Decision Science Letters 14 (2025) 435–448

Contents lists available at GrowingScience

Decision Science Letters

homepage: www.GrowingScience.com/dsl

Evolution and gaps in data mining research: Identifying the bibliometric landscape of data mining in management

Romel Al-Ali[a*], Sabri Mekimah[b], Rahma zighed[b], Rima Shishakly[c], Mohammed Almaiah[d], Rami Shehab[e], Tayseer Alkhdour[e] and Theyazn H.H Aldhyani[f]

[a]*Associate Professor, The National Research Center for Giftedness and Creativity, King Faisal University, Saudi Arabia*
[b]*Ecofima Laboratory, Department of Management Sciences, University of 20 Août 1955 Skikda, Algeria*
[c]*Management Department, College of Business Administration, Ajman University, Ajman 346, United Arab Emirates*
[d]*King Abdullah the II IT School, The University of Jordan, Amman 11942, Jordan*
[e]*College of Computer Science and Information Technology, King Faisal University, Al-Ahsa 31982, Saudi Arabia*
[f]*Applied College in Abqaiq, King Faisal University, Al-Ahsa 31982, Saudi Arabia*

C H R O N I C L E

Article history:
Received: July 10, 2024
Received in revised format: August 29, 2024
Accepted: September 202 2024
Available online: September 22, 2024

Keywords:
Data mining
Decision-making
Artificial intelligence
Forecasting
Sentiment analysis
Bibliometric

A B S T R A C T

This study conducts a bibliometric analysis of data mining publications in the Scopus database, examining the evolution of the field from 2015 to 2024. The study examines the bibliometric structure of data mining in management. Analyzing 2,942 publications, the research identifies significant growth in data mining studies. It reveals gaps in integrating data mining with decision-making, artificial intelligence, forecasting, and sentiment analysis. Despite a large number of publications, interdisciplinary applications of data mining are limited. The scientific publication on data mining and its relationship with decision-making, artificial intelligence, forecasting, and sentiment analysis is found to be weak, showing significant research gaps in these areas. China and the USA are prominent contributors, indicating geographical concentration. The study highlights the need for broader interdisciplinary exploration in data mining beyond traditional areas, urging global researchers to diversify contributions. The analysis focuses solely on publications indexed in Scopus, potentially excluding relevant studies from other databases or sources. This study provides insights into the evolution of data mining research and identifies areas for further interdisciplinary exploration, contributing to the advancement of the field's boundaries.

1. Introduction

In an era where the pace of digital transformation accelerates and data emerges as a pivotal element in shaping administrative methodologies and decision-making strategies, data mining serves as a cornerstone for enhancing administrative performance and achieving sustainability within organizations. The art and science of management require insight and precision in choosing the optimal options (Baek & Doleck, 2022), where qualitative data and the ability to analyze and interpret it play a central role. Through advanced mining techniques (Maryoosh & Hussein, 2022), leaders and managers can extract deep insights and complex patterns from vast datasets, enabling informed decision-making and the construction of effective competitive strategies (Ramageri, 2010). From decision theory to information theory, and from machine learning to organizational knowledge theory, a range of

* Corresponding author.
E-mail address ralali@kfu.edu.sa (R. Al-Ali)

ISSN 1929-5812 (Online) - ISSN 1929-5804 (Print)

doi: 10.5267/j.dsl.2024.12.011

theories appears as a theoretical foundation supporting the exploitation of data in the administrative process and decision-making (Mannila, 2000). Furthermore, data analysis technologies are continuously evolving, from machine learning and deep learning to data mining and statistical analysis, including text analysis and visual analytics (Ramageri, 2010). These technologies find their applications in improving decision-making processes, risk management, logistics operations, and targeted marketing, as well as analyzing complaints and feedback and improving inventory management (Furkan & Haide, 2014). By exploring these intertwined relationships and practical examples, data mining demonstrates its ability to support effective management and strategic decision-making, opening new horizons for innovation and contributing to the success and sustainable growth of organizations in the dynamic business environment of today (Mydyti, 2021). This environment requires a strategic use of data to foresee challenges and seize opportunities. Data mining, with its advanced techniques and ability to delve into data depths, provides indispensable tools for market analysis, understanding customer behavior, and overall performance enhancement (Chala et al., 2021). As technology advances, data becomes more complex and abundant, making the ability to analyze and extract insights from it increasingly important. Data mining not only aids in identifying patterns and trends but also in predicting the future, enabling companies to adapt and evolve in response to forthcoming changes (Graefe et al., 2013). The challenges faced by management today extend beyond mere data collection to analyzing and utilizing it in a way that serves the organization's goals. Here, data mining emerges as a central pillar in the data strategy, providing the necessary foundations for achieving strategic excellence and competitive advantage (Marian et al., 2023).

In conclusion, data mining serves as a potent tool, empowering organizations to delve deeply into their data. This exploration helps not only in understanding the current state but also in anticipating future opportunities and challenges. Through integrating advanced managerial theories and employing modern data analysis technologies (Altintas & Trick, 2014), companies can enhance their ability to make informed decisions, and build robust strategies ensuring a sustainable competitive edge in an ever-changing market (Rosamilha et al., 2023).

In light of the aforementioned, we are able to ask the following question:

RQ: Is there a sufficient scholarly publication on data mining and related topics that demonstrate the strong relationship between them to bridge the research gap in this field?

Our study aims to analyze the literature and scientific publications to understand the trends, developments, and research gaps in this field through identifying prevailing trends and key research topics within the realm of data mining. This provides an overview of the evolution of this field over time, revealing the relationships among different researches, topics, authors, and journals, thus aiding in identifying collaboration networks and mutual influence among researchers and institutions. It also allows evaluating the impact of published research in data mining by analyzing citations and reviews, helping to pinpoint gaps and research areas in need of further exploration and study, thereby guiding researchers toward new research opportunities.

Additionally, we aim to provide insights on how data mining can be used to improve decision-making processes in project management, explore effective methods and tools in data mining for risk analysis, outcome prediction, and overall project performance enhancement (Alam et al., 2023).

We also seek to understand the impact of data mining on innovation and product development within project management. Identifying recent trends and innovative technologies in the field of data mining that can contribute to enhancing scientific research and promoting research collaboration by identifying leading researchers and institutions in this field, and guiding researchers and students towards promising research topics by identifying current research gaps. Overall, through this study, we aim to achieve a deep understanding of the field, guiding research and practical efforts towards more effective and innovative uses of data mining.

2. Literature review

Data mining is an analytical process aimed at discovering and extracting previously unknown patterns, relationships, or valuable information from large and complex data sets, utilizing methods from computer science and statistics (Varajão et al., 2022). Data mining is a component of the wider process of Business Intelligence and data analysis. This process employs computer and statistical techniques to uncover trends, forecasts, and behavioral patterns that traditional analyses might not reveal (Bertoni & Larsson, 2017). The goal of data mining is to transform raw data into applicable knowledge, allowing organizations and companies to make strategic decisions based on data (Upadhyay, 2018). Data mining applications span diverse areas in decision-making, forecasting, artificial intelligence, and studying consumer behavior through sentiment analysis, demonstrating its wide importance in data analysis and extracting valuable information in business management (Samways dos Santos et al., 2019). Based on the foregoing, we can formulate the main hypothesis of our study:

MRH: There is extensive scientific publishing on the topic of data mining and related subjects, which indicates a strong relationship between them filling the research gap in this field.

2.1. Decision-making

Is an analytical and rational process involving the selection of an alternative from several possible options to solve a problem or achieve a specific goal (Santoso et al., 2022), This process includes several steps starting from problem identification, gathering relevant information and data, evaluating possible alternatives based on defined criteria and goals, and finally choosing and implementing the optimal alternative (Pacheco-Velázquez et al., 2023). Decision-making is not limited to individuals but is also a crucial part of administrative processes in organizations and companies, requiring precise evaluation of risks, benefits, and potential impacts of each decision (Guo et al., 2022). Decisions can range from simple choices like what to eat for lunch to complex ones like formulating a new strategy for the company in a competitive market (Marchisott et al., 2018). In the business context, decision-making is an essential skill that requires a deep understanding of the competitive environment, analysis of available data and information, and the ability to predict the outcomes and impacts of decisions (Millemann et al., 2022). The aim of decision-making is to achieve the strategic objectives of the organization and enhance its competitive position in the market (Bond et al., 2008). The relationship between data mining and decision-making is pivotal in the administrative and organizational process, manifesting in key points as follows:

- Data mining enables organizations to discover statistics, patterns, and trends that may be unclear in big data.
- Data mining enables the development of predictive models, helping management to understand future trends and make proactive decisions (Mydyti, 2021
- Data mining aids in delving deeper into data to search for value-added information that contributes to inform decision-making (Mydyti, 2021).
- Data mining supports decision-makers by providing detailed and specific information on particular topics (Mydyti, 2021).
- Insights gained from data mining help improve efficiency, effectiveness, and productivity within the organization (Gora, 2019).
- Data mining can identify potential risks and threats, allowing management to address them before they escalate (Da Silva et al., 2020).
- Using information extracted from data mining, organizations can more effectively allocate resources (Zhan, 2016).
- Information obtained through data mining can generate new ideas for products, services, and processes (Nagendhra & Jen, 2024).
- Data mining can lead to a better understanding of customer needs and behaviors, resulting in decisions that enhance customer experience and brand loyalty (Pynadath et al., 2023).
- Data mining enables companies to understand the market and competitors more deeply, contributing to more effective marketing and strategic decisions (AnhKhoa et al., 2019).

In general, data mining provides the necessary knowledge foundation for making objective decisions and enhances the competitive ability of the organization by transforming data into actionable knowledge (AnhKhoa et al., 2019|). Therefore, we can formulate the first sub-hypothesis of our study as follows:

SRH1: There is extensive scientific publishing that combines the terms data mining and decision-making, indicating a strong relationship between them that fills the research gap in this field.

2.2. Forecasting

A methodical process analyzes current and historical data and identifies prevailing trends to envision potential future events or developments. It aims to provide insights and predictions about the future to aid in strategic planning and decision-making (Taylor & Letham, 2018). Forecasting seeks not to predict the future with absolute accuracy but to explore possible future scenarios and assess how current decisions can affect future outcomes (Matta et al., 2021).

The forecasting process involves collecting and analyzing data, including economic, social, technological, and political trends, to identify factors that may affect the future of organizations or industries. Forecasting uses tools like scenarios, forecasts, and future mapping to explore potential opportunities and challenges. This empowers organizations to enhance their future planning, and develop flexible, adaptable strategies (Buchatskaya et al., 2015). The relationship between data mining and forecasting centers on using available data to accurately predict future events. Data mining is an essential part of the forecasting process, providing the tools and techniques necessary for analyzing large data sets and extracting complex patterns and relationships that may not be apparent (Guermoui et al., 2024). This relationship manifests in several ways:

- Data mining reveals historical patterns and trends in datasets, serving as a basis for future predictions (Altintas & Trick, 2014).
- Using machine learning techniques, data mining can analyze past records and learn from them to predict future outcomes with greater accuracy (Xiuyi & Yuxia, 2018
- Advanced data mining techniques allow for precise analysis of complex interactions between different factors, leading to improved prediction accuracy (Xiuyi & Yuxia, 2018).
- Data mining applies to time-series analysis to evaluate how variables change over time, crucial for forecasting future trends (Saad et al., 2021).
- Organizations can analyze data continuously, thus responding faster to changes in the market or competitive environment (Xiuyi & Yuxia, 2018).
- Data mining enables the creation of models that can simulate various scenarios and see how different decisions might affect the future (Maksood & Achuthan, 2016).
- Based on insights gained from data mining and future predictions, organizations can make informed strategic decisions that enhance their chances of success (Wiśniewsk, 2021). Therefore, we can formulate the second sub-hypothesis of our study as follows:

  SRH2: There is extensive scientific publishing that combines the terms data mining and forecasting, indicating a strong relationship between them that fills the research gap in this field.

2.3. Artificial Intelligence (AI)

Is a branch of computer science that develops systems and software capable of performing tasks that require human intelligence, such as learning, understanding, reasoning, and adapting (Marian et al., 2023). AI encompasses a wide range of technologies, including machine learning, deep learning, natural language processing, and computer vision, used in various applications from smart recommendation systems and virtual voice assistants to advanced robots and big data analytics (Zahlan et al., 2023). The relationship between data mining and artificial intelligence (AI) is integrative and interwoven, with both working together to analyze big data and extract patterns and knowledge. Key points highlighting this relationship include:

- Data mining involves extracting useful knowledge from large and complex data sets. AI refers to developing systems capable of performing tasks requiring human intelligence, such as learning, inference, pattern recognition, and decision-making (Leal Filho et al., 2023).
- Machine Learning, a branch of AI, acts as a bridge between data mining and AI, with algorithms used to analyze data and extract patterns, central to the data mining process (Bach et al., 2021).
- Data mining can be significantly enhanced using AI, especially in processing and analyzing big data, providing more powerful and accurate analysis tools, allowing for faster and more efficient data analysis (Dwivedi et al., 2023).
- AI expands the scope of data mining applications, enabling entry into areas like Deep Learning for image recognition and Natural Language Processing for understanding texts (Hearst & Hirsh,2020).
- Both data mining and AI benefit from ongoing technological advancements, leading to the innovation of new analysis methods and data extraction that can provide new and valuable insights for businesses and science (Zhisheng, 2023).
- Both fields raise questions about privacy and ethics, especially regarding data collection and analysis (Zhao et al., 2019). Therefore, we can formulate the third sub-hypothesis of our study as follows:

  SRH3: There is extensive scientific publishing that combines the terms data mining and artificial intelligence, indicating a strong relationship between them that fills the research gap in this field.

2.4. Sentiment Analysis

Also known as, emotion mining is the process of using natural language processing, text analysis, and artificial intelligence techniques to understand and analyze the sentiments and opinions expressed in written texts. The goal of sentiment analysis is to identify whether the expressed emotions in the text are positive, negative, or neutral, and sometimes to assess their intensity. Sentiment analysis is widely used in analyzing social media, consumer reviews, forums, and blogs to gather insights about customer opinions, preferences, and experiences with products or services. Data mining also supports market analysis, political research, and brand monitoring, facilitating strategic decisions through a deeper understanding of public sentiments and trends (Amiri et al., 2015). Data mining and sentiment analysis closely intertwine, boosting the capacity to comprehend data and extract valuable information, especially regarding emotions and opinions. Key points clarifying this relationship include:

- Data mining helps in extracting and analyzing big data, while sentiment analysis focuses on assessing and understanding the sentiments and trends from this data. Sentiment analysis increasingly processes texts from social media, online reviews, and articles to capture public opinions.
- While data mining can identify quantitative patterns and trends, sentiment analysis adds a qualitative dimension, enabling companies and researchers to understand the emotions and sentiments associated with these patterns (Moorjani et Sadath, 2019).
- Data mining with sentiment analysis can provide valuable insights to companies about how customers feel about their products or services, helping to improve customer experience and increase brand loyalty (Cristescu et al., 2022).
- Integrating data mining and sentiment analysis can support the decision-making process with most comprehensive information, helping companies develop more effective marketing and advertising strategies based on a deep understanding of customer sentiments and preferences (Moorjani & Sadath, 2019; Cristescu et al., 2022).
- Advancements in machine learning and artificial intelligence significantly enhance data mining and sentiment analysis, with algorithms analyzing texts and classifying emotions accurately and quickly (Kharde & Sonawane, 2016).
- Sentiment analysis, an advanced application of data mining, uses complex techniques to analyze texts and understand the contexts of words and phrases, identifying associated emotions. This includes the ability to distinguish between sarcasm, humor, seriousness, and to determine the intensity of the emotion.

Overall, the relationship between data mining and sentiment analysis represents an important intersection that provides a deeper understanding of data in a way that helps guide strategic decisions and improve interactions with customers and the target audience. Therefore, we can formulate the fourth sub-hypothesis of our study as follows:

SRH4: There is extensive scientific publishing that combines the terms data mining and sentiment analysis, indicating a strong relationship between them that fills the research gap in this field. Overall, through this study, we aim to achieve a deep understanding of the field, guiding research and practical efforts towards more effective and innovative uses of data mining.

3. Research methodology

For the analysis of results, we relied on bibliometric studies using the Vos-viewer program, as they are among the most prominent studies employing statistical methods and mathematical techniques to analyze data related to books, periodicals, articles, authors, and other elements of documentary communication. The purpose is to identify the most productive and cited authors in scientific research and the countries publishing the most in the study topics.

3.1. Corpus curation

We obtained the data for this bibliometric analysis from the Scopus database, the most important source of bibliographic information in publications. We studied a multitude of researchers, terms, and the most publishing and cited countries in the topic of data mining, showing a result of 141,656 studies from 2015 to 2024. Subsequently, we filtered these studies down to only 2,942 studies specializing in business management, finance, and economics. We also analyzed those studies published in English, focusing on articles related to data mining, as shown in Fig. 1.

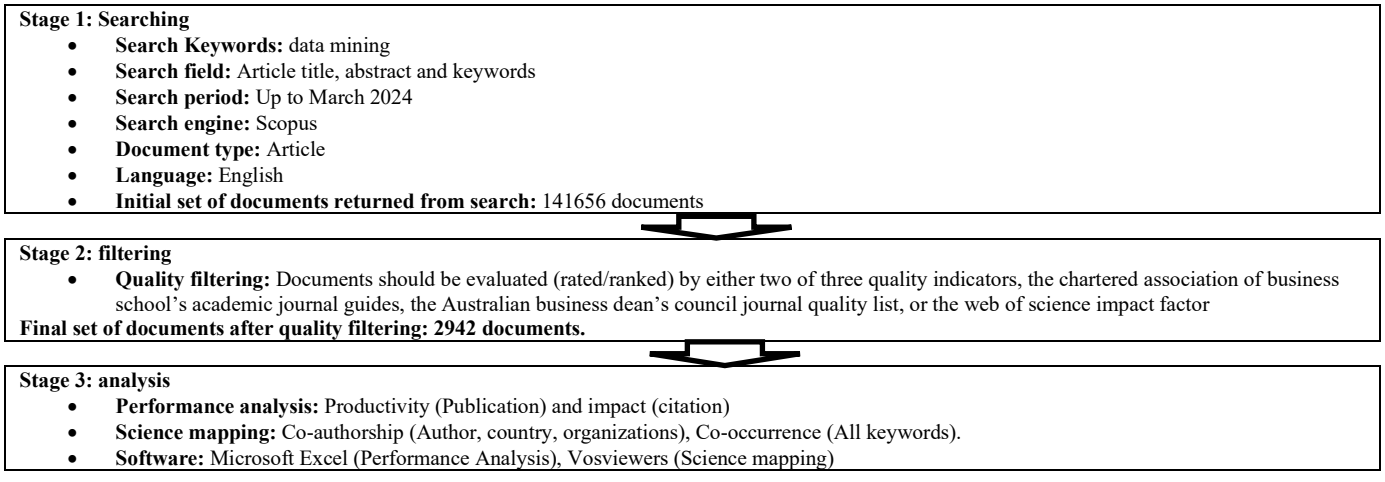

Fig. 1. Corpus curation

3.2. Data Mining Publishing During the Period (2015-2024)

The topic of data mining has seen a yearly evolution in publishing. Accordingly, we designated this element to track the cumulative number of research publications annually. Below, we will present the evolution of publishing over the last ten years, as shown in Fig. 2.

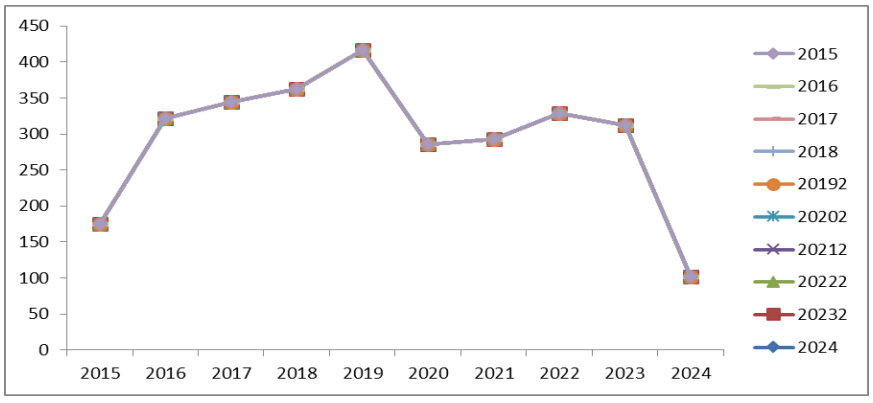

Fig. 2. Evolution of the Number of Publications on the Topic of Data Mining During the Period(2024-2015)

This figure shows the trend in the development of publications in the field of data mining research. There was not a large production of publications in the last year (2024), as it was at its lowest level (102 publications). From the year (2016) to (2019), the topic saw a significant increase in the number of publications, reaching a peak in (2019) with 416 publications, and then began to decline in (2020) with the number of publications that year being (286 publications). It increased again from (2021) to (2023), where the number of publications was continuously rising, but significantly decreased in (2024), which is attributed to the publication period being only three months for that year. The cumulative number of publications in this field during the period (2015-2019) was 1620 publications, taking five years, while the total number of publications during the last five years (2020-2024) was 1322 publications, indicating a decreased interest in data mining research, particularly noted during the three-month period of (2024).

3.3. Citation Analysis

We also conducted a citation analysis using the document analysis unit to highlight the most important articles based on the number of citations. The analysis revealed that among all the documents analyzed (Panța & Popescu, 2023), there were 50 documents containing more than 10 citations. The scarcity of citations might result from the recent publication of many research works within the year 2024, as shown in Fig. 3.

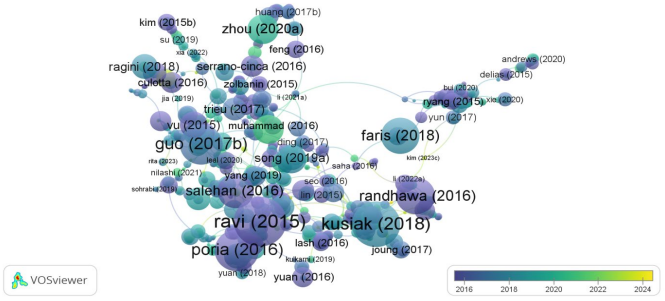

Fig. 3. Citation map of all analyzed documents

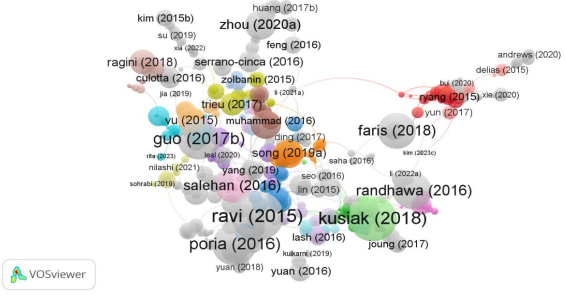

Fig. 4. Citation map of top cited documents

Given the abundance of articles, we aim to highlight the most significant documents that have a large number of citations. To simplify the visualization for the upcoming analysis stage, we have selected only those papers that have at least 50 citations. We will display this selection in Fig. 4 to highlight the most significant articles accordingly.

The most cited work belongs to Ravi and Ravi (2015), which had 958 citations in 2015 and was published in the Knowledge Based Systems journal. This includes a survey of tasks and applications in opinion mining and sentiment analysis. The second most cited article (810 citations) in 2018 is by Kusiak (2018) in the area of manufacturing.

The third article by Poria et al. (2016), ranked third in terms of citations (677 citations) published in Knowledge Based Systems, suggests its focus on opinion mining using deep convolutional neural networks.

The fourth article in terms of citations was by Guo et al. (2002) with 635 citations in 2017 published in Tourism Management, where the authors analyzed online ratings and reviews, focusing on tourist satisfaction using latent Dirichlet allocation.

The fifth article, with 458 citations in 2016 by Salehan and Kim (2016) discussed predicting the performance of online consumer reviews through a sentiment mining approach to big data analytics.

In summary, of the five articles, two directly or indirectly focus on sentiment mining and analysis, two on opinion mining and analyzing tourist satisfaction, and one solely on smart manufacturing. Therefore, the scarcity of studies necessitates a more in-depth review of the methodological literature on the same topic.

3.4. Countries and Their Cooperation

Many countries have actively engaged in the topic of data mining. This section will detail the countries that are most influential in this area of study, as shown in Table 1.

Table 1
Top 10 Most Influential Countries in Research on Data Mining

Rank	Country	Number of publications	Number of citations
1	China	883	**20453**
2	United states	493	**14369**
3	Australia	166	**6627**
4	United Kingdom	173	**5724**
5	India	373	**4767**
6	South Korea	128	**3973**
7	Spain	90	**2883**
8	Iran	126	**2724**
9	Japan	45	**2656**
10	Hong Kong	75	**2606**

Sixty-three countries worldwide, identified through the affiliations of the authors, have published on data mining-related topics in the Scopus database, allowing the same publication to appear in more than one country. China leads in the number of publications and citations, with 883 publications and 20,453 citations, followed by the United States with 493 publications and 14,369 citations. Hong Kong has the fewest citations, with 2,606.

4. Results

The subsidiary and main hypotheses will be tested to identify the scope of data mining and the research gaps according to the Scopus database for the period (2015-2024).

4.1. Testing the First Subsidiary Hypothesis

There is intense scientific publishing that combines the term data mining with decision-making and indicates the strength of the relationship between them, filling the research gap in this area, as shown in Fig. 5.

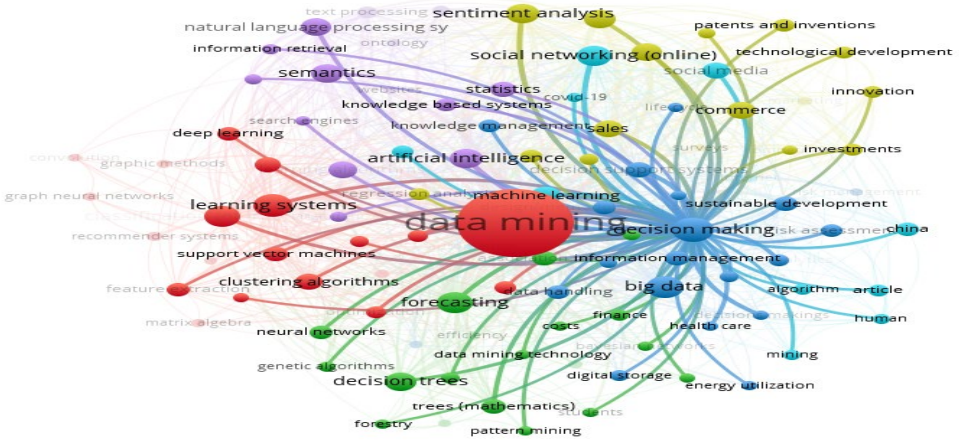

Fig. 5. Cluster of decision making

From the figure, we observe a cluster in red encompassing the term data mining, which includes a set of subsidiary terms primarily represented by: learning systems, deep learning, clustering algorithms, etc. There is also a cluster in blue encompassing the term decision-making, which also contains a set of subsidiary terms represented by: big data, information management, sustainable development, etc. We note a fundamental linkage between data mining, decision-making, and subsidiary linkages among various terms in these clusters. The number of articles published on this topic was estimated at 41 articles out of the total 2942 articles forming our study sample, representing 1.39%, which is a low percentage. Therefore, we reject this hypothesis and acknowledge a significant scientific gap between data mining and decision-making.

4.2. Testing the Second Subsidiary Hypothesis

There is intense scientific publishing that combines the term data mining with forecasting, indicating the strength of the relationship between them to fill the research gap in this area, as shown in Fig. 6.

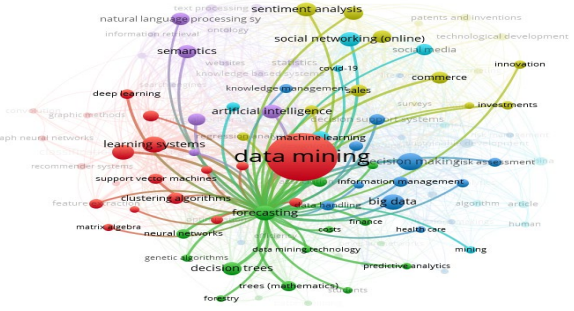

Fig. 6. Cluster of forecasting

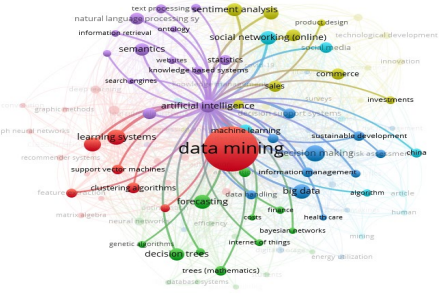

Fig. 7. Cluster of artificial intelligence

Through the analysis, we notice a red cluster that includes the term data mining, encompassing subsidiary terms like information management, deep learning clustering algorithms, etc. There is also a green cluster containing the term forecasting, which includes subsidiary terms like decision trees, data mining technology, neural networks, etc., indicating the cluster's limited scope. A fundamental connection exists between data mining and forecasting, with subsidiary linkages between the various terms in these two clusters. There were 8 articles published on this topic, out of a total of 2942 articles forming our study sample, representing a mere 0.27% a very low percentage. The most cited article, titled “Forecast of individual customer’s demand from a large and noisy dataset” by Murray et al. (2018) with 41 citations out of a total of 85 citations. Therefore, we reject this hypothesis and acknowledge a significant scientific gap between data mining and forecasting.

4.3. Testing the Third Subsidiary Hypothesis

There is intense scientific publishing that combines the term data mining with artificial intelligence, indicating the strength of the relationship between them to fill the research gap in this area, as shown in Fig. 7. From the analysis, we see a cluster in red encompassing the term data mining, including subsidiary terms like information management, deep learning, clustering algorithms, etc. Additionally, there is a cluster in purple that includes the term artificial intelligence, with subsidiary terms like statistics, knowledge-based systems, semantics, etc., showing the limited nature of this cluster. A fundamental linkage exists between data mining and artificial intelligence, with subsidiary linkages among the various terms in these clusters. There were 48 articles published on this topic, out of 2942 articles forming our study sample, which equates to 1.63%, a relatively low percentage. The most cited article, titled “Intelligent diagnostic prediction and classification system for chronic kidney disease” was published by Elhosney and Kumar (2019), had 128 citations out of a total of 645 citations. Therefore, we reject this hypothesis and acknowledge a significant scientific gap between data mining and artificial intelligence.

4.4. Testing the Fourth Subsidiary Hypothesis

There is intense scientific publishing that combines the term data mining with sentiment analysis, indicating the strength of the relationship between them to fill the research gap in this area, as shown in Fig. 8. From the analysis, we observe a red cluster that includes the term data mining, containing subsidiary terms like information management, deep learning, clustering algorithms, etc. Additionally, there is a yellow cluster encompassing the term sentiment analysis, which also contains subsidiary terms like social networking, commerce, investments, etc., illustrating the cluster's limited scope. A primary connection exists between data mining and sentiment analysis, with subsidiary linkages among various terms in these two clusters. There were 11 articles published on this topic, out of 2942 articles forming our study sample, amounting to 0.56%, a relatively weak percentage. The most cited article by Omuya et al., 2021), had 141 citations out of a total of 241 citations. Thus, we reject this hypothesis and acknowledge a significant scientific gap between data mining and sentiment analysis.

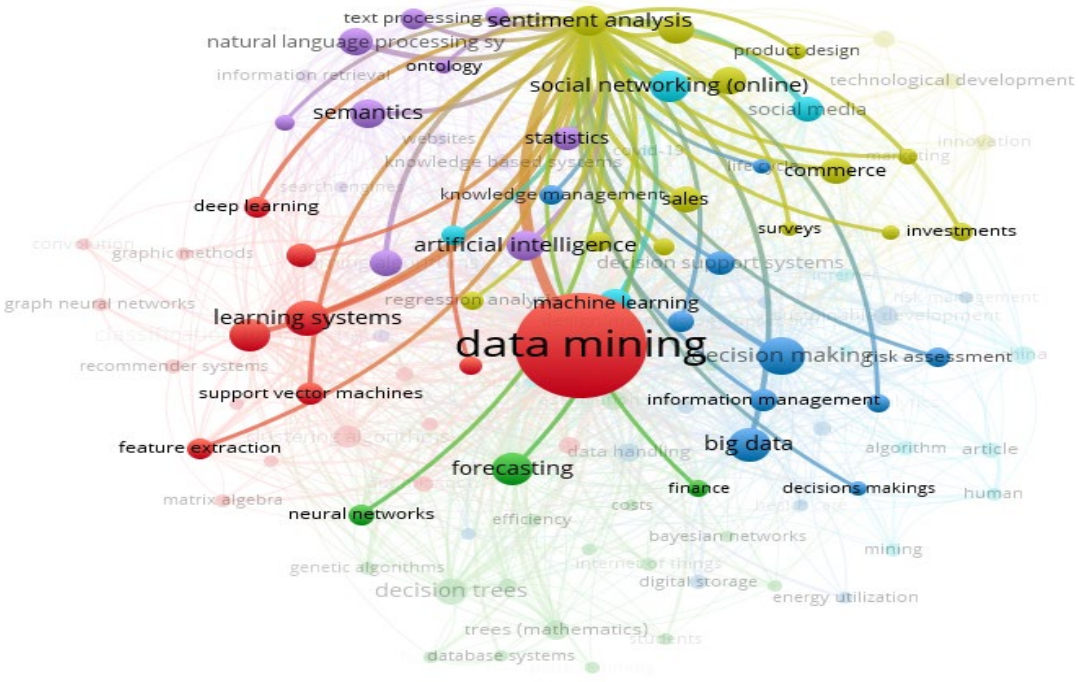

Figure 8. Cluster of sentiment analysis

4.5. Testing the Main Hypothesis

There is intense scientific publishing on the topic of data mining and related terms, indicating a strong relationship between them to fill the research gap in this area, as shown in Fig. 9.

Fig. 9. The concentration of data mining and related terms

Observing Fig. 9, it is evident that the publishing density on the topic of data mining is represented by the illuminated areas in the figure, encompassing the following terms in order: data mining, decision-making, artificial intelligence, forecasting, and sentiment analysis. The total number of articles published on the topic of data mining and related terms is estimated at 108 out of 2942, representing a mere 3.67%, which is a low percentage. Therefore, we reject this hypothesis and acknowledge the presence of a significant scientific gap between the topic of data mining and related subjects.

5. Discussion

The bibliometric analysis has led to several insights:

- The Scopus database contains a significant volume of scientific publications on data mining, totaling 2,942. This number contradicts the 2,162 studies estimated by Cheshmehsohrabi and Mashhadi (2022) and differs from the 291 studies counted by Baek and Doleck (2022).
- All research on data mining occurred in the last ten years (2015-2024), conflicting with the study by (Lundberg, 2023), which reported research from 2012 to 2022, and differing from (Kolling et al., 2021), focusing on the period from 1995 to 2020.
- There was an increase in scientific output in data mining research, with 55.06% in the first five years and 44.93% in the last five years, indicating a decreased interest in data mining in recent years, contrary to (Kolling et al., 2021), which found a recent increase of 72.99%.
- The paper published by Ravi and Ravi (2015) has the highest citation count of 958, followed by Kursiak, Andrew with 810 citations, showing significant contributions compared to other authors and differing from (Kolling et al., 2021), where Kolling had the highest with 945 citations.
- Publication diversity in Scopus about data mining includes articles, conferences, etc., aligning with studies by Kolling et al. (2021) and Lundberg (2023).
- The most publishing and interested countries in data mining are China first with 883 publications and 20,543 citations, followed by the United States with 493 publications and 14,369 citations, differing from (Cheshmehsohrabi & Mashhadi, 2022), where the United States led and from (Kolling et al., 2021) where the United States had 1973 publications followed by China with 923.
- Eight different author groups exist by citations, with leaders like Ravi and Ravi (2015) and Kusiak (2018) in various colored clusters.
- The terms related to data mining heavily relied on the term itself.

The scientific publication on data mining and its relationship with decision-making, artificial intelligence, forecasting, and sentiment analysis is found to be weak, showing significant research gaps in these areas. This finding aligns with several studies:

- The gap between data mining and decision-making is noted in Mydyti (2021), indicating an area that requires further research to understand and enhance the linkage between these fields.
- The connection between data mining and artificial intelligence, despite being intuitively strong, shows a substantial gap, as discussed by Dwivedi et al. (2023) and Yang et al. (2023), suggesting the need for more integrated and detailed research.
- The research linking data mining with forecasting has significant gaps, as highlighted in studies by (Guermoui et al., 2024) and (Altintas & Trick, 2014), pointing to the potential for developing predictive models and techniques.
- The relationship between data mining and sentiment analysis also shows a considerable gap, underscoring the need for more nuanced and in-depth studies in this area.

Despite the recognized weaknesses in scientific publishing, these gaps also represent promising research opportunities. The need for a deeper exploration of these relationships emphasizes the potential for innovative studies that can bridge these gaps, as supported by the cited studies, including (Mydyti, 2021; Dwivedi et al., 2023; Altintas & Trick, 2014). This highlights the ongoing evolution in the field of data mining and its interdisciplinary applications, suggesting fertile ground for future academic and practical advancements.

6. Conclusion

In conclusion, the field of data mining has been identified as both vital and ripe for further exploration. Through bibliometric analysis, we have discerned significant gaps in research, particularly in the integration of data mining with decision-making,

artificial intelligence, forecasting, and sentiment analysis. Despite the vast number of publications in databases like Scopus, the proportion of studies that deeply engage with these interconnections remains low, revealing a substantial opportunity for future research.

Leading countries like China and the USA dominate the field, yet there is a call for broader international contribution, particularly from nations like Japan and Hong Kong, to enrich the global landscape of data mining research. The role of prolific authors and the importance of publishing in influential journals are underscored as key factors in advancing the field's development.

As we move forward, researchers are encouraged to delve deeper into the multifaceted dimensions of data mining, leveraging advanced bibliometric methods to uncover rich veins of unexplored topics. The ultimate aim is to foster a more integrated and comprehensive understanding of data mining's potential to revolutionize decision-making processes, enhance predictive analytics, and interpret complex emotional data, thereby contributing to the broader sphere of knowledge and its application in real-world scenarios.

Based on the study's findings, the following recommendations can be made to researchers in the field of data mining:

- Emphasize the importance of data mining for enhancing decision-making and future forecasting, highlighting its significance in various fields.
- Employ bibliometric analysis to identify research gaps, collect data on the most cited authors, predominant countries, and commonly used terms related to data mining to guide future research directions.
- Given the currency and accessibility of information, prioritize citing and using sources in English and from electronic databases to stay updated with the latest developments.
- Encourage countries like Japan and Hong Kong to publish more on data mining to match the leading efforts of countries like China and the USA.
- Encourage authors to aspire to the levels of highly cited researchers like Ravi and Ravi (2015) and increase their scientific contributions to the field.
- Encourage the authorship of books and participation in international conferences on data mining to enhance the visibility and dissemination of research findings.
- Encourage more research on the interrelation of data mining with decision-making, artificial intelligence, sentiment analysis, and forecasting to fill the identified research gaps.

By adhering to these recommendations, researchers can contribute to advancing the field of data mining, addressing the existing gaps, and leveraging its potential for significant societal and technological advancements.

Acknowledgment

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Grant No. KFU241908).

References

Alam, S., Zardari, S., Abbas, M., Huda, N. U., & Khan, M. A. (2023). The Impact of Data Mining on Digital Libraries–A bibliometric Study. *VFAST Transactions on Software Engineering*, *11*(4), 124-137.

Altintas, N., & Trick, M. A. (2014). A data mining approach to forecast behavior. *Annals of Operations Research*, 216(1). https://doi.org/10.1007/s10479-012-1236-9

Amiri, F., Scerri, S., & Khodashahi, M. (2015, September). Lexicon-based sentiment analysis for Persian Text. In *Proceedings of the International Conference Recent Advances in Natural Language Processing* (pp. 9-16).

AnhKhoa, D. N., & others. (2019). Marketing intelligence from data mining perspective: A literature review. *International Journal of Innovation, Management and Technology*, 10(5). https://doi.org/10.18178/ijimt.2019.10.5.859

Baek, C., & Doleck, T. (2022). Educational data mining: A bibliometric analysis of an emerging field. *IEEE Access*, 10. https://doi.org/10.1109/ACCESS.2022.3160457

Bertoni, A., & Larsson, T. (2017). Data mining in product service systems design: Literature review and research questions. *Procedia CIRP*, 64. https://doi.org/10.1016/j.procir.2017.03.131

Bond, S. D., Carlson, K. A., & Keeney, R. L. (2008). Generating objectives: can decision makers articulate what they want?. *Management Science*, *54*(1), 56-70.

Buchatskaya, V., Buchatsky, P., & Teploukhov, S. (2015). Forecasting methods classification and its applicability. *Indian Journal of Science and Technology*, *8*(30), 1-8.

Chala, N., Voropai, O., & Pichyk, K. (2021). Using data mining to create innovations in education. *Електронне наукове фахове видання" Соціально-економічні проблеми і держава"*, (2 (25)), 21-28.

Cheshmehsohrabi, M., & Mashhadi, A. (2022). Using data mining, text mining, and bibliometric techniques to the research trends and gaps in the field of language and linguistics. *Journal of Psycholinguistic Research*, 52, 1–24.

Cristescu, M. P., Nerişanu, R. A., & Mara, D. A. (2022). Using Data Mining in the Sentiment Analysis Process on the Financial Market. *Journal of Social and Economic Statistics*, *11*(1-2), 36-58.

Da Silva, J. B. N., Senna, P., Chousa, A., & Coelho, O. (2020). Data mining and operations research techniques in Supply Chain Risk Management: A bibliometric study. *Brazilian Journal of Operations & Production Management*, *17*(3), 1-14.

dos Santos, B. S., Steiner, M. T. A., Fenerich, A. T., & Lima, R. H. P. (2019). Data mining and machine learning techniques applied to public health problems: A bibliometric analysis from 2009 to 2018. *Computers & Industrial Engineering*, *138*, 106120.

Dwivedi, Y. K., Sharma, A., Rana, N. P., Giannakis, M., Goel, P., & Dutot, V. (2023). Evolution of artificial intelligence research in Technological Forecasting and Social Change: Research topics, trends, and future directions. *Technological Forecasting and Social Change*, *192*, 122579.

Elhosney, M., & Kumar, U. (2019). Intelligent diagnostic prediction and classification system for chronic kidney disease. *Scientific Reports*, 9(1), 1–14.

Filho, W. L., Yang, P., Eustachio, J. H. P. P., Azul, A. M., Gellers, J. C., Gielczyk, A., ... & Kozlova, V. (2022). Deploying digitalisation and artificial intelligence in sustainable development research. *Environment, Development and Sustainability*.

Furkan, M., & Haide, A. S. (2014). An introduction to data mining technique. *International Journal of Advancement in Engineering Technology, Management & Applied Science*, 1(3).

Gora, A. A. (2019). The link between decision making process and performance: A bibliometric analysis. *Management and Economics Review*, 4(2). https://doi.org/10.24818/mer/2019.12-08

Guermoui, M., Fezzani, A., Mohamed, Z., Rabehi, A., Ferkous, K., Bailek, N., ... & Ghoneim, S. S. (2024). An analysis of case studies for advancing photovoltaic power forecasting through multi-scale fusion techniques. *Scientific Reports*, *14*(1), 6653.

Guo, L., Cheng, J., & Zhang, Z. (2022). Mapping the knowledge domain of financial decision making: A scientometric and bibliometric study. *Frontiers in Psychology*, *13*, 1006412.

Hearst, M., & Hirsh, H. (2017). AI's greatest trends and controversies. *IEEE Intelligent Systems*, 15(1).

Hu, Y., Yu, Z., Cheng, X., Luo, Y., & Wen, C. (2020). A bibliometric analysis and visualization of medical data mining research. *Medicine*, *99*(22), e20338.

Hussein, D. J., Rashad, M. N., Mirza, K. I., & Hussein, D. L. (2022). Machine learning approach to sentiment analysis in data mining. *Passer Journal of Basic and Applied Sciences*, *4*(1), 71-77.

Islam, S., Ab Ghani, N., & Ahmed, M. (2020). A review on recent advances in Deep learning for Sentiment Analysis: Performances, Challenges and Limitations. *Compusoft*, *9*(7), 3775-3783.

Islam, S., Ab Ghani, N., & Ahmed, M. (2020). A review on recent advances in Deep learning for Sentiment Analysis: Performances, Challenges and Limitations. *Compusoft*, *9*(7), 3775-3783.

Karno, M. R. (2015). *Facilitating resource allocation decision through bibliomining: the case of UTM's library* (Doctoral dissertation, Universiti Teknologi Mara).

Kharde, V. A., & Sonawane, S. (2016). Sentiment analysis of Twitter data: A survey of techniques. *International Journal of Computer Applications*, 139(11), 5–15.

Kolling, M. L., Furstenau, L. B., Sott, M. K., Rabaioli, B., Ulmi, P. H., Bragazzi, N. L., & Tedesco, L. P. C. (2021). Data mining in healthcare: Applying strategic intelligence techniques to depict 25 years of research development. *International journal of environmental research and public health*, *18*(6), 3099.

Kusiak, A. (2018). Smart manufacturing. *International journal of production Research*, *56*(1-2), 508-517.

Leal Filho, W., Yang, P., Eustachio, J. H. P. P., Azul, A. M., Gellers, J. C., Gielczyk, A., ... & Kozlova, V. (2023). Deploying digitalisation and artificial intelligence in sustainable development research. *Environment, development and sustainability*, *25*(6), 4957-4988.

Li, W., Huang, Z., Jia, X., & Cai, X. (2016). Neighborhood based decision-theoretic rough set models. *International Journal of Approximate Reasoning*, *69*, 1-17.

Lundberg, L. (2023). Bibliometric mining of research directions and trends for big data. *Journal of Big Data*, *10*(1), 112.

Luo, T., Chen, S., Xu, G., Zhou, J., Luo, T., Chen, S., & Zhou, J. (2013). Sentiment analysis. *Trust-based collective view prediction*, 53-68.

Maksood, F. Z., & Achuthan, G. (2016). Analysis of data mining techniques and its applications. *International Journal of Computer Applications*, 140(3).

Mannila, H. (2000). Theoretical frameworks for data mining. *ACM SIGKDD Explorations Newsletter*, 1(2). https://doi.org/10.1145/846183.846191

MARcHISottI, G. G., Domingos, M. D. L., & Almeida, R. L. D. (2018). Decision-making at the first management level: The interference of the organizational culture. *RAM. Revista de Administração Mackenzie*, *19*(3), eRAMR180106.

Mariani, M. M., Hashemi, N., & Wirtz, J. (2023). Artificial intelligence empowered conversational agents: A systematic literature review and research agenda. *Journal of Business Research*, *161*, 113838.

Maryoosh, A. A., & Hussein, E. A. (2022). A review: Data mining techniques and its applications. *International Journal of Computer Science and Mobile Applications*, 10(3). https://doi.org/10.47760/ijcsma.2022.v10i03.001

Matta, C. E. D., Bianchesi, N. M. P., Oliveira, M. S. D., Balestrassi, P. P., & Leal, F. (2021). A comparative study of forecasting methods using real-life econometric series data. Production, 31, e20210043.

Millemann, J. A., De Waal, G. A., & Maritz, A. (2022). Connecting the dots: A bibliometric analysis on the consumer innovation–decision process. *International Journal of Innovation Management*, *26*(04), 2250031.

Moorjani, G., & Sadath, L. (2019). Sentiment analysis: A tool for data mining in big data analytics. *International Journal of Innovative Technology and Exploring Engineering*, 8(9), 2125–2131.

Murray, P. W., Agard, B., & Barajas, M. A. (2018). Forecast of individual customer's demand from a large and noisy dataset. *Computers & industrial engineering*, *118*, 33-43.

Mydyti, H. (2021). Data mining approach improving decision-making competency along the business digital transformation journey: A case study – Home appliances after-sales service. *SEEU Review*, 16(1). https://doi.org/10.2478/seeur-2021-0008.

Nagendhra, R. Y., & Jen, C. C. (2024). Bibliometric insights into data mining in education research: A decade in review. *Contemporary Educational Technology*, 16(2). https://doi.org/10.30935/cedtech/14333

Omuya, E. O., Okeyo, G. O., & Kimwele, M. W. (2021). Feature selection for classification using principal component analysis and information gain. *Expert Systems with Applications*, *174*, 114765.

Pacheco-Velázquez, E. A., Vázquez-Parra, J. C., Cruz-Sandoval, M., Salinas-Navarro, D. E., & Carlos-Arroyo, M. (2023). Business decision-making and complex thinking: A bibliometric study. *Administrative Sciences*, *13*(3), 80.

PANȚA, N., & POPESCU, N.-E. (2023). Charting the course of AI in business sustainability: A bibliometric analysis. *Studies in Business and Economics*, 18(3), 214–229.

Pejic-Bach, M., Dumičić, K., Žmuk, B., & Ćurlin, T. (2020). Data mining approach to internal fraud in a project-based organization. *International Journal of Information Systems and Project Management*, *8*(2), 81-101.

Poria, S., Cambria, E., & Gelbukh, A. (2016). Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*, *108*, 42-49.

Pynadath, M. F., Rofin, T. M., & Thomas, S. (2023). Evolution of customer relationship management to data mining-based customer relationship management: a scientometric analysis. *Quality & quantity*, *57*(4), 3241-3272.

Ramageri, B. M. (2010). Data mining techniques and applications. *Indian Journal of Computer Science and Engineering*, 1(4).

Ramezani, F., Strasbourg, M., Parvez, S., Saxena, R., Jariwala, D., Borys, N. J., & Whitaker, B. M. (2024). Predicting quantum emitter fluctuations with time-series forecasting models. *Scientific Reports*, *14*(1), 6920.

Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-based systems*, *89*, 14-46.

Rosamilha, N., Silva, L., & Penha, R. (2023). Competence of project management professionals according to type of project: a systematic literature review. *International Journal of Information Systems and Project Management*, *11*(4), 54-100.

Saad, H., Nagarur, N., & Shamsan, A. (2021). Analysis of Data Mining Process for Improvement of Production Quality in Industrial Sector. *arXiv preprint arXiv:2108.07615*.

Salehan, M., & Kim, D. J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, *81*, 30-40.

Santoso, B., Hikmawan, T., & Imaniyati, N. (2022). Management information systems: bibliometric analysis and its effect on decision making. *Indonesian Journal of Science and Technology*, *7*(3), 583-602.

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, *PeerJ Preprints*. https://doi.org/10.7287/peerj.preprints.3190v2

Umarani, V., Julian, A., & Deepa, J. (2021). Sentiment analysis using various machine learning and deep learning Techniques. *Journal of the Nigerian Society of Physical Sciences*, 385-394.

Upadhyay, A. (2018). A review on bibliometric analysis of data mining. *International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)*, 5(4).

Varajão, J., Trigo, A., Pereira, J. L., & Moura, I. (2021). Information systems project management success. *International Journal of Information Systems and Project Management*, *9*(4), 62-74.

Wiśniewsk, J. W. (2021). Forecasting in small business management. *Risks*, 9(69). https://doi.org/10.3390/risks9040069

Xiuyi, T., & Yuxia, G. (2018). Research on application of machine learning in data mining. In *IOP Conference Series: Materials Science and Engineering*, 392(6). https://doi.org/10.1088/1757-899X/392/6/062202

Zahlan, A., Ranjan, R. P., & Hayes, D. (2023). Artificial intelligence innovation in healthcare: Literature review, exploratory analysis, and future research. *Technology in society*, 102321.

Zhan, M. (2016). Exploring the feasibility of applying data mining for library reference service improvement: A case study of Turku Main Library. Finland: ÅboAkademi University.

Zhao, L., Chen, L., Liu, Q., Zhang, M., & Copland, H. (2019). Artificial intelligence-based platform for online teaching management systems. *Journal of Intelligent & Fuzzy Systems*, *37*(1), 45-51.

Zhisheng, C. (2023). Artificial intelligence virtual trainer: Innovative didactics aimed at personalized training needs. *Journal of the Knowledge Economy*. https://doi.org/10.1007/s13132-022-00985-0