struggling to learn meaningful representations from randomly initialized embeddings, benefits significantly more from pre-trained embeddings compared to BLSTM.

The LSTM model achieved the highest average accuracy of 96.22% with the random embedding type, surpassing the other models. This result indicates that LSTM is more effective at learning contextual relationships without relying on pre-trained embeddings. One possible reason is that LSTM's simpler structure, compared to bidirectional architectures such as BLSTM and BLSTMA, allows it to generalise better when embeddings are not pre-trained.

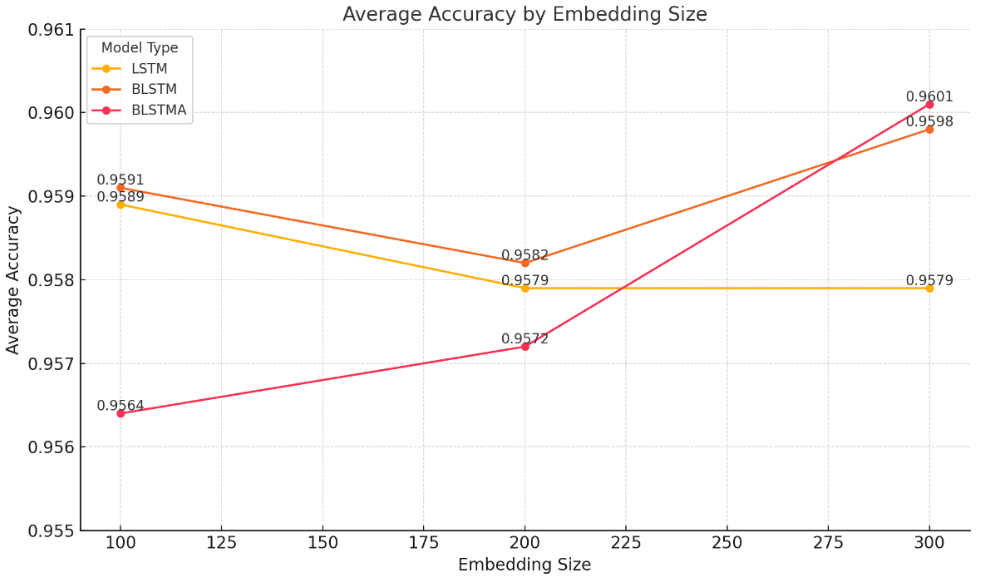


Figure 9. *Average accuracy values for embedding dimensions*

In Figure 9, the BLSTMA model exhibited lower accuracy rates compared to other models with smaller embedding dimensions. However, when the embedding dimension was set to 300, it surpassed the average accuracy values of the LSTM and BLSTM models. This result indicates that the BLSTMA model benefits more from larger embedding sizes due to its bidirectional structure, which can leverage richer feature representations when more parameters are available.

The LSTM model maintained a relatively stable performance as the embedding dimension increased, indicating that it does not significantly benefit from higher-dimensional word embeddings. Meanwhile, the BLSTM model showed an increasing trend, aligning with findings in previous studies that suggest that bidirectional architectures perform better with larger embedding sizes due to their ability to capture forward and backward dependencies more effectively.

The results show that the embedding dimension significantly affects how well the BLSTM and BLSTMA models provide accuracy. However, it is important to remember that overfitting may result from an overabundance of the embedding dimension augmentation. To evaluate their effect on model performance, the embedding dimensions of 100, 200, and 300 were carefully chosen for this work based on experimental considerations.

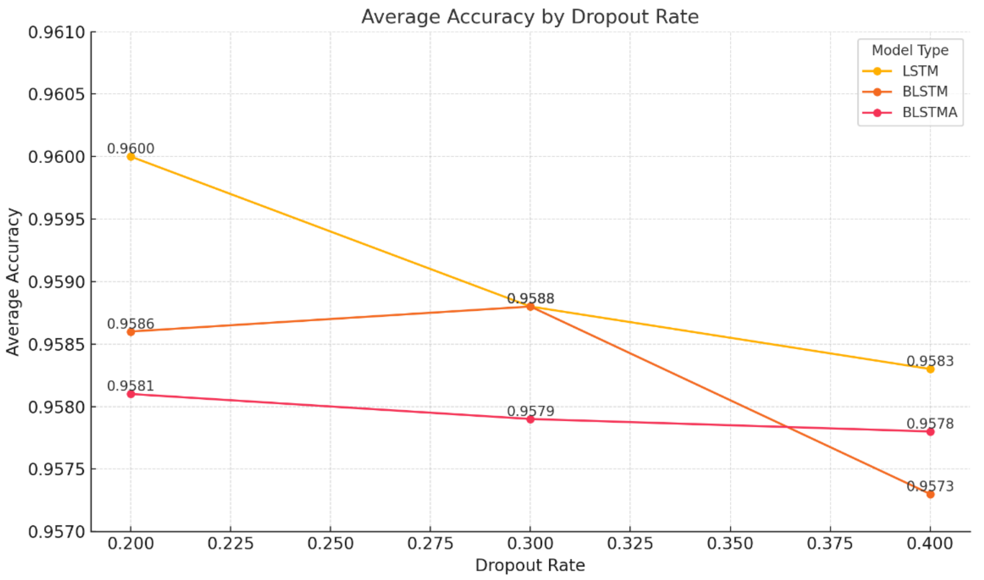


Figure 10. *Average accuracy values for dropout rates*

For the dropout rate parameter, values of 0.2, 0.3, and 0.4 were used. While the LSTM model exhibited a noticeable decline in accuracy beyond a dropout rate of 0.3, both the BLSTM and BLSTMA models began to show reductions in accuracy starting from 0.2. This indicates that the LSTM model experienced significant information loss after 0.3, whereas the BLSTM and BLSTMA models began losing information at 0.3.

The optimal parameter values for these models correspond to the points immediately before the decline in performance begins. As seen in Figure 10, the LSTM model fell below the BLSTMA model in accuracy at a dropout rate of 0.4. Among all the parameters, the BLSTM model consistently provided the best results. This demonstrates that the BLSTM model achieves better generalisation and exhibits a more balanced performance compared to the other models.

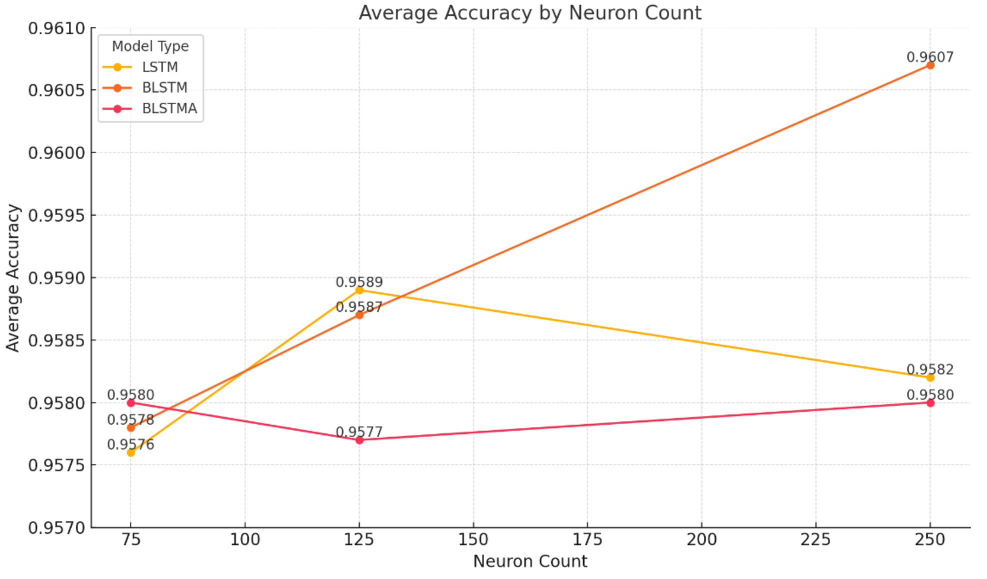


Figure 11. *Average accuracy values for the number of neurons*

As demonstrated in Figure 11, there is a clear indication of how the neuron count affects the performance of the LSTM, BLSTM and BLSTMA models. The BLSTM model's consistent enhancement in accuracy with rising neuron numbers can be ascribed to its capacity to effectively capture the bidirectional context. As the number of neurons escalates, the model evidently gains from the augmented capacity to depict intricate patterns, culminating in superior performance compared to the LSTM and BLSTMA models This indicates that the bidirectional structure in the BLSTM scales effectively with more computational resources. In contrast, the BLSTMA model demonstrates less dependency on the neuron count, achieving competitive accuracy with fewer neurons but failing to leverage additional neurons to the same extent as BLSTM. This could indicate that the attention mechanism primarily enhances local feature learning rather than relying heavily on the increased model capacity. The findings further indicate that the LSTM model demonstrates a decline in accuracy beyond 128 neurons, which may suggest that it experiences overfitting or diminishing returns when scaled due to its lack of bidirectional or attention-enhanced mechanisms to effectively utilise the additional neurons.

The observed differences between these algorithms underscore the merits of the bidirectional context in BLSTM in leveraging higher neuron counts, while the BLSTMA model's attention mechanism underscores efficiency with fewer resources. These observations underscore the necessity for selecting a model architecture that considers not only the characteristics of the data set but also the available computational resources.

Table 4. *Machine Learning Models Results*

Model Type	Vectorisation Method	Accuracy	Precision	Recall	F1 Score
Logistic Regression	TF-IDF	0.9465	0.9454	0.9465	0.9453
Naïve Bayes	TF-IDF	0.9435	0.9438	0.9435	0.9436
SVM	TF-IDF	0.9530	0.9528	0.9530	0.9529
Random Forest	TF-IDF	0.9533	0.9529	0.9533	0.9531
XGBoost	TF-IDF	0.9566	0.9563	0.9566	0.9564
Logistic Regression	Bag Of Words	0.9527	0.9523	0.9527	0.9525
Naïve Bayes	Bag Of Words	0.9319	0.9457	0.9319	0.9351
SVM	Bag Of Words	0.9581	0.9578	0.9581	0.9579
Random Forest	Bag Of Words	0.9569	0.9566	0.9569	0.9567
XGBoost	Bag Of Words	0.9584	0.9585	0.9584	0.9584
Logistic Regression	Word2Vec	0.9212	0.9194	0.9212	0.9200
SVM	Word2Vec	0.9292	0.9311	0.9292	0.9300
Random Forest	Word2Vec	0.9367	0.9369	0.9367	0.9368
XGBoost	Word2Vec	0.9393	0.9397	0.9393	0.9395

These results offer a useful opportunity to investigate how well various algorithms perform when classifying tweets of emergencies. Strong accuracy is demonstrated by machine learning models like XGBoost and SVM, which achieve above 95% accuracy when the Bag of Words (BoW) and TF-IDF approaches are used. This implies that word frequency-based techniques are useful for identifying emergency messages, most likely due to the fact that urgent tweets frequently follow particular patterns. Word2Vec-based models' comparatively worse performance of the Word2Vec-based models raises the possibility that pre-trained word embeddings could not adequately convey the context or urgency of these signals. Although SVM and XGBoost perform well, they may have trouble understanding more complex or context-dependent emergency tweets because of their emphasis on basic frequency patterns.

On the other hand, the deep learning models—BLSTMA with FastText embeddings, in particular—performed better, achieving 96.64% accuracy. This implies that FastText embeddings help the model understand context, allowing it to identify crises even in various language usages. Deep learning models also have stronger recall, which lowers the likelihood of overlooking important signals. Deep learning models exhibit a remarkable ability to navigate the intricate and informal nature of Twitter discourse, which improves their dependability for real-world emergency detection, even though machine learning models are more efficient in terms of processing speed and efficiency.

Conclusion

The results showed that while the BLSTMA model had the best accuracy (96.64%) and F1 score (0.9116), conventional machine learning techniques like XGBoost and SVM. Using Bag of Words vectorisation, SVM obtained 95.81% accuracy and an F1 score of 0.9579, whereas XGBoost earned 95.84% accuracy and an F1 score of 0.9584. By demonstrating the usefulness of the BLSTMA model in real-time disaster response and the complementary advantages of conventional approaches in the analysis of complex disaster data, these findings highlight the significance of customising machine learning and deep learning approaches to particular tasks.

Even though deep learning models demonstrated excellent generalisation skills, the need for further improvement is highlighted by their comparatively poorer accuracy and recall compared to conventional techniques. Additionally, the study's conclusions are not as broadly applicable to other languages or catastrophe situations due to its dependence on a dataset of Turkish tweets. To improve these models' worldwide applicability, future studies should examine how well they can adapt to multilingual datasets and various settings.

Furthermore, there are difficulties in processing large data streams (e.g., the vast number of messages shared on social media during disaster events), noisy data (e.g., misleading, incomplete, or irrelevant information within the shared messages), and the requirement for quick categorisation (e.g., the need to analyse and classify data in real time to support immediate decision-making) when implementing these models in real-time disaster response systems. Overcoming these obstacles requires combining real-time processing capabilities (e.g., enabling models to process and classify incoming data as it is collected) and increasing computing efficiency (e.g., optimising the speed and resource usage of models for faster performance). By addressing these issues, it will be possible to create frameworks for evaluating social media data connected to disasters that are more reliable and scalable, greatly improving emergency response and disaster management initiatives throughout the globe.

Ethics Committee Approval: Ethics approval was not required, as the analysis relied solely on publicly available Twitter data, ensuring anonymity and no use of personal information.

Peer-review: Externally peer-reviewed.

Author Contributions: Conception/Design of Study- M.S.; Data Acquisition- M.S.; Data Analysis/Interpretation- M.S., A.Ö.; Drafting Manuscript- M.S., A.Ö.; Critical Revision of Manuscript- M.S., A.Ö.; Final Approval and Accountability- M.S., A.Ö.; Technical or Material Support- M.S., A.Ö.; Supervision- M.S., A.Ö.

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: The authors declared that this study has received no financial support.

References

Al Jazeera (2023). “Quake victims stuck under rubble take to social media for help.” Available: https://www.aljazeera.com/news/2023/2/7/quake-victims-stuck-under-rubble-take-to-social-media-for-help

Euronews (2023). “How Twitter helped find survivors trapped beneath rubble after Turkey’s earthquakes.” Available: https://www.euronews.com/next/2023/02/10/how-twitter-helped-find-survivors-trapped-beneath-rubble-after-turkeys-earthquakes

Powers, C. J., Devaraj, A., Ashqeen, K., Dontula, A., Joshi, A., Shenoy, J., & Murthy, D. (2023). Using artificial intelligence to identify emergency messages on social media during a natural disaster: A deep learning approach. *International Journal of Information Management Data Insights*, *3*(1), 100164. https://doi.org/10.1016/j.jjimei.2023.100164

Bhere, P., Upadhyay, A., Chaudhari, K., & Ghorpade, T. (2020). Classifying Informatory Tweets during Disaster Using Deep Learning. *ITM Web of Conferences*, *32*, 03025. https://doi.org/10.1051/itmconf/20203203025

Kumar, A., Singh, J. P., & Saumya, S. (2019). A Comparative Analysis of Machine Learning Techniques for Disaster-Related Tweet Classification. In *2019 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)* (pp. 1-6). IEEE. https://doi.org/10.1109/R10-HTC47129.2019.9042443

Behl, S., Rao, A., Aggarwal, S., Chadha, S., & Pannu, H. S. (2021). Twitter for disaster relief through sentiment analysis for COVID-19 and natural hazard crises. *International Journal of Disaster Risk Reduction, 55*, 102101. https://doi.org/10.1016/j.ijdrr.2021.102101

Madichetty, S., & Sridevi, M. (2020). A stacked convolutional neural network for detecting the resource tweets during a disaster. *Multimedia Tools and Applications, 80*, 3927-3949. https://doi.org/10.1007/s11042-020-09873-8

Sit, M.A., Koylu, C., & Demir, I. (2019). Identifying disaster-related tweets and their semantic, spatial and temporal context using deep learning, natural language processing and spatial analysis: a case study of Hurricane Irma. *International Journal of Digital Earth, 12*, 1205-1229. https://doi.org/10.1080/17538947.2018.1563219

Mijwel, M. M. (2015, April). *History of Artificial Intelligence*. *Computer Science*, 3–4

Kang, Y., Cai, Z., Tan, C., Huang, Q., & Liu, H. (2020). Natural language processing (NLP) in management research: A literature review. *Journal of Management Analytics*, *7*(2), 139–172. https://doi.org/10.1080/23270012.2020.1756939

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press. (pp. 1–20)

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review, 65*(6), 386–408. https://doi.org/10.1037/h0042519

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press. (pp. 168–171).

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press. (pp. 373–383).

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Graves, A., Mohamed, A.-R., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 6645–6649). IEEE. https://doi.org/10.1109/ICASSP.2013.6638947

Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., & Xu, B. (2016). Attention-based bidirectional long short-term memory networks for relation classification. In K. Erk & N. A. Smith (Eds.), *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 207–212). Association for Computational Linguistics. https://doi.org/10.18653/v1/P16-2034

Zhong, G., Lin, X., Chen, K., Li, Q., & Huang, K. (2020). Long short-term attention. In *Lecture Notes in Computer Science* (pp. 45–54). Springer. https://doi.org/10.1007/978-3-030-39431-8_5

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," in *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673-2681, Nov. 1997, doi: 10.1109/78.650093

Vaswani, A., Shazeer, N.M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. *Neural Information Processing Systems*.

Journal of Data Applications
Issue: 3, 2024
E-ISSN: 2980-3357

RESEARCH ARTICLE

Investigation of the Reasons for Failure in Blockchain Applications Using Bibliometric Analysis

Muzaffer ŞENLİK[1], Melih ENGİN[2]

DOI: 10.26650/JODA.1615884

ABSTRACT

The utilisation of blockchain technology has gained significant traction across multiple sectors, notably finance, healthcare, supply chain management and real estate, by offering decentralised, secure and transparent data management systems. Introduced in 2008 with Bitcoin, this technology has garnered attention for its smart contracts, various data tracking applications, and cryptocurrencies. The decentralised structure and high security features of the blockchain have led to its emergence as an alternative to traditional systems. The present study aims to examine the reasons for the failures encountered in blockchain applications through bibliometric analysis by reviewing and analysing the extant literature to reveal the conditions and reasons for blockchain technology's failure. The researchers obtained a dataset containing 2779 documents by searching the Scopus database using the keywords 'blockchain' and 'failure'. These data were analysed using Biblioshiny, a bibliometric analysis tool. The analysis provides a comprehensive statistical perspective by revealing quantitative data such as the number of publications on the research topic, the number of citations by years, the number of publications by countries, the most cited documents and their abstracts, the word cloud, the periodic distribution of the most used words, the thematic map of keywords and the association network. The bibliometrix library of the R programming language was utilised to analyse and visualise the data. The results indicate that factors such as scalability issues, cross-platform incompatibility, network security concerns, lack of user experience, inadequate real-world testing, and legal uncertainties complicate the implementation of blockchain technology. This analysis provides findings to guide future blockchain applications.

Keywords: Blockchain, Failure, Bibliometric Analysis

* This study was produced from the master's thesis prepared by Muzaffer ŞENLİK under the supervision of Dr. Melih ENGİN.

[1]Bursa Uludağ University, Management Information Systems Scientist, Bursa, Türkiye
[2]Bursa Uludağ University, İnegöl Business Faculty, Management Information Systems, Bursa, Türkiye

ORCID: M.Ş. 0009-0006-0341-4622; M.E. 0000-0002-4953-6119

Corresponding author:
Muzaffer ŞENLİK,
Bursa Uludağ University, Management Information Systems Scientist, Bursa, Türkiye
E-mail: muzaffer.senlik96@gmail.com

Submitted: 08.01.2024
Accepted: 22.01.2024
Published Online: 30.01.2025

Citation: Şenlik, M., & Engin, M. (2024). Investigation of the Reasons for Failure in Blockchain Applications Using Bibliometric Analysis. *Journal of Data Applications, 3*, 85-105.
https://doi.org/10.26650/JODA.1615884

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International License

Introduction

The concept of blockchain technology was first introduced by Satoshi Nakamoto in 2008 with the publication of his article 'Bitcoin: Peer-to-Peer Electronic Cash Payment System' (Nakamoto, 2008). Researchers distinguish this technology by the encryption of records held by several network participants in a distributed database, as opposed to a centralised database. The records are individually labelled at the time of generation and cannot be modified retroactively. Each new transaction is verified by numerous participants and occurs autonomously, negating the necessity for intermediaries (Nakamoto, 2008). Consequently, blockchain technology can be used effectively across diverse sectors.

The growing interest in blockchain technology has also increased academic research examining the potential applications and consequences of this technology. A number of studies have been conducted in a variety of disciplines, including finance, information technology, law, healthcare and logistics, which examine the potential benefits of blockchain technology for different sectors. However, some academic studies make exaggerated claims about blockchain technology and offer untested solutions (Zile & Strazdina, 2018). This situation emphasises the necessity for future research on blockchain technology to be based on more concrete findings. In this context, the aim of this research is to examine the reasons for the failures in blockchain applications and to provide guidance for future applications.

Literature Review

Despite the fact that blockchain technology was introduced to the market as a cryptocurrency, with the rapid development of technology, it has become a technological development with applications in many different sectors, not only as a financial transaction. It offers revolutionary solutions in many areas, from health services to supply chain management, from public services to production (Ceylan & Işık, 2023).

Decentralised Finance (DeFi)

Decentralised finance (DeFi) is a structure developed with blockchain technology that offers a financial ecosystem accessible to all by eliminating intermediary institutions. The digitalisation of traditional financial services has enabled activities such as deposits, loans, insurance and stock exchange transactions to be conducted without the need for a central authority. Applications such as cryptocurrencies, non-fungible tokens (NFTs) and metaverses are platforms built on blockchain technology that meet the need for decentralisation. DeFi applications facilitate financial transactions in a secure and transparent manner through the utilisation of smart contracts and decentralised applications. This facilitates direct interaction between users and the platforms, enabling them to engage in financial decision-making processes without the involvement of intermediaries (Parlar, 2022).

Health sector

It is evident that certain domains within the healthcare sector hold considerable potential for technological transformation facilitated b78y blockchain protocols. This section discusses the current and potential uses of blockchain protocols in healthcare. The sub-use scenarios encompass health information transfer, health research integrity, personal health records, health data storage, billing, damage records, drug supply chain, and pandemic situations (Aydar & Çetin, 2020).

Public/Government Services

A plethora of states, international organisations and the private sector are closely monitoring blockchain technology, conducting research activities and developing various projects. A diversity of countries and organisations are employing a combination of a competitive and collaborative approach to the utilisation of blockchain in domains other than cryptocurrencies. For instance, the US state of Delaware has adopted blockchain technology for the purpose of company incorporations, while Sweden, in collaboration with banking institutions and land registry authorities, is undertaking the testing of a blockchain-based land registry application that enables buyers and sellers to view and approve transactions in real time (Tüfekçi & Karahan, 2019).

Property Sector

Blockchain technology is a decentralised system that records transactions in a secure and unalterable way thanks to its distributed structure. This technology can create a paradigm shift within the domain of land and real estate records. The integration of blockchain technology into land registry records has the potential to enhance transparency and reliability in the tracking of property rights. Furthermore, it has the capacity to expedite transfer processes and reduce intermediary costs. The transparency characteristic of blockchain technology ensures that all records are accessible to the public, thereby equalising access to information and enhancing transparency within the property market. Consequently, this will contribute to the creation of a fairer and more competitive market (Atzori, 2018).

Supply Chain Sector

The blockchain technology utilised ensures the immutability of records, thereby guaranteeing complete transparency concerning the provenance, quality and reliability of the products. The use of smart contracts enables the automation of supply chain processes, facilitating more expeditious and dependable payments and product deliveries. Furthermore, the use of blockchain-based platform fosters enhanced data sharing among supply chain partners, thereby optimising operational efficiency and reducing costs (Treiblmaier, 2018).

Production

The supply chain, from the procurement of raw materials to the delivery of the final product to the end user, is meticulously recorded on the blockchain in a secure and unalterable manner. This provides a comprehensive and transparent account of the origin of the products, the materials utilised and the manufacturing conditions. Consequently, the prevalence of counterfeit and substandard products can be mitigated, thereby empowering consumers to make informed decisions. The utilisation of smart contracts facilitates the automation of production processes and expedites transactions between suppliers, manufacturers and distributors, ensuring reliability (Güven, 2023).

Education

The potential of blockchain technology in the context of educational applications is significant, particularly in domains that require a reliable information infrastructure. This includes the recognition of learning histories, the management of open and distance education courses, the processing of on-campus applications, and the implementation of learning management systems. In all these contexts, the primary requirement is the creation of secure systems that individuals can use in the future. In particular, learning management systems and trust-based systems are potential application areas of blockchain technology in education. Furthermore, the integration of blockchain technology with other technologies, such as the Internet of Things, augmented reality, and artificial intelligence, can enhance open and distance learning environments. This integration can enhance the learning experience by making it more engaging, personalised and effective (Yıldırım, 2018).

Method

In this study, the data obtained from the Scopus database were subjected to a rigorous filtering process in accordance with the criteria delineated in Table 1, ensuring their suitability for the research objectives. Subsequently, the data underwent a comprehensive analysis through the lens of bibliometrics. A bibliometric analysis provides a comprehensive statistical perspective, revealing data such as the number of publications, frequency of citations, and authors’ countries (Donthu et al., 2021). This qualitative research method is interpreted and provides meaningful results. The bibliometrix library of the R programming language was used for the analysis and visualisation of the data obtained (Aria & Cuccurullo, 2017).

Table 1. *Flowchart for the determination of data.*

The following search was conducted on the Scopus database:	(TITLE-ABS-KEY (blockchain OR block-chain) AND TITLE-ABS-KEY (fail OR failure))
	The aforementioned query was employed in the search of all relevant publications.
	Results: 3089 documents.
Language Selection	Documents written in languages other than English were not included in this study. The following languages are excluded: Chinese (77), Russian (3), Portuguese (2), Korean (2), Spanish (1), and Japanese (1). This equates to 86.
	The results yielded a total of 3,003 documents.
2011-2023 Elections	The research conducted in all years revealed that the initial period started in 2011.
	The current year was constrained to 2023 by subtracting 2024.
	The results yielded 2,779 documents.

Method

The study was conducted using the Scopus database, which includes a range of terms related to blockchain and failure. These terms were used in conjunction with one another to identify relevant research. The data extracted from the Scopus database was obtained through text mining using the Biblioshiny library of the RStudio program.

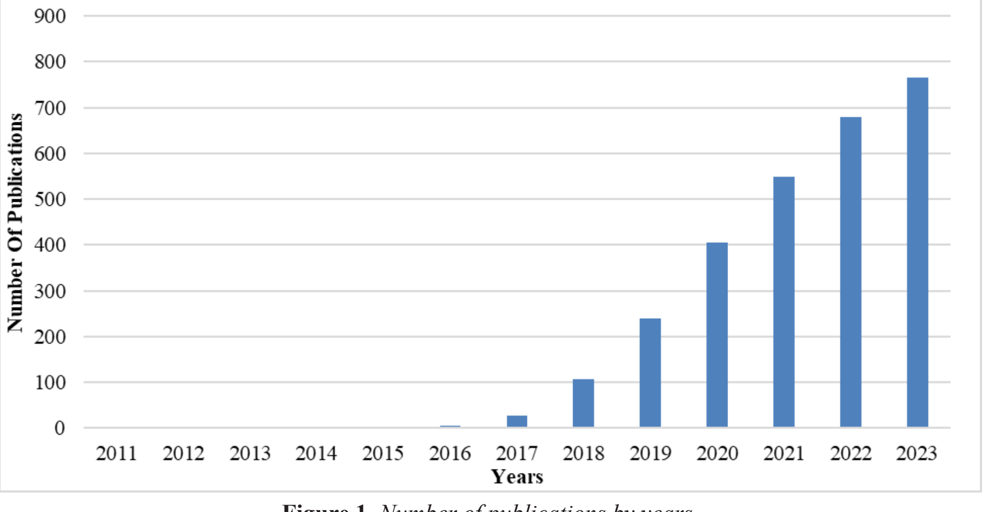


Figure 1. *Number of publications by years.*

Figure 1 illustrates the number of publications by years. Accordingly, the initial notable increase occurred in 2017. In 2018, there were more than 100 publications. There has been a notable surge in activity, particularly over the past five years. This upward trajectory has continued unabated annually. In 2023, the highest number of publications was recorded. The upward trajectory started with 27 publications in 2017 and culminated in 765 annual publications by 2023. Figure 1 illustrates that the number of errors or failure examples of blockchain applications has increased. The growth in research activity in this area is of significant value to both developers of blockchain applications and businesses seeking to use this technology. By analysing the areas of weakness identified in the aforementioned examples of failed blockchain applications, it is possible to gain insight into the key challenges and potential pitfalls that must be addressed to ensure the successful deployment of blockchain technology in practice.

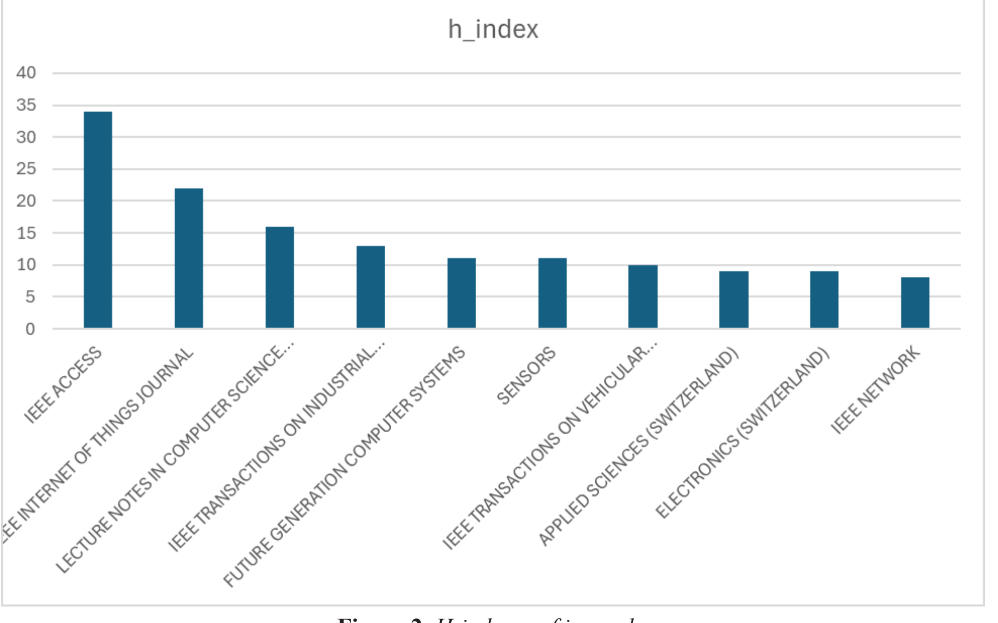


Figure 2. *H-indexes of journals.*

Figure 2 presents a comparative analysis of the average number of citations received by the publications under examination. This comparison allows for evaluating the academic impact of different publications. It can be stated that the publications of IEEE Access receive a significantly higher number of citations than other sources, indicating a greater level of influence. However, it should be noted that some publications, such as Lecture Notes in Computer Science, may have a lower number of citations due to the inclusion of more lecture notes and conference proceedings. While such publications are not direct research outputs, they play a crucial role in supporting the accumulation of knowledge and academic discourse.

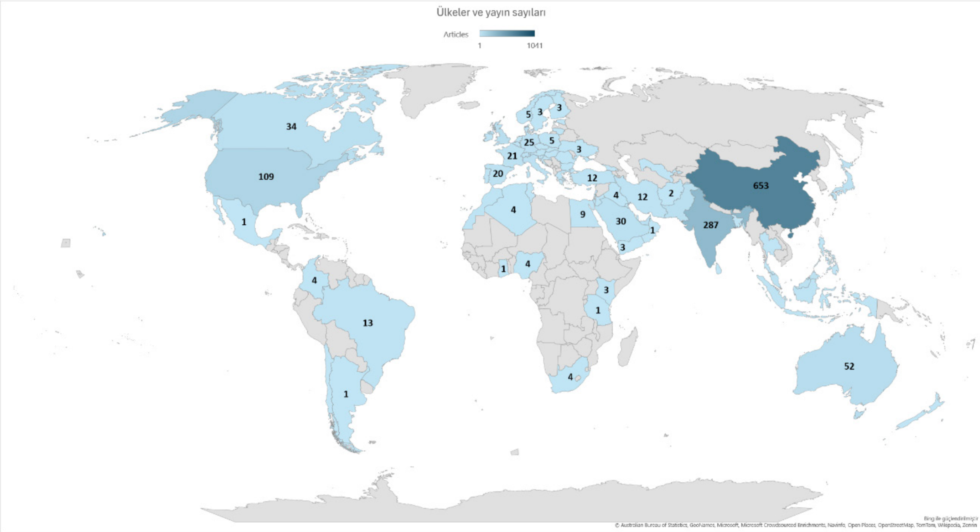


Figure 3. *Number of publications of countries.*

Table 2. *Number of publications in the top 10 countries and distribution of single (SCP) and multi (MCP) corresponding authors by Country.*

Country	Article	SCP	MCP
CHINA	653	500	153
INDIA	287	240	47
USA	109	76	33
AUSTRALIA	52	24	28
KOREA	49	33	16
UNITED KINGDOM	42	20	22
United Arab Emirates	38	26	12
CANADA	34	21	13
PAKISTAN	34	9	25
TOTAL	**1041**	**823**	**218**

As illustrated in Table 2 and Figure 3, China is the country with the highest number of publications on blockchain and failure. Of the 653 publications originating from China, 500 were single-authored, while 153 were multi-authored. This indicates that researchers in China engage in both independent and collaborative research activities. While India ranks second with 287 publications, 240 of these are single-authored and 47 are multi-authored. This indicates that blockchain and failure research in India is primarily conducted as individual studies. In the United States, 76 of the 109 publications are single-authored, while 33 are multi-authored. In other countries, although the number of publications is lower, it is generally observed that multi-author publications exceed single-author publications. This demonstrates that international collaboration on blockchain and failure-related matters is pervasive, with researchers from diverse countries engaged in joint endeavours, disseminating and exchanging their insights and expertise.

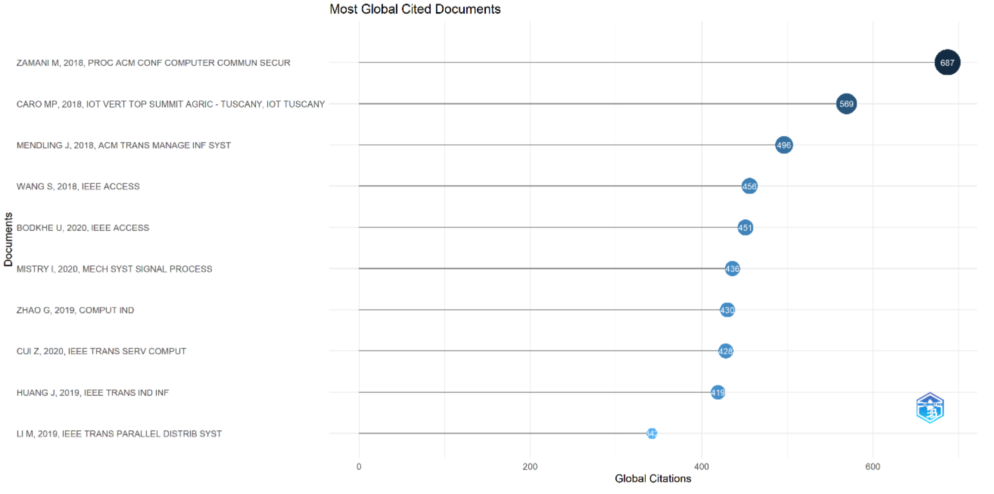


Figure 4. *Most globally cited documents*

Table 3. *Review of the most cited articles*

Article Name	Authors	Abstracts
RapidChain: Scaling the Blockchain via Full Sharding	Mahdi Zamani, Mahnush Movahedi, and Mariana Raykova.	This paper introduces RapidChain, the inaugural Byzantine fault-tolerant public blockchain protocol that employs comprehensive sharding to overcome the performance and scalability constraints of extant blockchain protocols. RapidChain fully shards the transaction processing, communication, computation, and storage overhead without assuming a trusted setup (Zamani et al., 2018).
Blockchain-based Traceability in Agri-Food Supply Chain Management: A Practical Implementation	Miguel Pincheira Caro, Muhammad Salek Ali, Massimo Vecchio, and Raffaele Giaffreda.	This paper presents a fully decentralised, blockchain-based traceability solution for agriculture and food supply chain management, designated as AgriBlockIoT. The solution was evaluated and benchmarked using the Ethereum and Hyperledger Sawtooth blockchain applications (Caro et al., 2018).

Table 3 (Continue). *Review of the most cited articles*

Article Name	Authors	Abstracts
Blockchains for Business Process Management: Challenges and Opportunities	Jan Mendling, Ingo Weber, Wil Van Der Aalst, Jan Vom Brocke, Cristina Cabanillas, Florian Daniel, Søren Debois, Claudio Di Ciccio, Marlon Dumas, Schahram Dustdar, Avigdor Gal, Luciano García-Bañuelos, Guido Governatori, Richard Hull, Marcello La Rosa, et al.	This paper provides a summary of the challenges and opportunities presented by blockchain technology for business process management (BPM). It discusses the potential for blockchains to be used within the traditional BPM lifecycle and considers their potential to become a significant factor in BPM beyond the scope of BPM. (Mendling et al. 2018).
A Blockchain-Based Framework for Data Sharing With Fine-Grained Access Control in Decentralised Storage Systems	Shangping Wang, Yinglong Zhang, and Yaling Zhang.	This paper examines the data storage and sharing scheme for decentralised storage systems and proposes a framework that integrates the decentralised storage system, the interplanetary file system, the Ethereum blockchain and the ABE technology. In this framework, the data owner is able to encrypt shared data by defining the distribution of secret keys and access policy for data users, and the scheme provides detailed control over data access (Wang et al., 2018).
Blockchain for Industry 4.0: A Comprehensive Review	Umesh Bodkhe, Sudeep Tanwar, Karan Parekh, Pimal Khanpara, Sudhansu Tyagi, Neeraj Kumar, and Mamoun Alazab.	This paper presents a systematic review of various blockchain-based solutions and their applicability in various Industry 4.0-based applications. It explores the latest solutions in blockchain technology for smart applications, demonstrates the reference architecture used for blockchain applicability in Industry 4.0 applications, discusses the advantages and disadvantages of traditional security solutions in comparison with their countermeasures, and compares existing blockchain-based security solutions using various parameters (Bodkhe et al., 2020).
Blockchain for 5G-enabled IoT for industrial automation: A systematic review, solutions, and challenges	Ishan Mistry, Sudeep Tanwar, Sudhansu Tyagi, and Neeraj Kumar.	This paper provides an overview of the integration of 5G-enabled IoT into blockchain-based industrial automation and discusses its potential industrial applications. Furthermore, it addresses the issues of scalability, interoperability, and other research challenges in 5G-enabled IoT for blockchain applications (Mistry et al., 2020).
Blockchain technology in agri-food value chain management: A synthesis of applications, challenges and future research directions	Guoqing Zhao, Shaofeng Liu, Carmen Lopez, Haiyan Lu, Sebastian Elgueta, Huilan Chen, and Biljana Mileva Boshkoska.	This study makes a significant contribution to the existing literature on the applications, challenges, and future research directions of blockchain technology in agri-food value chain management. The findings demonstrate that blockchain technology, in conjunction with advanced information and communication technology and the Internet of Things, has been adopted in four main areas (traceability, information security, production and sustainable water management) for the enhancement of agri-food value chain management (Zhao et al., 2019).

Table 3 (Continue). *Review of the most cited articles*

Article Name	Authors	Abstracts
A Hybrid BlockChain-Based Identity Authentication Scheme for Multi-WSN	Zhihua Cui, Fei Xue, Shiqiang Zhang, Xingjuan Cai, Yang Cao, Wensheng Zhang, and Jinjun Chen.	This paper proposes a blockchain-based authentication scheme for multiple wireless sensor networks (WSNs) in the context of the Internet of Things (IoT). A hierarchical network is constituted by the division of IoT nodes into three categories: base stations, cluster head nodes, and ordinary nodes, according to their differing capabilities. A hybrid blockchain model comprising a local chain and a public chain between different types of nodes is established. In this hybrid model, the identity of the nodes is authenticated in various communication scenarios (Chu et al., 2020).
CrowdBC: A Blockchain-Based Decentralised Framework for Crowdsourcing	Ming Li, Jian Weng, Anjia Yang, Wei Lu, Yue Zhang, Lin Hou, Jia-Nan Liu, Yang Xiang, and Robert H. Deng.	This paper introduces CrowdBC, a decentralised crowdsourcing framework. CrowdBC permits a group of workers to complete a requester's task without the necessity of entrusting it to a third party. It ensures user privacy and necessitates only minimal transaction fees. The framework provides a concrete scheme where smart contracts are used to facilitate the entire crowdsourcing process, including tasks, rewards, and other aspects. The usability and scalability of CrowdBC have been demonstrated through the implementation of a prototype on the Ethereum tested (Li et al., 2019).
Towards Secure Industrial IoT: A Blockchain System With a Credit-Based Consensus Mechanism	Junqin Huang, Linghe Kong, Guihai Chen, Min-You Wu, Xue Liu, and Peng Zeng.	This paper presents a blockchain system with a credit-based consensus mechanism for the Industrial Internet of Things (IIoT). The proposed mechanism is a credit-based proof-of-work (PoW) mechanism for power-constrained IoT devices, which can simultaneously guarantee security and transaction efficiency (Huang et al., 2019).

Figure 4 shows the names of the 10 most cited authors and the number of citations. Furthermore, also Table 3 shows the names, authors, and abstracts of the 10 most cited documents. The most cited study is Mahdi Zamani et al. with 687 citations and it introduces RapidChain. The top 10 most cited documents were written between 2018 and 2020.

Figure 5. *Index keyword cloud (Keywords Plus)*

Table 4. *Top 10 words in ındex keyword cloud*

Words	f
Internet of Things	524
Network security	461
Digital storage	350
Smart contract	292
Authentication	276
Decentralised	272
Cryptography	245
Single point	231
Security	201
Knowledge Management	196

As illustrated in Figures 5 and Table 4, the journals in question have identified a set of keywords that are used to categorise the publications. The keywords were obtained by content analysis of the articles and are therefore independent of the author's determination. They are defined as keywords plus. The terms "blockchain" and "block-chain" have been excluded from the list of keywords. The rationale for the exclusion of these terms is that the term "blockchain" lacks semantic coherence, as it is rendered superfluous by the existence of alternative equivalents in the literature. Table 4 presents the ten most frequently occurring terms. The term “Internet of Things” (IoT) is referenced 524 times. This demonstrates a notable interest in the conjunction of blockchain and the concept of the Internet of Things. As the concepts of failure and error are also included in the search, it can be assumed that most applications within this field may be defective, with examples of failed applications readily available. The second most frequently occurring term is “network security,” which was referenced 461 times. This is pertinent to the security aspect, which is a fundamental feature of blockchain technology. Blockchain technology is employed to enhance security. It is evident that this is a highly preferred option. It may also be assumed that some applications in question are unsuccessful because of being searched with the concepts of failure and error. The third word is “digital store.” The term was employed 350 times. The preference for storing digital data with blockchain technology is based on the assumption that it is a secure method of data storage. If the storage of important data in this way is not adequately planned, it can result in significant financial implications. The fourth term smart contracts represents a fundamental pillar of the blockchain, with a documented usage of 292 instances. The development of smart contracts requires collaboration between individuals and businesses. The irreversible nature of the blockchain recording system underscores the necessity for businesses and individuals to exercise caution when preparing such contracts. Errors or failures have been observed in this context. A review of the remaining six words reveals that the first is “authentication,” occurring 276 times. The seventh word is “Decentralised.” The term was referenced 272 times. The term “cryptography” is referenced 245 times, “single point” is mentioned 231 times, and “security” is discussed 201 times. These words share a commonality in that they pertain to the security of data. The necessity of encrypting data and avoiding the storage of data at a single point is underscored. In the absence of these considerations during the creation of a blockchain, the security of the data may be compromised, and errors may be introduced. The term “information management” is referenced 196 times. The decentralised structure of blockchain technology and the formation of secure blocks are important concerning information management. In this context, an error that may occur or an unsuccessfully developed application may present challenges to the effective management of information.

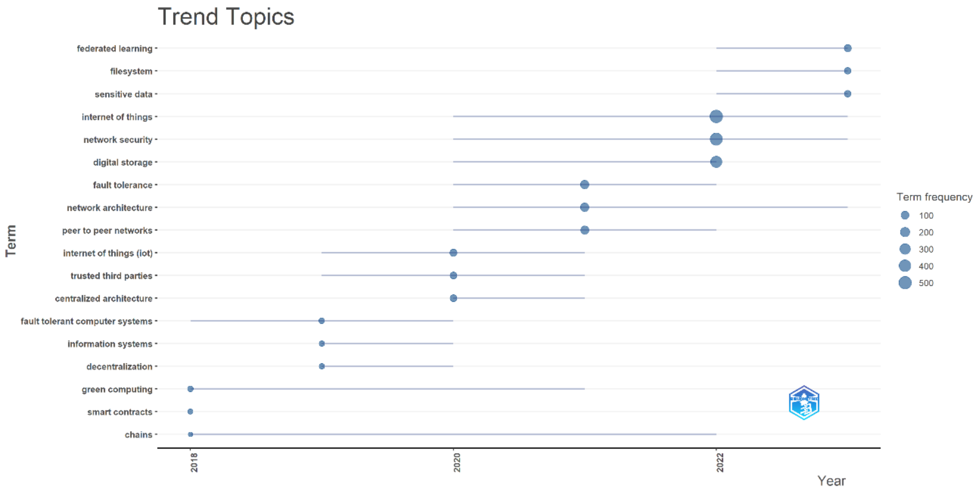


Figure 6. *Periodic distribution of the most used concepts (trend topics)*

Figure 6 illustrates the variable frequencies of concepts in publications employing the terms “blockchain” and “fail” and “failure” according to the year of publication. It is evident that certain concepts have declined in usage over time, while new concepts have emerged and gained prominence. While the frequency of certain concepts, such as “security” and “decentralisation,” was high in 2018, it subsequently declined towards 2022. This illustrates the maturation of blockchain technology over time, accompanied by a shift in the focus of researchers from fundamental concepts such as security and decentralisation towards more specific and application-oriented issues. The emergence of concepts such as “sensitive data,” “unified learning,” and “file system” recently indicates the advent of novel applications and potential issues associated with blockchain technology. The storage and processing of sensitive data in blockchain technology presents novel challenges with regard to the protection of privacy and the assurance of security. The emergence of concepts such as unified learning and the file system demonstrates the potential of blockchain technology to be utilised in a multitude of domains, while also highlighting the potential challenges and failure scenarios that may arise. These developments demonstrate that blockchain technology is undergoing constant evolution and that its integration into diverse sectors is ongoing.

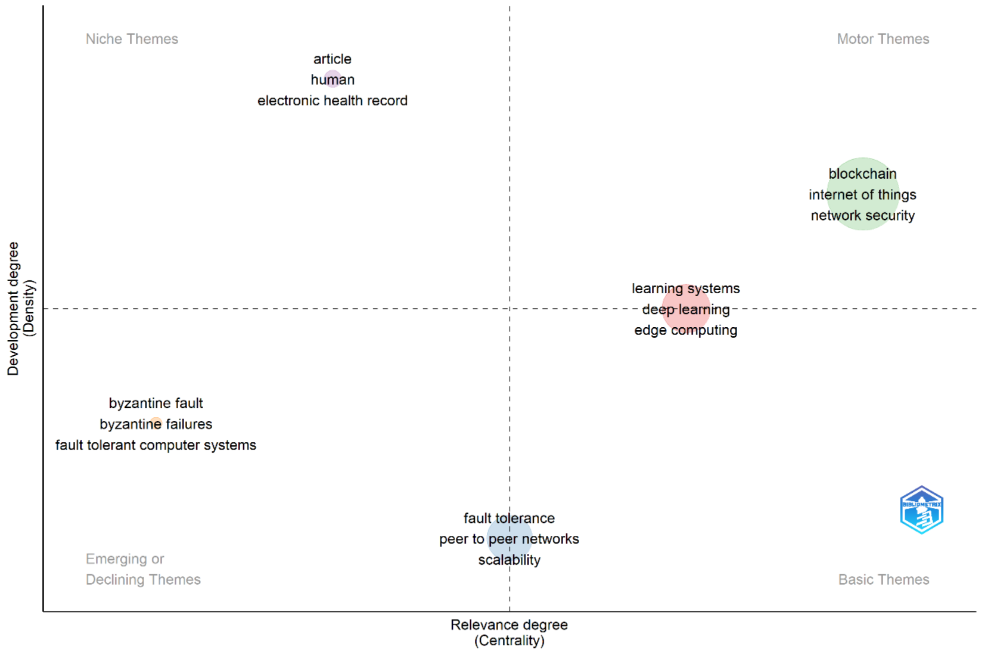


Figure 7. *Index keywords thematic map.*

Figure 7 illustrates the thematic map, which forms a cluster comprising the KeyWord Plus indexes of the publications in the dataset that emerged by searching the terms (fail) and "failure" together with the blockchain. In this instance, the terms "blockchain" and "Blockchain" are once again combined. Thematic map analysis is a form of word analysis that focuses on the examination of auxiliary words. Thematic representation of a research area is achieved through its ecentraliz (Waltman & Van Eck, 2013). As illustrated in Figure 7, the terms employed in this investigation have been classified into five distinct categories. The word groups were observed to fall into one of four categories: niche, motor, basic, and rising or falling themes.

The terms "article," "human," and "electronic health record" in the "niche themes" category in the upper left quadrant are topics that have undergone significant development but are of limited relevance. It is possible that the article and people themes will not be concentrated in specific research areas, given that they are generally very broad and abstract topics. Although the electronic health record is a significant topic within the healthcare sector, it may not be as central a theme as other technological innovations. These topics appeal to specific areas of expertise and thus do not receive widespread attention and are therefore ecentraliz as niche themes.

The upper right corner of the motor theme category illustrates the high degree of development and relevance of the terms “blockchain,” “IoT,” and “network security.” These themes occupy a prominent position within the fields of technology and digital security, exerting a significant influence over a range of related subjects. Blockchain technology is transforming numerous sectors, including finance and data security. The Internet of Things (IoT) offers a plethora of applications through the interconnection of devices, and the security of these connections is of paramount importance. These subjects occupy a central position in the field of technology and are acknowledged as being at the vanguard of innovation.

The terms “Byzantine error,” “Byzantine failures,” and “fault-tolerant computer systems,” located in the lower left-hand corner and classified as emerging or discredited themes, pertain to significant challenges encountered in domains such as distributed systems and blockchain technology. The terms “Byzantine error” and “Byzantine failure” collectively refer to errors that arise from malfunctioning or malicious system components, which can result in the rest of the system providing incorrect information. The management of such errors is a complex process, and the range of available solutions is limited. Fault-tolerant computer systems are defined as systems that are capable of maintaining operational integrity even in the event of the failure of one or more components. Nevertheless, the intricacy and expense of these systems render their extensive implementation challenging. These subjects are classified as either emerging or discredited themes, given that they represent issues that have not yet reached a sufficient level of maturity or gained sufficient acceptance due to the technical challenges and research requirements involved.

The terms “fault tolerance,” “peer-to-peer networks,” and “scalability,” which are ecentraliz as core themes in the bottom right corner, are core topics of high relevance, although they have a low degree of development. Fault tolerance can be defined as the resilience of systems to failures, and it is a vital quality for systems that are required to function reliably. Peer-to-peer networks represent a distributed network structure and play an important role in the sharing and communication of data. The term “scalability” is used to describe the capacity of systems to maintain their performance despite increasing loads. These topics are fundamental to the development of technology and computing infrastructure and have a wide range of potential applications, which is why they are ecentraliz as core themes.

The terms “learning systems,” “deep learning,” and “edge computing,” which are situated centrally within the table, are topics of high relevance but relatively moderate development. The development of learning systems and deep learning is of great importance in the fields of artificial intelligence and machine learning. Edge computing facilitates the decentralized processing of data on local devices, which is particularly pertinent for the Internet of Things (IoT) and time-critical applications. These topics occupy a significant position within the technological domain, yet they are at a central point due to their relative immaturity.

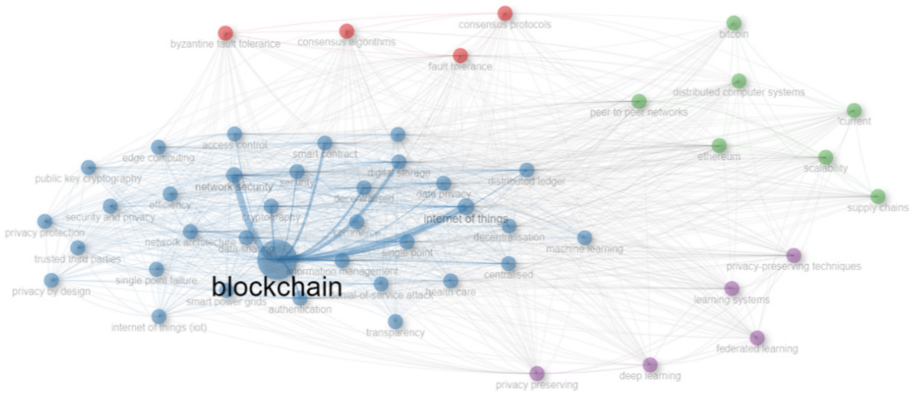


Figure 8. Co-occurrence network.

Figure 8 illustrates the network cluster of authors' fields of study in publications that encompass the themes of blockchain and error or failure. The terms "blockchain" and "block-chain" are used interchangeably. The network comprises four distinct clusters, which are represented by different colours. Blockchain, the Internet of Things, network security and smart contracts are all part of a more dense and interconnected structure, indicating that these areas are used more frequently than others. An evaluation of these areas based on a dataset derived from the failure literature reveals that they are the areas that require the most attention. The blue association network demonstrates the close relationship between these terms.

In the context of the green theme, the use of peer-to-peer networking with distributed ledger technology reveals that the structure between businesses and individuals can be intricate and multifaceted. Such relationships are frequently intended to be conducted using smart contracts, with Ethereum being the most commonly known smart contract provider. In the context of the green area, the supply chain theme also demonstrates that this technology is most extensively utilised within this sector.

In the red area, studies on the prevention of further technical errors in the context of error and failure have been conducted. A high degree of fault tolerance, attributable to its decentralised structure, characterises blockchain technology. This enables it to maintain operational continuity even in the event of the failure of some nodes within the network. Blockchain offers fault tolerance through the utilisation of consensus protocols and algorithms. The aforementioned protocols and algorithms guarantee that all nodes within the network possess identical information and that transactions are duly verified. In particular, Byzantine fault tolerance ensures the correct operation of the blockchain in the presence of malicious or faulty nodes.

The purple area encompasses studies that concentrate on the issues of privacy and security as they pertain to the applications of blockchain technology. In particular, the protection and security of sensitive data is of paramount importance for the adoption and diffusion of blockchain technology. The objective of studies in this area is the development of techniques that ensure the confidentiality of data and thus increase the reliability of blockchain technology while simultaneously minimising potential risks. The application of machine learning methods, including deep learning and machine learning, can facilitate the detection of anomalies and the identification of vulnerabilities through the analysis of data stored on the blockchain. In this manner, blockchain systems can be rendered more secure and users' privacy can be safeguarded in a more comprehensive manner.

Discussion and Conclusion

This study aims to analyse the extant literature on blockchain and failure by bringing together the concepts of blockchain and failure. In this context, the study will analyse the reasons for the failure of blockchain applications and provide ideas to other researchers who will contribute to this field. For this purpose, 2779 documents in the Scopus database were examined using the Biblioshiny library of the R studio programme. The analysis of the downloaded documents revealed that the earliest study was conducted in 2011, while the most recent was conducted in 2023. documents written in languages other than English were excluded from the analysis. Scopus was selected as the primary database due to its status as one of the largest databases available. The tables obtained from the downloaded data were analysed in the finding section.

As demonstrated in the findings related to blockchain technology, there has been a substantial increase in the number of publications utilising the terms "blockchain" and "failure" since 2017. Despite the rapid advancements in blockchain technology and the expanding range of its applications, meticulous observation of its development is imperative to ensure the optimal utilisation of the technology (Carson et al., 2018).

Blockchain networks are subject to considerable limitations, particularly about processing capacity and speed. These limitations result in a restriction of the number of transactions that the network is capable of processing per second, leading to network congestion and delays, particularly during periods of high demand. The aforementioned issues are compounded by the increasing number of nodes and transactions within the blockchain network, as well as the requirement for each node to verify and store every transaction. To illustrate this point, consider the fact that Bitcoin's transaction throughput is limited to just 7 transactions per second (TPS), in contrast to the up to 400 TPS achieved by traditional payment systems such as Visa. Moreover, the validation of blocks in Bitcoin can require up to 10 minutes, resulting in transaction delays (Khan et al., 2021). The analysis reveals that concepts such as scalability and network security are frequently cited in publications related to blockchain and failure. This observation indicates that scalability issues play a pivotal role in the failure of blockchain applications. A further analysis of the periodic distribution graph reveals a notable increase in the use of the term ‘scalability’ since 2021, suggesting that scalability problems have begun to garner increased attention with the expanding use of blockchain technology.

The utilisation of distinct protocols, data formats and smart contract languages by each blockchain platform engenders significant challenges in the transfer of data and assets between disparate platforms, thereby impeding the development of a cohesive blockchain ecosystem. The diverse application domains of blockchain technology have promoted the development of bespoke blockchain systems by various organisations, each tailored to its specific requirements. This has consequently produced a plethora of blockchain projects that employ disparate protocols and architectures. These projects, which rely on varied technologies and consensus protocols, cater to a range of use cases or applications. However, the proliferation of these projects has resulted in the fragmentation of blockchain developments and limited interoperability between different blockchain projects (Mohanty et al., 2022). The thematic map demonstrates that the concept of decentralisation is a prominent feature. The inherently decentralised nature of blockchain technology may impede the assurance of interoperability between disparate platforms. Furthermore, an analysis of the association network reveals that the integration of peer-to-peer and distributed ledger technology concepts poses significant challenges in establishing effective communication and interaction between disparate blockchain networks.

The concept of smart contracts, which is frequently referenced in the extant literature, posits that the absence of standards in the development and implementation of smart contracts, in conjunction with the incompatibility between platforms, may contribute to the failure of blockchain applications, given that each smart contract operates on its own blockchain platform. These findings demonstrate that the incompatibility of different blockchain platforms is a significant factor that can lead to the failure of blockchain applications.

To resolve this issue, it is essential to establish standards between different blockchain platforms and develop protocols to ensure interoperability. Inadequate testing and implementation processes appear to play a significant role in the failure of blockchain applications. Due to the lack of experience in blockchain and the immaturity of the technology, unsuccessful applications are persisting. Consequently, it can be concluded that this technology is still in the development stage (Treiblmaier, 2019). The intricacy of the technology’s structure and the challenges associated with its integration into diverse sectors necessitate meticulous testing and implementation processes.

Despite the significant promises of the blockchain, particularly in terms of its decentralised structure and security assurances, it has not lived up to expectations, particularly within the legal context. The technology encounters various challenges and uncertainties in the legal domain due to its decentralised architecture and the anonymity of its users. This results in ambiguities concerning the legal status of blockchain-based organisations and financial assets, national security regulations and the limited liability of shareholders (Quintais et al., 2019). The blockchain's insensitivity to off-chain events, such as fraud, coercion and theft, which may affect the validity and enforceability of legal transactions, results in a discrepancy between the law and the blockchain. While a comprehensive reorganisation of the legal framework is necessary to address this incompatibility, it may jeopardise the key features of the blockchain and hinder its adoption (Schilling, 2021). Furthermore, the reluctance of other nations to adopt this technology, due to concerns regarding its reliability, could potentially result in its prohibition. This ambiguity is a cause for concern for investors and developers, potentially hindering the financial support and progress of blockchain-based projects.

For blockchain technology to realise its full potential, issues must be addressed. Future research should concentrate on developing solutions to increase the success of blockchain applications by focusing on issues such as scalability, interoperability, security, energy efficiency, user experience, standards, testing and verification, regulations, and legal uncertainties. In addition, more case studies should be conducted on the challenges and reasons for the failure of blockchain technology applications in different sectors (finance, health, supply chain, energy, etc.).

Ethics Committee Approval: Authors declared that this study does not require ethics committee approval.

Peer-review: Externally peer-reviewed.

Author Contributions: Conception/Design of Study- M.Ş., M.E..; Data Acquisition- M.Ş., M.E.; Data Analysis/ Interpretation- M.Ş., M.E.; Drafting Manuscript- M.Ş., M.E.; Critical Revision of Manuscript- M.Ş., M.E.; Final Approval and Accountability- M.Ş., M.E.; Technical or Material Support- M.Ş., M.E.; Supervision- M.Ş., M.E.

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: The authors declared that this study has received no financial support.

References

Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive bibliometric analysis of scientific literature. *Journal of Informetrics*, *11*(4), 959-975.

Atzori, M. (2018). Blockchain technology and its potential applications in the real estate industry. *Journal of Corporate Real Estate*, *20*(4), 232-245.

Aydar, M. & Çetin, S. C. (2020). Blokzincir Teknolojisinin Sağlık Bilgi Sistemlerinde Kullanımı. *Avrupa Bilim ve Teknoloji Dergisi*, (19), 533-538.

Bodkhe, U., Tanwar, S., Parekh, K., Khanpara, P., Tyagi, S., Kumar, N., & Alazab, M. (2020). Blockchain for industry 4.0: A comprehensive review. *IEEE Access*, *8*, 79764-79800.

Caro, M. P., Ali, M. S., Vecchio, M., & Giaffreda, R. (2018). Blockchain-based traceability in Agri-Food supply chain management: A practical implementation. In *2018 IoT Vertical and Topical Summit on Agriculture-Tuscany (IOT Tuscany)* (pp. 1-4). IEEE. https://doi.org/10.1109/IOT-TUSCANY.2018.8373021

Ceylan, O., & Işık, A. H. (2023). Blokzinçir teknolojisi ve uygulama alanları. *Uluborlu Mesleki Bilimler Dergisi*, 6(1), 129-154.

Cui, Z., Xue, F., Zhang, S., Cai, X., Cao, Y., Zhang, W., & Chen, J. (2020). A hybrid blockchain-based identity authentication scheme for multi-WSN. *IEEE Transactions on Services Computing*, *13*(2), 241-251.

Carson, B., Romanelli, B., Walsh, P. and Zhumaev, A. (2018), "Blockchain beyond the hype: what is the strategic business value?", available at: https://www.mckinsey.com/business-functions/digital mckinsey/our-insights/blockchain-beyond-the-hype-what-is-the-strategic-business-value (accessed 7 January 2019).

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, *133*, 285-296.

Güven, C. (2023, 6 July). *Blockchain nedir? Üretimi nasıl etkiliyor?* [Blog yazısı]. ERP.tc. https://www.erp.tc/post/blockchain-nedir-uretimi-nasil-etkiliyor

Huang, J., Kong, L., Chen, G., Wu, M. Y., Liu, X., & Zeng, P. (2019). Towards secure industrial IoT: Blockchain system with credit-based consensus mechanism. *IEEE Transactions on Industrial Informatics*, *15*(6), 3680-3689.

Khan, S. N., Loukil, F., Ghedira-Guegan, C., Benkhelifa, E., & Bani-Hani, A. (2021). Blockchain smart contracts: Applications, challenges, and future trends. *Peer-to-peer Networking and Applications*, *14*, 2901-2925.

Li, M., Weng, J., Yang, A., Lu, W., Zhang, Y., Hou, L., ... & Deng, R. H. (2019). CrowdBC: A blockchain-based decentralized framework for crowdsourcing. *IEEE Transactions on Parallel and Distributed Systems*, *30*(6), 1251-1266.

Mendling, J., Weber, I., van der Aalst, W., vom Brocke, J., Cabanillas, C., Daniel, F., ... & Zhu, L. (2018). Blockchains for business process management-challenges and opportunities. *ACM Transactions on Management Information Systems (TMIS)*, *9*(1), Article 4.

Mistry, I., Tanwar, S., Tyagi, S., & Kumar, N. (2020). Blockchain for 5G-enabled IoT for industrial automation: A systematic review, solutions, and challenges. *Mechanical Systems and Signal Processing*, *135*, 106382.

Mohanty, D., Anand, D., Aljahdali, H.M., & Villar, S.G. (2022). Blockchain Interoperability: Towards a Sustainable Payment System. *Sustainability*, *14*(913), 2-20. https://doi.org/10.3390/su14020913

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.

Parlar, T. (2022). Blokzincir Teknolojisi ve Merkeziyetsiz Finans Uygulamaları [Blockchain Technology and Decentralized Finance Applications]. *Journal of Politics, Economy and Management*, *5*(2), 165-174

Schillig, M. (2021). 'Lex Cryptographia', 'Cloud Crypto Land' or What? - Blockchain Technology on the Legal Hype Cycle. King's College London Law School Research Paper. https://doi.org/10.2139/ssrn.3804197

Treiblmaier, H. (2018). Blockchain for supply chain management: Opportunities, challenges and research agenda. *Proceedings of the 51st Hawaii International Conference on System Sciences*.

Treiblmaier, H. (2019). Toward More Rigorous Blockchain Research: Recommendations for Writing Blockchain Case Studies. *Frontiers in Blockchain*, 2(3), 1-15. https://doi.org/10.3389/fbloc.2019.00003.

Tüfekçi, A., & Karahan, Ç., (2019). Blokzincir Teknolojisi ve Kamu Kurumlarınca Verilen Hizmetlerde Blokzincirin Kullanım Durumu. *Verimlilik Dergisi*, *4*, 157-193.

Wang, Y., Su, Z., Ni, J., Zhang, N., & Shen, X. (2022). Blockchain-Empowered Space-Air-Ground Integrated Networks: Opportunities, Challenges, and Solutions. *IEEE Communications Surveys & Tutorials*, *24*(1), 160-211. https://doi.org/10.1109/COMST.2021.3131711.

Yıldırım, H. (2018). Açık ve uzaktan öğrenmede blokzincir teknolojisinin kullanımı. *AUAd*, *4*(3), 142-153.

Zamani, M., Movahedi, M., & Raykova, M. (2018). RapidChain: Scaling Blockchain via Full Sharding. In *ACM SIGSAC Conference on Computer and Communications Security (CCS '18)*, October 15–19, 2018, Toronto, ON, Canada. ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3243734.3243853

Zhao, G., Liu, S., Lopez, C., Lu, H., Elgueta, S., Chen, H., & Boshkoska, B. M. (2019). Blockchain technology in agri-food value chain management: A synthesis of applications, challenges and future research directions. *Computers in Industry*, *109*, 83-99. https://doi.org/10.1016/j.compind.2019.04.002.

Zīle, K., & Strazdiņa, R. (2018). Blockchain use cases and their feasibility. *Applied Computer Systems*, *23*(1), 12-20. https://doi.org/10.2478/acss-2018-0002.

Quintais, J. P., Bodo, B., Giannopoulou, A., & Ferrari, V. (2019). Blockchain and the law: A critical evaluation. *Stanford Journal of Blockchain Law & Policy.*