



# Data Mining Tools with Effective Roles in Supporting the Decision-Making Process in the Bank and Financial Sectors: A Critical Review

Hawazen Awad Metair<sup>1\*</sup>, Ibrahim M. M. El Emary<sup>2</sup>

<sup>1</sup> MIS Department, University of Business & Technology, SAUDI ARABIA.

<sup>2</sup> Department of Information Science, College of Arts and Humanities, King Abdulaziz University, SAUDI ARABIA.

\*Corresponding Author (Email: [hmetair@gmail.com](mailto:hmetair@gmail.com)).

**Paper ID: 13A10R**

**Volume 13 Issue 10**

Received 19 April 2022

Received in revised form 09 July 2022

Accepted 16 July 2022

Available online 23 July 2022

## Keywords:

Information harvesting;  
Data mining tools;  
Knowledge management; Artificial intelligence; Banking decision making;  
Financial institutions

## Abstract

Technology is one of the most important factors affecting the banking business environment, especially in modern banking organizations, as it allows adding new capabilities to support the decision-making process at all organizational levels that have an impact on the immediate work environment. The use of data mining technology enhances the knowledge base and provides all the necessary data that helps banks make strategic decisions supported by real and accurate information. Hence, there is an arising need to conduct research that deals with the development of data mining tools. This research aims at shedding light on the role of data mining tools in supporting the decision-making process in Saudi banks. To achieve this goal, the descriptive analytical approach was used, which depends on the evaluation of many works and research papers, from 2012 to 2022. After conducting this survey and evaluating many studies, the study concluded that there is a strong impact of the use of the latest technologies and data mining tools on decision-making and forecasting risks. In fact, these reached results not only have a positive impact but also support decision-making processes, which increases the efficiency and effectiveness of banking operations, especially those related to security, fraud risks, and customers' personal information theft and alike.

**Disciplinary:** Management (Banking, MIS).

©2022 INT TRANS J ENG MANAG SCI TECH.

## Cite This Article:

Metair H.A., Emary I.M.M.E. (2022). Data Mining Tools with Effective Roles in Supporting the Decision-Making Process in the Bank and Financial Sectors: A Critical Review. *International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies*, 13(10), 13A10R, 1-11. <http://TUENGR.COM/V13/13A10R.pdf> DOI: 10.14456/ITJEMAST.2022.207

# 1 Introduction

Data is one of the most important resources for banks, as they have to know how to reveal the value of knowledge hidden in raw data. Data mining tools allow extracting knowledge from historical data and predicting the outcome of future events, which helps improve business decisions. The amount of data collected by banks has increased rapidly in recent years. Accordingly, current statistical data analysis tools that do not rely on data mining may not be able to keep pace with the huge volumes of data that banks currently collect and analyze accurately. This huge growth in data volumes has led to the need for new data analysis techniques that rely on data mining in order to find the information hidden in this data and then transform it into knowledge.

Many previous studies dealt with data mining, its tools, techniques, and applications. To the knowledge of the researcher, all studies focused on some of the mining tools in each study. The research gap here lies in the fact that there is still no study that combines all mining techniques and tools in the same study. Henceforth, the urgent need for this study arose, as it collects and analyzes the studies most relevant to the topic of the impact of data mining tools on decision-making in the banking sector. Indeed, it highlights the studies that carry the greatest amount of data mining techniques and their high ability in decision-making.

For this reason, this study is a starting point for many studies that are concerned with knowing the most appropriate and latest technologies that extract knowledge from large data in the databases of banks and banking institutions, to support strategic decisions and raise efficiency and competitiveness among banks. This study answers the question that many researchers are concerned with: What is the role of the latest data mining techniques and their relationship to the decision support process in various financial institutions?

## 2 The Concept of Data Mining

Data mining as explained by Taylor [19] is the process of finding potentially useful patterns from large data sets. In addition, an interdisciplinary skill that uses machine learning, statistics, and artificial intelligence to extract information to assess the likelihood of future events. Insights from data mining are used for marketing, fraud detection, scientific discovery, etc. Data mining is concerned with discovering hidden, unexpected, previously unknown but valid relationships between data stored in an organization's databases.

### 2.1 Data Mining Implementation Stages

The process of implementing data mining goes through six stages as explained by many authors such as Gupta [6], Rohanzadeh & Bamani [16], and Taylor [19]. Figure 1 presents the data mining implementation stages.

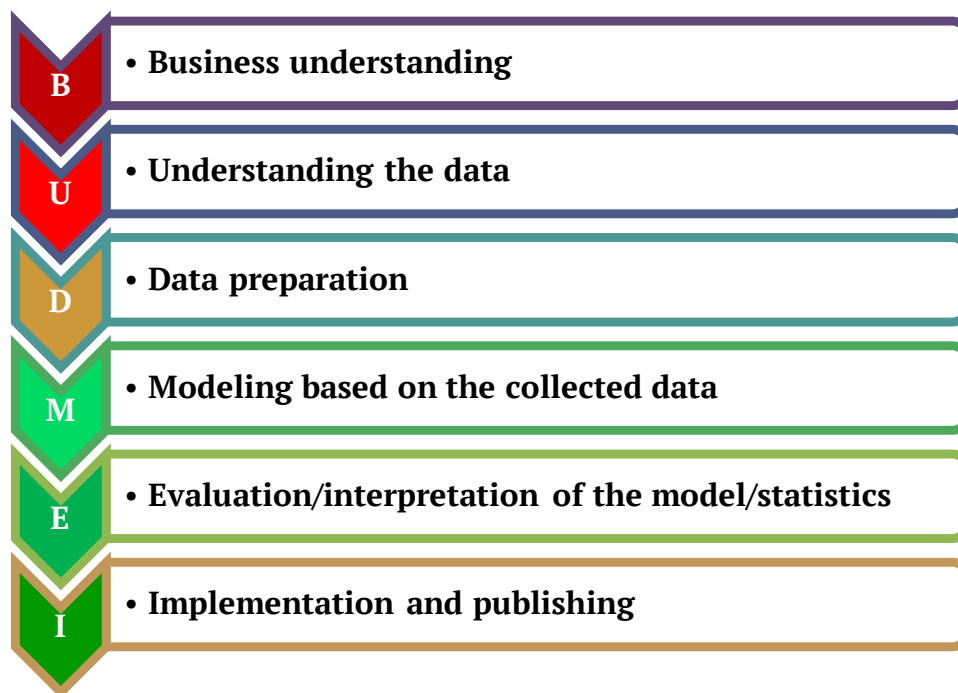


Figure 1: Data Mining Implementation Stages.

### 3 Data Mining Techniques and Algorithms

Data mining is concerned with extracting useful information from a large amount of data, whereby some techniques and methods are applied to large sets of data to extract this useful information.

There are eight data mining techniques [1,8], as the following.

**1. Classification:** This data mining method helps classify data into different categories that are easy to use and search for later. One of the most famous algorithms designed for classification technology is the decision tree algorithm. There are many examples of decision tree algorithms such as CART, C4.5, Quest, and Chaid.

**2. Clustering:** Clustering analysis is a technique used to extract and identify similar data together. One of the most famous algorithms for clustering is the Artificial Neural Network (ANN) algorithm. ANN is a set of algorithms loosely based on human brain functioning and designed for pattern recognition [8].

**3. Association Analysis:** This data mining technique helps in finding the association between two or more items and discovering a hidden pattern in the data set. Perhaps the most famous algorithm is the statistical modeling algorithm such as the *Apriori Algorithm* which was introduced to find sets of repeating elements in a logical association database dataset. The sets of recurring items identified by Apriori can be used to define association rules that highlight general trends in the database and are frequently used in, for example, shopping cart analysis.

**4. Deviation Detection:** This type of knowledge extraction technique refers to the observation of data elements in a data set that match an expected pattern or behavior. This technology can be used in a variety of areas such as fraud detection, error detection, etc.

**5. Dependency Modeling:** This knowledge extraction technique helps in discovering or identifying similar patterns or trends in transaction data for a particular period. One of the most popular algorithms in dependency modeling is genetic algorithms. [19]

6. **Prediction:** Prediction uses a combination of other techniques to extract information such as trends, sequential patterns, grouping, classification, etc. This technique analyzes past events or states in a correct sequence to predict a future event.

7. **Visualization:** It is the representation of data in a drawn way using visual elements such as charts, graphs, maps, tables, and others. The most important algorithms of this technology are:

8. **Summarization:** Summarization refers to the breakdown of data blocks into metrics that provide a general description of variables and their relationships. Some programs are used to access a summary of the data, and the most famous of these are WEKA and Orange. [5]

Figure 2 shows the types of data mining divided into two parts: Descriptive and Predictive.

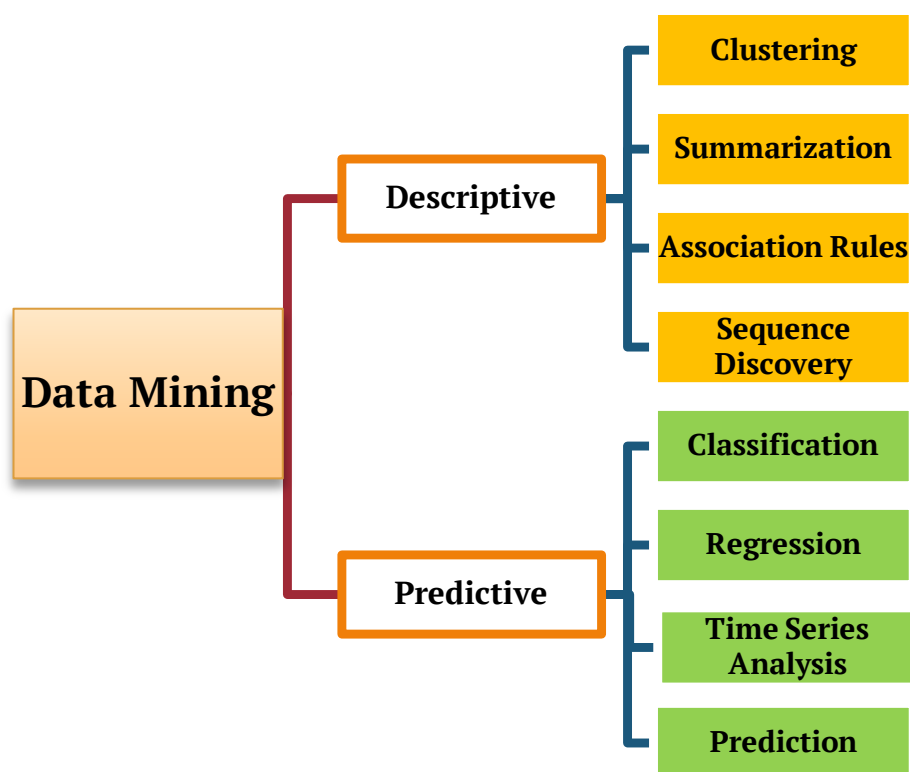


Figure 2: Types of data mining techniques.

## 4 Previous Studies

This section includes the collection of recent previous literature, which shed light on the tools of exploration in the banking environment and the latest technologies used and their impact on supporting decision-making in banks, raising the efficiency of banking services, and increasing competition among banks. These studies cover the period from 2012 to 2022.

Nobanee et al. [12] reviewed the existing literature on the applications of big data in banks using the bibliometric approach. The most important finding of the research is that there are 60 papers related to big data in banks as the sector is growing at a tremendous speed and the number of research outputs in this field is limited. This study is very close to ours, as both of them look at the number of previous research works in the field of studying exploration tools in the banking sector. But the difference between the two pieces of research is that our study looks at studies that

include techniques used in data mining in the banking environment, and not only in mining tools in general, as does the study of Nobanee et al. [12]. The study recommended the necessity of paying attention to future research in the banking sector and big data analytics using data mining tools.

Hanif [7] developed a digital dashboard to facilitate interaction with the results of Explainable Artificial Intelligence (XAI) algorithms and to discuss how the proposed XAI method can be significantly improved. The main challenge for users is to understand and trust the results of AI algorithms and methods appropriately, which explains the need for this study. The study relies on descriptive and experimental methods. The study concluded that the proposed evidence-based process set is a set of processes that go forward to support the organization with a data mining tool to facilitate daily interaction with the results of the machine learning system and help in obtaining results with the powerful visualization system. This study assists our current research by providing a working model for creating a digital dashboard. The study recommended the necessity of developing the prototype and improving the project by obtaining several techniques for selecting the necessary process from the control panel to compare the results of different features in the same setup.

Cheng et al [4] integrated the decision tree algorithm with the Apriori algorithm and explore the relationship between financial ratio, corporate governance, and stock return to establish a stock investment decision model. The need for this study was raised from the fact that the financial data specific to the share price of the facility in the stock market reflects the most direct information for the company, and therefore this share price is an important reference indicator for assessing the value of the company. The study is based on the analytical method, which performs a narrative statistical analysis of the collected data. One of the most prominent outputs and conclusions of the study is the presentation of a proposed systematic model to explore the relevant data through the use of the decision tree algorithm and the Apriori algorithm to detect implicit investment knowledge. The study recommended the completion of research in data mining techniques such as the Bayesian network or the deep learning algorithm.

Amerian [2] focused on the democracy of data, which can be defined as open data within the organization, which aims to justify the importance and role of data democracy in the organization becoming data dependent with a focus on the case of a Swedish bank. The need for this study appeared in that it bridges the gap in the previous literature, which is related to the democracy of data within organizations. The study followed the experimental and analytical approach by studying a sample of data in a Swedish bank. The results of the study show that there is a strong relationship between the benefits of empowering different actors in the organization with the required knowledge of data and accelerating the journey of data-driven transformation. The study recommended the need to share data and the availability of data to a larger number of stakeholders within the organization to better understand the various aspects of problems, simplify data-based decision-making, and make the organization more data-dependent.

Birant [3] proposed a model that improves the decision tree. The researcher followed the experimental method, where the study tested a bank data set and compared it to a decision tree. The study concluded that the proposed approach is superior to the decision tree method when dealing with unbalanced banking data or when it is difficult to find the relationship between them, such as the relationship between high divorce rates and demand for real estate loans. The study recommended the necessity of completing studies on approaches similar to the proposed model, in addition to improving the proposed model for the ordinal classification problem.

Musa et al. [11] examined the role of applying data mining tools in industry and decision-making in Jordanian commercial banks. To achieve this goal, the researchers used the descriptive analytical method based on a questionnaire distributed to the study community. The importance of this study lies in stressing the importance of the role of Jordanian commercial banks in the Jordanian national economic system, and the importance of data mining techniques in managing decision-making processes. The study concluded that there is a significant role for information exploration mechanisms in discovering, distributing and sharing knowledge systems for industry and decision-making in Jordanian commercial banks. The study recommended the need to use data mining techniques, which help in providing useful information and contribute to making sound decisions. This study is very important as it has the same goal, which is to reveal the importance of exploration tools in supporting decision-making processes in banks. However, the difference between this study and our current study is that this study adopted the experimental method using a questionnaire to know the opinions of employees and users of programs in the Banks of Jordan while our current study reveals the importance of prospecting tools in supporting decisions in the financial environment by analyzing and comparing previous studies and applying the empirical approach by comparing and analyzing many prospecting tools.

Preethi & Vijayalakshmi [13] analyzed different data mining techniques and concepts that can be applied to the banking sector to improve its performance. The researchers relied on the descriptive and analytical approach by reviewing previous studies and extracting from them the types of tools used in the banking sector and analyzing these tools to reach their various functions, which would help in conducting the decision-making process quickly and accurately. The study concluded that by using data mining tools, the bank would achieve huge profits with minimal effort and time. The study recommended the necessity of conducting research that includes analyzing algorithms in various exploration tools to raise the efficiency of banks and increase competitiveness.

Mocanu [9] discussed the areas that can be applied in various banking fields such as risk management, credit risk, financial risk, etc. In fact, various data stored in bank databases help to represent the most important and profitable assets. The importance of this paper lies in solving problems in financial and banking institutions by developing patterns of correlation between information so that the results are clear because the volume of data is very large. The researcher used the descriptive method in this paper. The study concluded that many industries, including

banking and telecommunications, use data mining because the used applications contribute to detecting credit card fraud and predicting customer behavior in banking services. The study stressed the need to use data mining tools in banks and financial institutions because of their impact on detecting fraud, acquiring new customers, analyzing current models, and achieving their own market trends, by analyzing data mining tools and focusing on the best applications that contribute to reading and analyzing predictions.

Shree [18] aims at shedding light on the technological methods, support tools and applications used in extracting data from the banking sector. Banking systems collect huge amounts of data daily, whether it is customer information, transactions or risk details, personal files, credit card details and other miscellaneous banking transactions and much other compliance and anti-money laundering information. The importance of this study arose to understand data mining, and how to use suitable tools to meet this challenge. The researcher has adopted the comparative descriptive approach in this paper to describe the data mining tools and compare them with each other. The researcher concluded that the best program in data mining operations is the Natural Language Toolkit (NLTK), because it is the most accurate program in data processing and the fastest in analyzing results, in addition to being easy to learn with a friendly, convenient, and clear user interface. The study recommended the necessity of researching various tools of data mining, commending the previous work in this field, and developing these tools.

Chitra & Subashini [5] analyzed data mining techniques and their applications in the banking sector such as fraud prevention and detection, customer retention, marketing and risk management. The researchers represented the flow of data mining techniques in their system model. This study relied on the analytical and empirical approach, as it analyzed several algorithms to support its objectives. The importance of the study lies in that it highlights the best algorithms in data mining techniques, which greatly help in improving customer relationship management and understanding their behavior, a strategy that can help them build long-term relationships with their customers and increase revenue and profits. The study concluded that data mining techniques are very useful for the banking sector to improve targeting and gaining new customers as well as retention of the most valuable customers, and the automatic approval of credit requests which are used to prevent and detect fraud in real-time. Our current study is related to this research, as both of them discuss the importance of mining tools in the banking environment, especially in supporting decision-making, but this research differs from our study in that it scrutinizes algorithms more than reviewing and analyzing the efficiency of different software.

Pulakkazhy & Balan [15] explored different data mining techniques that can be applied in banking fields in addition to understanding and analyzing information extraction techniques and procedures and how to use these techniques in banking fields to make the decision-making process easier and more productive. The need for this study arose as it sheds light on ways to extract interesting patterns and knowledge from this huge volume of data that can be used in the decision-making process. The study followed the descriptive and empirical approach to test this data mining

technique. One of the most prominent results discussed in the study is that the correlation analysis technique succeeded by 83% and was superior to the techniques used in the study such as clustering, classification and deviation detection. The study concluded that these proposed patterns, which are represented in the statistical modeling algorithm, help the bank predict future events that can help in decision-making processes and that many banks invest in data mining techniques to be more competitive. The study recommended the need to pay attention to the development of data mining techniques and the work of several studies that explore the best techniques for extracting data from banking databases.

Moin & Ahmed [10] overviewed the concept of data mining and highlights the applications of data mining to improve the performance of some basic business operations in the banking industry. The need for this study arose as the banking industry began to realize the need for techniques such as data mining that could help them compete in the banking market. The study followed the empirical approach as it studied 300 cases of credit card fraud in an Indian bank. The study came out with an important result, which is that the classification technique proved its effectiveness and accuracy in 281 fraud cases out of the total number of 300 fraud cases, which indicates its success by 94%. The study concluded that data mining has a great benefit in order to build a data-mining environment for the process of making data to decide which will bring great benefit and competitive advantage in the future. The study recommended the need to support the use of various techniques for data mining to target and gain customers, detect fraud, and provide banking services.

## **5 The Study Problem and the Used Methodology**

The successive developments in technology related to financial institutions, as well as the increase in the flow of large data and related information, have led to the necessity of owning and managing large databases with high efficiency. The literature has indicated the lack of research that examines, analyzes, and compares data mining tools and its most important software used to extract effective knowledge that contributes to supporting decision-making and achieving more competitiveness in the banking sector.

The problem is evident in the banking sector through the high proportion of risks on depositors' money, where banks provide loan and financing facilities of all kinds and then collect money from borrowers. At this level, the wrong decision to finance people with a weak reputation or give them sums that exceed their ability to pay back leads to huge problems and losses for banks, which negatively affects their reputation and the reluctance of the customers to deal with them. Based on these facts, the problem of this paper can be defined as follows: What is the role of data mining tools in extracting the necessary knowledge "from big bank data" to detect credit risks in banks?

This paper relied on the use and application of the descriptive analytical approach by collecting previous studies that cover the period from 2012 to 2022. All of these studies deal with the topic of data mining tools and their impact on decision-making in the banking environment in



order to raise the efficiency of the decision-making process and achieve competitiveness among banks both locally and internationally. The researchers arranged the studies according to the year of their publication, starting with the most recent studies.

## 6 Results and Discussion

All the literature mentioned in this paper agrees on the importance of the role of data mining tools in banks and financial institutions for several reasons, the most important of which is the ability of these tools to predict the behavior of investors in the stock markets and reduce risks such as fraud prevention and detection, customer retention and marketing. All studies indicated that the main reason why the banking sector invested in developing data mining tools is the presence of huge data and little useful information, and this means that institutions and banks cannot use the data in their possession and transform it into information that can be used in decision-making processes, increasing efficiency, and raising competition among them. Therefore, all studies emphasized the use of mining tools and the investment in developing them so that they are able to convert data into useful information that supports decision-making processes, contributes to developing future estimations, and supports the expectations of banks in the financial market.

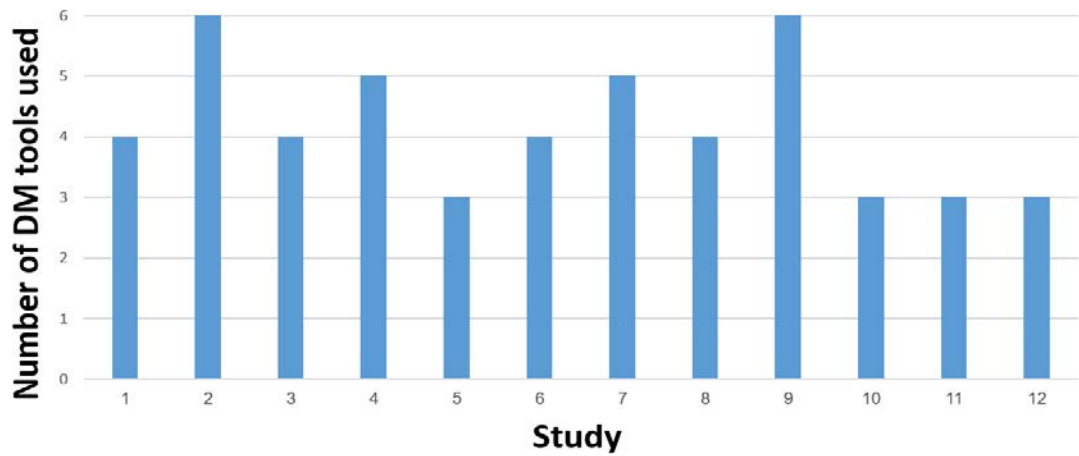
The study [2] was unique in analyzing the democracy of data within the organization, as it explained the method of information movement within organizations and explained the importance of this topic because most studies are concerned with developing data mining tools, oblivious to the tasks performed by these tools. **The study [12]** supported this study, but from the perspective of analyzing the data path within the framework of daily banking operations, and added to the previous one a proposed model for the flow of data.

Studies [8], [12], and [5] explained the high rate of risks in banks, which is the failure of the banking sector in understanding how to use data mining tools and not investing in developing research that sheds light on the mechanisms of digital development in these tools.

The study [12] is the only one of its kind that analyzes users' behavior to find out what's going on in their minds and feelings, which includes investors' fear of an event or users' increased interest in an expected event at an expected time.

The studies [2, 6, 11] differed from previous studies in that they did not focus on data mining tools only, but included analysis of the banking environment and highlighted the exploration of the factors affecting the algorithms of mining tools.

Thus, this study finds an answer to the main question: what is the importance of using the latest data mining techniques and their relationship to the decision support process in various financial institutions? This is illustrated in the graph in Figure 3, which shows the extent of each study's interest in the mining tools that banks rely on in their banking operations.



**Figure 3:** The number of mining tools used in the twelve studies presented in this research

## 7 Conclusion

Although the analysis of data mining tools differed in previous studies, there was not a single study that combines and compares all data mining tools in one paper. In fact, there are eight data mining tools, which are classification, clustering or segmentation, imaging, prediction, correlation analysis, deviation detection, dependency modeling, and summarization. Each of these tools has its own algorithms and programs. This study concluded that there is not yet a single study that combines all eight-tool programs, specifically in the banking and financial sector environment. This in itself represents a deep gap in the world of studies and technical development.

One of the most important future studies is related to the analysis and comparison of all mining tools in one study. This would give a final answer that includes all eight mining tools without exception to know the best data-mining tool for financial and banking institutions that would support the decision to reach the top of competitiveness in the financial sector.

## 8 Availability of Data and Material

Data can be made available by contacting the corresponding author.

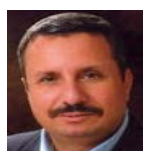
## 9 References

- [1] Ahmed, S., Haque, S. & Tauhid, S. (2014). a detailed Study of Data Transformation for Privacy Preserving in Data Mining. *Int. Journal Science & Engineering Res.*, 5, 1068-1072.
- [2] Amerian, I. (2021). Improving Data-Driven Decision Making Through Data Democracy: Case Study of a Swedish Bank. MM-DD. 41K50E.
- [3] Birant, D. (2020). Data Mining in Banking Sector Using Weighted Decision Jungle Method. *Intechopen*, 91836.
- [4] Cheng, K., Huang, M., Fu, C. Wang, K., Wang, H. and Lin, L. (2021). Establishing a Multiple-Criteria decision-Making Model for Stock Investment Decisions Using Data Mining Techniques. DOI: 10.3390/su13063100
- [5] Chitra, K. and Subashini, B. (2013). Data Mining Techniques and its Applications in Banking Sector. *IJETAE*, 3(8).

- [6] Gupta, G. (2014). Introduction to data mining with case studies. PHI Private learning. 3<sup>rd</sup> Ed.
- [7] Hanif, A. (2021). Towards Explainable Artificial Intelligence in Banking and Financial Services. *CS Journal*, 3(2).
- [8] Linoff, G. (2016). *Data Analysis Using SQL and Excel*. John Willey & Sons.
- [9] Stancu, A. M. R., & Mocanu, M. (2016). Data Mining in the Banking SYSTEM. *Cogito*, 8(1).
- [10] Moin, K. and Ahmed, Q. (2012). Use of Data Mining in Banking. *IJERA*, 2(2), 738-742.
- [11] Musa, K., AlShehadeh, A. R., & Alqerem, R. (2019). The Role of Data Mining Techniques in the Decision-Making Process in Jordanian Commercial Banks. In *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)* (pp. 726-730). IEEE. DOI: 10.1109/JEEIT.2019.8717461
- [12] Nobanee, H., Dilshad, M., Dhanhani, M., Neyadi, M. Qubaisi, S. and Alshamsi, S. (2021). Big Data Applications in the Banking Sector: A Bibliometric Analysis Approach. *SAGE Journal*, 3(4), 1-18.
- [13] Preethi, M. and Vijayalakshmi, M. (2017). Data Mining in Banking Sector. *IJANA*, 8(5), 1-4.
- [14] Pruegkarn, R., Wong, K. W., & Fung, C. C. (2017). A review of data mining techniques and applications. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 21(1), 31-48.
- [15] Pulakkazhy, S. and Balan, S. (2013). Data Mining in Banking and its applications – A review. *Journal of CS*. DOI: 10.1.1.677.4076
- [16] Rohanizadeh, S. & Bameni, M. (2009). A proposed data mining methodology and its application to industrial procedures. *Journal of Industrial Engineering*, 4(4), 37-50.
- [17] Schlagwein, D., Conboy, K., Feller, J., Leimeister, J. M., and Morgan, L. (2017). Openness' with and without Information Technology: A Framework and a Brief History. *Journal of Information Technology*, 32(4), 297–305.
- [18] Shree, D. (2015). Data Mining Tools and Techniques used in Banking Sector.. *CS Journal*, 1(1), 103-110
- [19] Taylor, D. (2022). Data Mining Tutorial: What is Data Mining? Techniques, Process and Tools. *Guru99*, 2(4), 3-9.



**Hawazen Metair** directs the initiative for research in artificial intelligence (AI) at UBT where she works as a lecturer at the College of Business. She was the founder of Saudi Schools in Halifax, NS, Canada. She earned her Bachelor's Degree in MIS from Southern Illinois University Edwardsville, IL, USA. She graduated from Dalhousie University with a master's degree in E-Commerce10. She is currently a PhD candidate at King Abdulaziz University, Jeddah, Saudi Arabia.



**Professor Dr. Ibrahim M. M. El Emary** received a Dr. Eng. Degree from the Electronic and Communication Department, Faculty of Engineering, Ain shams University, Egypt. He is a visiting Professor at King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia. His research interests cover various analytic and discrete event simulation techniques, performance evaluation of communication networks, application of intelligent techniques in managing computer communication networks, and performing comparative studies between various policies and strategies of routing, congestion control, and subnetting of computer communication networks. He published more than 200 articles covering: Computer Networks, Artificial Intelligent, Expert Systems, Software Agents, Information Retrieval, E-learning, Case-Based Reasoning, Image Processing and Pattern Recognition and Robotic Engineering, including 84 articles with high citations according to the Scopus index curve. He is too interested in making a lot of scientific research in wireless sensor networks from the viewpoint of enhancing its algorithms of congestion control and routing protocols.