Information Sciences Letters
An International Journal

http://dx.doi.org/10.18576/isl/120529

Adopting Data Mining as a Knowledge Discovery Tool: The Influential Factors from the Perspectives of Information Systems Managers

A. H. Al-Faouri

Department of Management Information Systems, Al-Balqa Applied University, Al-Salt, Jordan

Received: 22 Feb. 2023, Revised: 22 Apr. 2023, Accepted: 23 Apr. 2023.
Published online: 1 May 2023.

Abstract: Data mining is the process of discovering patterns from large sets of data, based on methods at the intersection of machine learning, statistics, and database systems. As a form of knowledge discovery, the process uncovers concealed patterns to forecast possible results. To meet this objective, this study has applied a cross-sectional quantitative research approach. The data was gathered from managers in the fields of Information Technology (IT) and information systems (IS) of large companies operating in different e-commerce, digital businesses, and marketing in Jordan. The data was then gathered and analyzed. With a total of 309 responses collected in this study, the results were reached using structural equation modeling via Analysis of Moments Structure (AMOS V.21). The proposed conceptual model confirmed that all the identified variables associated with positive coefficients of data mining adoption with data warehouse, data accuracy, perceived usefulness, perceived ease of use, as well as Information System performance. Moreover, the study concluded with research insights related to this topic with further suggested research directed to expand the grasp in this field, and provide deeper understanding of the data mining related issues.

Keywords: Adapting Data mining, Data accuracy, Knowledge discovery, Perceived ease of use, Perceived usefulness.

1 Introduction

The wide spread of information technology, its large usage, and availability have led to a pre-emptive increase in the information volume, that the world hasn't previously witnessed. This has consequently led to emerging issues in the field of the large size of data available on the Internet which has become a controversial topic, in terms of the feasibility of its existence in this random way. Data mining means considering the unimaginable amount of various types and sources of data with a size of hundreds of terabytes or even petabytes [1]. Information exists everywhere and on various topics. However, data importance in terms of the realized facts of the gained information from the data available, only constitutes a small portion: up to 11% compared to the unstructured information that constitutes the rest [2]. This may lead to growing needs that could develop powerful tools that would help in data analysis and extracting key information and knowledge.

Moreover, the traditional tools and methods with the old-fashioned statistics can hardly deal effectively with the huge amount of information. Thus, advanced smart tools are currently applied and used to manage and process the data to make the right decisions [3]. Data mining as an emerged method refers to a technique that mainly derives important knowledge from many different data sources using various mathematical algorithms. This is the base of data mining that can be derived from several science fields such as statistics, mathematics, artificial intelligence, and information systems [1][4]. Smart and non-traditional fields of science encompass areas such as pattern recognition, machine science, and other related fields. Furthermore, data mining appeared in the late 80s and proved its important existence as a good solution for analyzing huge amounts of data by converting it from mere accumulated and incomprehensible information (data) into valued information that can exploit and provide benefits [5][6].

Data mining has attracted a lot of attention among the research community over the past decade, aiming to develop algorithms that are both scalable and adaptable to the increasing amount of data, in search of meaningful cognitive patterns [7]. The algorithms bundle and the software available as well have exponentially grown over the past decades. This has consequently extended its expansion as well as the difficulties arising in the field to keep track of the technologies available to solve a given task [8]. It is worth mentioning that there is a noticeable amount of data coming from various professional resources while searching for definitions of the information mining concept. The many definitions of the concept of data mining indicate that it as a process of searching for data over the available large sources of data to uncover the previously undetected relations between data aspects and elements which are known as Knowledge Discovery Database (KDD) [9] [2].

*Corresponding author e-mail: abeerfaouri1@gmail.com

With the successive technological developments in the field of information systems among different businesses and sectors, and with the driving forces towards digital transformation, companies seek to transform their main operations into electronic aspects [10]. The purpose is to raise the efficiency of the information systems and the ability to manage risks, which would contribute to strengthening the financial positions to keep pace with the changes that take place in the structure of economic activities at the international level [11]. In this regard, research conducted regarding this point found an urgent need to study and evaluate the use of various "data mining" methods adopted in the companies in predicting the perceived benefits to reduce the risks that may result from non-performing decision considering data mining as one application of artificial intelligence that helps in the prediction and accuracy of decision-making. It is based on downloading data and creating quality of the different decisions [12]. Consequently, the research aims to address data mining adoption and the influential factors affecting it from the perspectives of IT and IS Managers in Jordan.

Data mining is the process of discovering patterns from large sets of data, based on methods at the intersection of machine learning, statistics, and database systems [13] [14]. Another definition of this concept presents it as the procedure of data analysis from diverse perspectives and the discovery of the anomalies, data patterns, and correlations that have insightful and useful prediction to assist the organizations and individuals make the right decisions [15]. Since data mining regards the process of searching for data essential, the firms should firstly gather the required data from its sources, prepare it, and store it in a place during data mining. At present, firms store data in what is called a data warehouse that is used by data mining departments to increase their revenues, improve spending, target more customers, provide the best customer services, and listen to feedback to achieve greater competitive intelligence [16].

The discovery of knowledge in the databases is not an easy process, some might think it would stop while collecting and managing data. It would rather extend to the analysis process, to the prediction, and the prediction of what might happen in the future [17]. The forms of data mining are parts of knowledge discovery and they are considered the most comprehensive. The process of discovering knowledge encompasses several steps including data discovery. This stage of data collection includes detection, identification, and characterization of the available data as well as filtering and purifying the data. During the stage of data cleaning, no important data is removed while conflicting and inconsistent data is deleted [18]. While the data integration process focuses on collecting similar and related data from multiple data sources and merging them together, the stage of data selection focuses on the appropriate data to be retrieved from the data set as well as searching for data and retrieving it by completion summary or aggregation operations [19], [20].

However, data mining also refers to the processes of discovering the hidden patterns to predict the potential outcomes. It was previously referred to as "knowledge discovery," as the concept of "data mining" wasn't coined until the 90s. Its foundation overlapped with many different scientific disciplines such as statistics, artificial intelligence, machine learning, and business knowledge [21]. The data that companies gather is meaningless until they uncover insights from it, and data mining enables companies to identify previously hidden patterns and trends within datasets that were not previously discernible [22]. Whenever the insights are uncovered, this might lead to clear decisions, which will benefit both the companies and their customers, and serve other key stakeholders.

Knowledge discovery is also called data mining stemming from the broader arguments that information comes from a variety of media according to the different needs of knowledge access [20]. Knowledge discovery is intended to protect the user from any boring details regarding the original data where meaningful raw data is extracted, and concise knowledge extracted from reports is connected directly to the users [7] [21]. Knowledge-based discovery and data mining are still confusing, and the two terms are usually used interchangeably. Based on knowledge discovery, data will be reduced to a high-level knowledge of the entire process [23]. Furthermore, knowledge discovery can be simply defined as a process to identify valid, novel and potentially useful data, while being the basic pattern of a given operation to be understood [24]. In data mining, data can be thought of as patterns or models of extraction, which would be a general interpretation of the data mining process. Although data mining is the core of knowledge discovery, it is usually seen as one part of the knowledge discovery [25]. Therefore, data extraction is only one step in the entire process of knowledge discovery because there would be many steps to be performed and any step might be included in this process without a precise definition. However, there should be access to a general raw data entry process to select the important data items and limit data set processing and focus. The data is also converted into an appropriate form to find patterns and explain the results [26].

The process of extracting insights from numerous data sources and databases, often housed in large-scale data warehouses, is driven by the objective of identifying meaningful relationships and patterns. This entails sifting through substantial volumes of data and organizing the relevant findings in designated repositories [27]. To facilitate this process, organizations can utilize recognized techniques and statistical methods to analyze data trends and patterns, thereby enabling them to manage these activities seamlessly [28]. This approach is commonly referred to as knowledge exploration over databases. The key step in this process is discovering the knowledge over the databases and the analysis methods included such as neural networks, genetic algorithms, decision network, and technology used in order

to identify those trends, patterns and relations in the data processing setting. Furthermore, the technical procedures that use the algorithms, for instance, to analyze the data from different perspectives to extract the meaningful patterns, can also be applied to predict the future behaviors of the users [29] [30]. In the market analysis, for example, the information systems recommend producing new products to the market considering their previous purchasing behaviors, and this is an example of how data mining is used in the marketing world [31]. Accordingly, it could postulate the following research hypothesis:

Hypothesis 1: Data mining adoption is positively associated with the data warehouse:

The literature also discussed the influence of the quality of the perceived usefulness of data mining. It suggested that the outputs quality is an important aspect and a factor in the use of information systems [24] [32]. Thus, the output quality refers to the extent levels the individuals believe that the information systems perform their job tasks [33]. The perceived usefulness of data mining is also indicated through some research work and the findings supporting this issue through linking other factors that significantly influence the process of data mining adoption and applications [34]. The use of integrated information systems helps to develop product quality levels. It also increases the organization's response speed to the area that occurs in its external environment. The benefits of data mining adoption through databases available to users and controlled through its several techniques, also increases the level of cost information exchange, and contributes to achieving an optimal structure of your product [30]. The increasing use of data mining as one of the information technology innovative tools has positive results. So, the study proposes the following hypotheses:

Hypothesis 2: Data mining adoption is positively associated with the perceived usefulness.

The results of some studies indicate that Technology Acceptance Model (TAM) can assist decision makers in the organization in planning, evaluating, and implementing the use of the e-learning system as it explains and interprets the workers' intentions towards technology and e-learning [35]. On the other hand, the continuous growth of data mining adoption in the market has raised a lot of controversy about self-efficacy and its importance [36]. The success of information technology in decision making and business running provides the users with a certain degree of self-efficacy that may affect information integrity. Therefore, it is important for the provider to explore ways to improve the methods of data mining, and to realize the importance of competency expectations through the contribution of adopting data mining systems [37]. The model also interprets user behavior towards new technologies as a function of perceived ease of use and the perceived usefulness of this technology [38]. Accordingly, it could postulate the following research hypothesis:

Hypothesis 3: Data mining adoption is positively associated with the perceived ease of use.

The accuracy pertains to whether the collected data is precise and genuinely reflects what it is intended to represent. This can measure against the data sources and validity against users defining the business rules. The timeliness means that data is available and expected to facilitate the data-driven for decision making. Data is also the lifeblood of the organizations that forms the base for many important business actions and decisions [39]. However, contemporary organizations focus on the quality of the extensive data and process to ensure a flow process with valid and accurate data to help a clear delivery of valuable results [40]. In today's business environment, companies are leveraging the exponential growth of big data to gain valuable insights. However, before extracting actionable information from the data, companies are realizing the importance of utilizing data quality management frameworks. Such frameworks ensure data consistency and validity, enabling managers to make well-informed decisions based on accurate, current, and complete information.

It is important to note that using high-quality data assists managers in identifying new opportunities and maintaining competitive advantage. Inaccurate financial data, such as overstated earnings, can lead to misleading financial ratios and incorrect evaluations of past outcomes and performance. Therefore, it is essential to base important decisions on accurate and reliable data, including those related to potential target markets and price changes [41].

Furthermore, updated financial statements can assist companies in identifying profitable market segments. This information can help managers explore new opportunities and perform data migration projects more efficiently and accurately. Overall, the adoption of data quality and accuracy frameworks is crucial for companies to make informed decisions, maintain competitiveness, and achieve success in their respective industries [42]. Accordingly, it could postulate the following research hypothesis:

Hypothesis 4: Data mining adoption is positively associated with data accuracy.

Data mining is the discovery of knowledge, it indicates a process used to analyze the main data gathered from different perspectives and sources to extract the associations between them then summarize them into useful information, such as the information contributing towards generating more profit and reducing costs [40] [43]. This approach also represents

a process of finding useful required information through using sets of tools in the information systems. The tools include ordinary statistical tools and artificial intelligence [44]. The data mining life cycle as a type of technological emerged aspect is not clear as the experts take longer time and exert so much effort in conducting research to reach out its maturity and acceptance. Information performance requires analytical tools that depend on data warehouse and analytical processing of information that might occur before data mining, which relies on a multi-dimensional presentation with emphasis on information efficiency and a speedy meeting of the beneficiaries' demands [25] [45]. Accordingly, it could postulate the following research hypothesis:

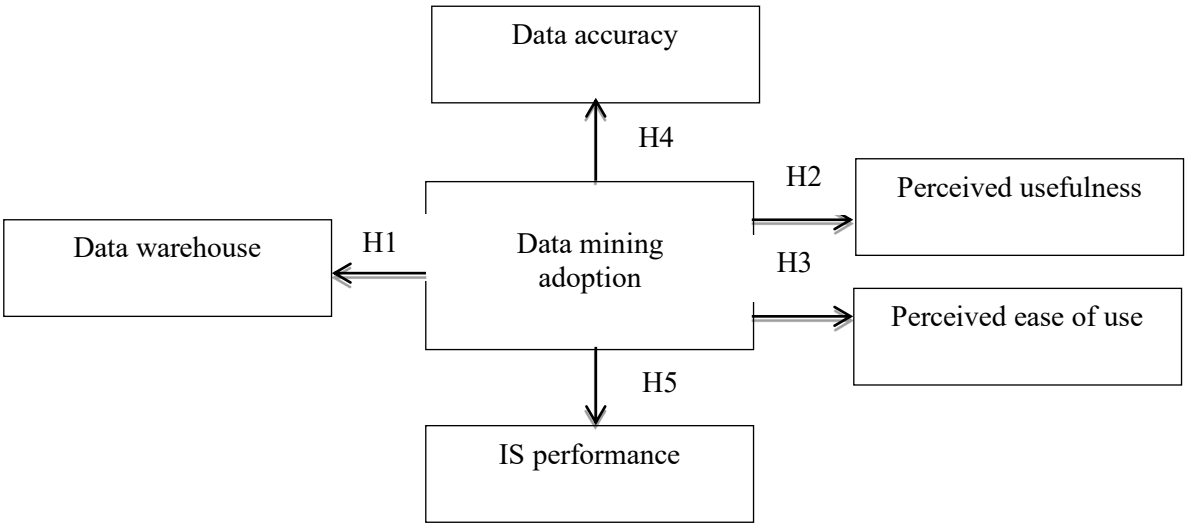

Fig. 1: Conceptual Framework.

2 Materials and Methods

The present research adopted the descriptive method due to its appropriateness for the research and its objectives.

The study is conducted through using quantitative research analysis and approach of cross-sectional by using a survey questionnaire for data collection purposes. Furthermore, the study is connected with the IT and information system IS managers in Jordan due to their real experiences with the different forms of data mining applications and adoption. The overall participants involved in this study reached 309 managers and head of IT department in large companies operating in different industries such as e-commerce, digital businesses, and marketing in Jordan who were invited to participate in this study to fulfil the survey questionnaire. The study instrument was content validated by sending it to the professional in the study scope and area. The experts (e.g., IT specialists) provided important modifications and suggestions to finalize the final draft for a more obvious understanding of the participants. The questionnaire is edited and prepared only in Arabic to ensure lack of uncertainty in the statements hence the respondents are aware of the study objectives. It is divided into two sections related to the demographics and constructing measuring items.

The debates of structural equation modelling analysis considered a sample size of more than 200 observations as an ideal size to conduct such a study according to the approach of maximum likelihood estimation parameters. On the other hand, the study was concerned with analyzing the data through using the Analysis of Moments Structure (AMOS) which assists the statistical procedures though the IBM-AMOS V.21. Thus, the procedures of the data analysis in this approach are generally illustrated with latent variables (or observed based on the nature of constructs model) arrowed into the respective items to show the shape of the relationship between the different variables. Structural Equation Modeling (SEM) analysis is concerned with analyzing the relations between latent research variables and their measurement errors. Since all those variables should be connected and converted into the indicators that measure and rank the responses using the interval scale from 1 to 5, the researcher formulated the interested causal model which is built based on the available relevant theoretical consideration. The path analysis of this model is appropriate to test the research hypotheses which presents the direction of the relations. Furthermore, the study was concerned with conducting descriptive statistics that is performed through the Statistical Package for Social Sciences (SPSS-21) and provided the common outputs in this analysis including mean and standard deviation.

3 Results and Discussions

In this section, the study conducted the key analysis procedures including examining both reliability and validity of the conceptual model through the common tests of composite reliability CR and average variance extracted AVE with

Cronbach's Alpha. The following sections present those important analysis results for further understanding of these aspects in the research study.

3.1 Model Reliability and Validity

The statistical descriptions and results of the study proposed model is given in Table 1. The important aspect in this analysis is called confirmatory factor analysis that was mainly conducted in the SEM analysis to assess the features of reliability and validity of the study measurements. The major results of the conducted analysis confirmed that all model variables are valid and reliable based on the values of CR and Cronbach's Alpha. The results of the reliability were more than 0.70 as suggested by the reliable studies and professionals in this field e.g. [46]. The instrument validity was ranked by using a five Likert scale (1–5). The discriminant validity in this study is determined by the value of average variance extracted where value represents the correlations between the variables. The recommended threshold of the measurement model has an appropriate discriminant validity which should not exceed a value of 0.85. Meanwhile, the value of AVE indicates a different type of validity known as convergent validity, which should be greater than 0.50. There are sets of important fitness indices that should be considered and involved in SEM analysis using AMOS analysis which include chi square divided by the degree of freedom ($\chi 2/df = 2.803$), comparative fit index (CFI = 0.909), Tucker-Lewis index (TLI = 0.912), and Root Mean Square Error f Approximation (RMSEA = 0.077). The results of those indices showed that the study model had a statistical fitness of all respective indices that importantly illustrated the goodness of the model, and this confirmed the construct validity (Table 1).

Table 1: Constructs' Validity

Variable and Item	Factor Loading	CR	AVE	Cronbach's Alpha
Data mining		0.88	0.65	0.84
DM1	0.77			
DM2	0.89			
DM3	0.82			
DM4	0.74			
Data accuracy		0.87	0.69	0.79
DA1	0.87			
DA2	0.82			
DA3	0.81			
IS performance		0.92	0.81	0.80
IS1	0.94			
IS2	0.86			
IS3	0.90			
Data warehouse		0.76	0.52	0.83
DW1	0.74			
DW2	0.71			
DW3	0.72			
Perceived usefulness		0.80	0.58	0.84
PU1	0.69			
PU2	0.78			
PU3	0.81			
Perceived ease of use		0.87	0.69	0.85
PEU1	0.87			
PEU2	0.82			
PEU3	0.81			

3.2 Structural Model Analysis

Before conducting path analysis and hypothesis testing, it is crucial to perform an initial process. This involves conducting a multicollinearity test to ensure that both the model and data adhere to the assumptions of regression analysis. [47]. Regarding this assumption, the study performed an analysis to assess the correlation among the model construct by using one of the common criteria in this aspect known as the Variance Inflation Factor VIF that is associated with the value of Tolerance of all involved variables. The analysis results correlate with the debates stating that the value of the Variance Inflation Factor (VIF) of the respective variable is greater than (10) which indicates in turn that this analysis has no higher levels of correlation among the model variables. Thus, this shows an existing problem within the regression analysis [48]. Therefore, the rule in this issue mainly depends on a problematic issue known as the "Multicollinearity" of these variables. As given in Table 2, the results include the study variables based on the obtained values of the VIF and Tolerance for each variable. It is noted that the respective value of this analysis is

less than (10) and ranged (4.264 - 4.926) which in turn indicates no high variables correlations. Moreover, this confirmed that the model had a normal distribution of the data since the coefficients of the Skewness are lower than (1), the model validity has asserted the analysis of testing the research hypotheses.

Table 2: Multicollinearity test

Variable	VIF	Tolerance	Skewness
Data mining	4.276	0.234	0.883
Data warehouse	4.807	0.208	0.94
Perceived usefulness	4.509	0.222	0.982
Perceived ease of use	4.264	0.234	0.988
Data accuracy	4.926	0.203	0.924
IS performance	4.541	0.220	0.962

The study assessed the structural model that is generally concerned in the SEM analysis and studies. It is mainly performed to check the interested model paths and the relations between the variables, and to test the study hypotheses. The conceptual model (Figure 1) was examined and tested by using SEM analysis through AMOS software with the approach of the maximum likelihood estimate for the parameters of the model. The results of this model through the involved parameters included the weights of the standardized regression (β) and the significance level (p-value) of all the different involved paths. The major analysis results are given in Table 3 and Figure 2. The results revealed the standardized regression and coefficients interpret the direct association among the study variables. It hence asserted the suggested hypothesized relations and hypotheses. Based on the results, data mining is positively associated with the data warehouse ($\beta = 0.27$, $t = 2.460$, $p = 0.000$), and this finding supports H1, so that it is accepted. Data mining is also positively associated with the perceived usefulness ($\beta = 0.39$, $t = 3.093$, $p = 0.000$), and this finding supports H2, so that it is accepted. Data mining is also positively associated with the perceived ease of use ($\beta = 0.32$, $t = 2.088$, $p = 0.001$), and this finding supports H3, so that it is accepted. Data mining is also positively associated with data accuracy ($\beta = 0.40$, $t = 3.937$, $p = 0.000$), and this finding support H4, so that it is accepted. Data mining is also positively associated with the IS performance ($\beta = 0.36$, $t = 3.862$, $p = 0.000$), and this finding supports H5, so that it is accepted. Based on the results given in Figure 2, the structural model is also satisfactory with statistically fit of the measures with the fitness indices and validity. Hence, it can be concluded that the study structural model has confirmed the data.

Table 3: Testing Hypotheses

Hypotheses	Estimates (β)	t-statistic	P-value	Decision
Data mining → data warehouse	0.27	2.460	0.000	Supported
Data mining → perceived usefulness	0.39	3.093	0.000	Supported
Data mining → perceived ease of use	0.32	2.088	0.001	Supported
Data mining → data accuracy	0.40	3.937	0.000	Supported
Data mining → IS performance	0.36	3.862	0.000	Supported

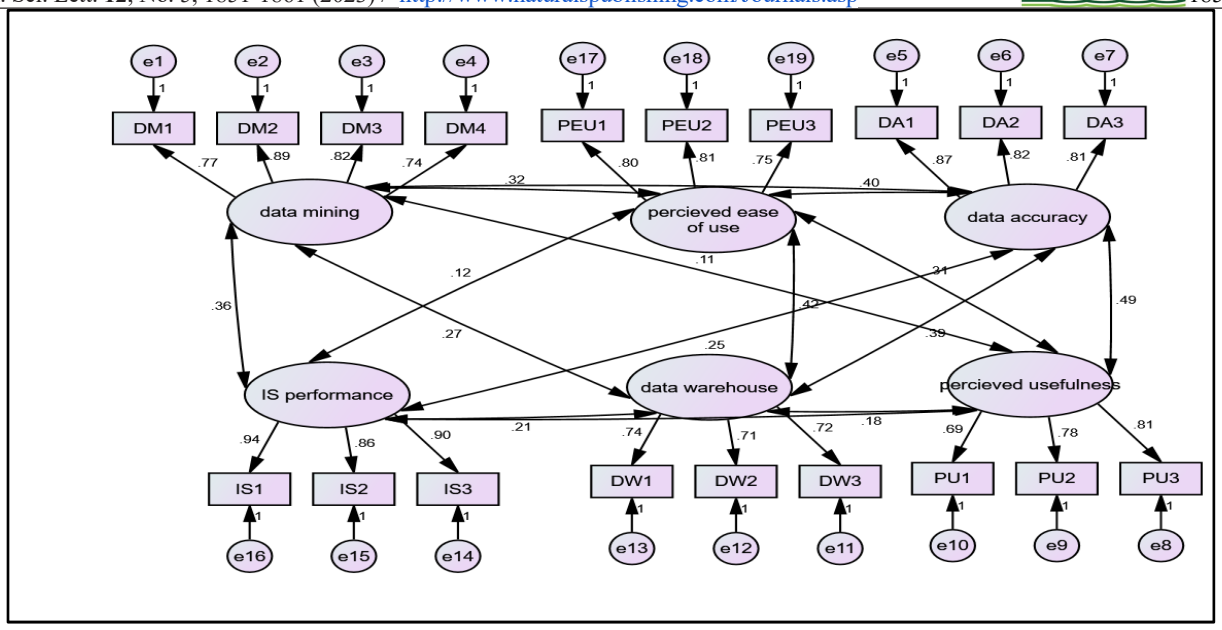

Fig. 2: Structural Model

Based on the gained results, the study found positive coefficients of the standardized weights of the regression. The study findings confirmed the importance perceived of both the usefulness and ease of use of the data mining adoption among large companies in Jordan. The perspectives of the managers in those companies also confirmed the essential aspects of data accuracy and data warehouse through the process of data mining using several methods and adopting it. However, the results of the study provided additional support for the growing concern among modern businesses and firms regarding the implementation of this mechanism. It was found that such implementation led to superior business outcomes and decision-making processes. [40]. Similarly, the study found increasing implementation of data mining strategies while running the key operation to avoid potential failure of making a wrong decision or business direction, since the information and data are the core of successful business activities. Through the obtained findings, the literature also stands in line with the pivotal effects if considering effective methods of data mining with the focus on the factors associated with integrating this mechanism [49]. The present study's findings align with the relevant literature that has established the significant positive impact of data mining in various industries. This confirms the theoretical basis of data mining's significance and supports the current work's results [14]. Based on the foregoing results, it could be inferred that data mining, as a tool for knowledge discovery, contributes to improving the quality of the decisions taken. The possibility of making the best decisions and the decision-making process takes place based on the best information that the individual shares with others. Managers of enterprises usually need support when making their daily important decisions. This is done through providing information about the available options and expected risks. The required data and information differ according to usefulness. When determining whether data is sufficient for decision-making, the process may appear straightforward. However, the complexity arises when making critical decisions, as the type of data required for such decisions differs significantly [50]. The findings would corroborate the hypothesis that the reduction of the decision cycle can be facilitated by the abundance of data and information available, leading to an expeditious decision-making process. This, in turn, results in a reduced time frame for critical decision-making by acquiring new knowledge, predominantly with the aid of computers, capitalizing on the significant advancements in information technology, and leveraging the speed and accuracy of data for the execution of diverse applications. As anticipated, these findings are inexorable [51].

In the same vein, the availability of knowledge that easily leads decision-makers not to assume a cost that exceeds the potential benefits. Data mining is important as it leads to the accumulation of knowledge that increases over time. This means that the more experienced managers need less information which would lower the cost of the decision made, it is consistent with [32] who also confirmed this significance. Data mining is a reconsideration of data from another perspective. Thus, it added value to obtaining more useful information. It has also been used in many fields and more broadly in different sectors e.g., banks, insurance, trade, pharmaceutical companies, etc. As a process, data mining has many techniques through which hidden trends and patterns can be discovered in large amounts of data. The results also debated that data warehouse has been developed to serve as a single database that includes many data while maintaining different databases used to store large-scale of information from different sources and departments in the organization. Then the data analysts deal with the characteristics of data stores, which are objectively oriented and complementary linking data and relationships, as well as non-modifiable and work according to time. Data becomes amenable to be

integrated to overcome any differences or duplication that may result from the different environments of the data sources, which would ensure consistency and accuracy of the results.

Data mining is commonly understood as the process of discovering patterns or generating knowledge from vast amounts of data from various sources such as databases, stores, the web, or other information repositories. The primary purpose of data mining is to identify the causes of certain phenomena and verify theories by analyzing data, thereby identifying new relationships. To achieve this goal, modern enterprises use complex techniques such as statistics, neural networks, and genetic algorithms to extract or mine data. Additionally, the architecture of data mining is an essential aspect that explains the need to perform functions related to data mining. However, creating an environment for data mining systems requires the availability of various units, including data warehouses, as pointed out in [34].

In this study, we also discuss data mining and knowledge discovery, as data mining is a stage in the process of knowledge discovery in databases. The process of data mining involves extracting concepts or deriving unclear knowledge. It is important to note that the retrieval process in data mining is not limited to the data stored in databases but extends to inferential knowledge, which is reached logically through different methods such as deduction and induction. The knowledge extracted from data mining can be directed towards decision makers, virtual decision makers, as well as operational systems and databases, making it a crucial aspect of modern enterprises.

In conclusion, data mining provides accurate, correct, and rapid information. The other advantages that can be reviewed in different sectors would help the decision-maker to activate the interdependence between the various departments and businesses because this is positively reflected in the interest of the business. It can facilitate dealing with advanced information technologies and help measure the effectiveness and productivity of the various sub-information systems through providing accurate and correct information. Furthermore, it assists in the effective use of available data and resources, and in planning for the improvement and development of the used information systems. The findings confirmed the significant role of data mining adoption in revealing the enterprises' ability to grow and keep pace with development, and their need to develop the sources of information systems used as well as the technical and human systems. The enterprise managers have the capability to plan the continuity of the examination procedure, pinpoint issues and their components, and exercise complete authority over the information and data files that are essential for critical decisions.

In recent years, there has been increasing interest among top management in various organizations and businesses regarding the potential benefits and applications of data mining. As a result, there is growing concern about the effects and roles of data in these contexts. Data mining has particularly gained prominence in the distribution sector and has since expanded to other industries. Its initial applications focused on managing customer relationships by analyzing their behaviors and preferences to increase loyalty and introduce new products to meet their needs. Research has demonstrated that data mining has several applications, including marketing activities, utilizing artificial neural networks in business fields such as market shares. These methods provide marketing practitioners and planners with an approach to customer allocation based on demographic factors, such as gender, age, and purchasing patterns. Data mining methods have proven effective in forecasting sales, multiple market variables, and customers' abilities based on their purchasing habits.

Adopting data mining by companies in Jordan has enhanced tremendous outcomes that reflect on the performance among the business operations. Based on the study results, it is suggested for the information systems managers to focus more on the essential experiences and factors associated with the emerging IT revolution through focusing on online media channels to write product reviews, and review consumers' feedback to increase business outcomes such as customers' satisfaction and loyalty. Furthermore, they should exert more effort to integrate data mining methods as knowledge tools that will enable company managers to develop the stakeholders' outcomes. A greater implementation of data mining adoption will in turn develop data processing and management, which would be reflected on the overall business performance. To achieve this purpose, the managers should develop their business enterprises capabilities, as well as the staff's capability to involve within data mining procedures in terms of gaining long-term profitable results with a sustainable competitive edge and increase the market share. Limitations and future research stated in this research are related to the stated aim of this work that was conducted to address the data mining adoption as a tool for knowledge discovering in the companies of Jordan. Hence, the study has some various limitations linked with the nature of research setting, the sample size and the model measurements. To enhance the understanding of this topic, it is recommended that forthcoming research endeavors incorporate a larger sample size and incorporate novel factors that influence the adoption of data mining in diverse contexts, while considering the varying characteristics of the sample.

Conflict of interest

The author declares that there is no conflict regarding the publication of this paper.

References

[1] C. Zhang and J. Han, Data Mining and Knowledge Discovery. *Urban Informatics*, 797-814 (2021).

[2] X. Shu and Y. Ye, Knowledge Discovery: Methods from data mining and machine learning. *Social Science Research*, 102817 (2022).

[3] K. J. Cios and L.A. Kurgan, Trends in data mining and knowledge discovery. *Advanced techniques in knowledge discovery and data mining*, 1-26 (2005).

[4] J. Yang, Y. Li, Q. Liu, L. Li, A., Feng, T. Wang, ... and J. Lyu, Brief introduction of medical database and data mining technology in big data era. *Journal of Evidence-Based Medicine*, **13**(1), 57-6 (2020).

[5] V. Plotnikova, M. Dumas and F. Milani, Adaptations of data mining methodologies: A systematic literature review. *Peer J Computer Science*, **6**, e267 (2020).

[6] A. Viloria, J. Li, J.M. Sandoval and J.V. Villa. Database *knowledge discovery in marketing companies*. In Proceedings of 6th International Conference on Big Data and Cloud Computing Challenges: ICBCC 2019, UMKC, Kansas City, USA, 65-75, (2020).

[7] P. Bhatia, Data mining and data warehousing: principles and practical techniques, *Cambridge University Press*, (2019).

[8] K. Verma, S. Bhardwaj, R. Arya, U.L. Islam, M. Bhushan, A. Kumar and P. Samant, Latest tools for data mining and machine learning, (2019).

[9] R.R. Asaad and R.M, The Concept of Data Mining and Knowledge Extraction Techniques, *Qubahan Academic Journal*, **1**(2), 17-20 (2021).

[10] A.A Freitas, A survey of evolutionary algorithms for data mining and knowledge discovery, *Advances in evolutionary computing: theory and applications*, 819-845 (2003).

[11] F.A Rahman, R. Kassim, Z. Baharum, H.A.M. Noor and N.A. Haris, Data Cleaning in Knowledge Discovery Database-Data Mining (KDD-DM), *International Journal of Engineering and Advanced Technology*, *8*(6 Special Issue 3), 2196-2199 (2019).

[12] M.M. Gaber, A. Aneiba, S. Basurra, O. Batty, A.M. Elmisery, Y. Kovalchuk, M.H.U. Rehman, Internet of Things and data mining: From applications to techniques and systems, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, **9**(3), e1292 (2019).

[13] P. Ristoski, P AND H. Paulheim, Semantic Web in data mining and knowledge discovery: A comprehensive survey, *Journal of Web Semantics*, **36**, 1-22(2016).

[14] Jiang, Y., Yang, G., Li, H., & Zhang, T. (2023). Knowledge driven approach for smart bridge maintenance using big data mining. *Automation in Construction*, *146*, 104673.

[15] N. Kumar, S. Jain, and K. Chauhan, Knowledge Discovery from Data Mining Techniques, *International Journal of Engineering Research & Technology (IJERT)*, *7*(12), 1-3(2019).

[16] C.M. Huerta, A.S. Atahua, J.V. Guerrero, L. Andrade-Arenas, Data mining: Application of digital marketing in education, *Advances in Mobile Learning Educational Research*, **3**(1), 621-629 (2023).

[17] G. Amaral, F. Baiao, and G. Guizzardi, Foundational ontologies, ontology-driven conceptual modeling, and their multiple benefits to data mining, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, **11**(4), e1408(2021).

[18] A. Azevedo, A, Data mining and knowledge discovery in databases. In *Advanced methodologies and technologies in network architecture, mobile computing, and data analytics*, 502-514 (2019).

[19] A.DogaN and D. Birant, Machine learning and data mining in manufacturing. Expert *Systems with Applications*, **166**, 114060 (2021).

[20] J. Fan and D. Li, An overview of data mining and knowledge discovery. *Journal of computer science and technology*, **13,** 348-368 (1998).

[21] A. Viloria, J. Li, J.M. Sandoval, J.V. Villa, Database knowledge discovery in marketing companies. In *Proceedings of 6th* International *Conference on Big Data and Cloud Computing Challenges: ICBCC 2019, UMKC, Kansas City,* USA Springer Singapore, 65-75 (2020).

[22] N. Bhandari and P. Pahwa, Comparative analysis of privacy-preserving data mining techniques. In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2018 Springer Singapore,* **2** .535-541 (2019).

[23] A. Bartschat, M. Reischl and R. Mikut, Data mining tools. *Wiley* Interdisciplinary *Reviews: Data Mining and Knowledge Discovery*, **9**(4), e1309 (2019).

[24] S, A. Mohd Selamat, S. Prakoonwit and W. Khan, A review of data mining in knowledge management: applications/findings for transportation of small and medium enterprises, *SN Applied Sciences*, **2**, 1-15 (2020).

[25] N. Walia and A. Kalia, Tools in Data Mining A Comparative Analysis, **12**, 6(2020).

[26] D. Hartama, A.P. Windarto and A. Wanto, The Application of Data Mining in Determining Patterns of Interest of High School Graduates, In *Journal of Physics: Conference Series IOP Publishing*, **1339**(1), p. 012042(2019, December).

[27] W.W. Chu, Data mining and knowledge discovery for big data. *Studies in Big Data*, **1,** 153-192 (2014).

[28] K. S. Hlaing and Y.M. Thaw, Applications, Techniques and Trends of Data Mining and Knowledge Discovery Database, *Int. J. Trend Sci. Res. Dev*, **3**(5), 1604-1606 (2019).

[29] E. Triantaphyllou and G. Felici, *Data mining and knowledge discovery approaches based on rule induction techniques* Springer Science & Business Media, **6** (2006).

[30] M.I. Al-Twijri and A.Y. Noaman, A new data mining model adopted for higher institutions, *Procedia Computer Science*, **65**, 836-844 (2015).

[31] W.M. Yafooz, Z.B. Bakar, S.A. Fahad and M. A, Mithun. *Business intelligence through big data analytics, data mining and machine learning*, In Data Management, Analytics and Innovation: Proceedings of ICDMAI 2019 Springer Singapore*,* **2**, pp. 217-230, 2020.

[32] M. G. Majumder, S.D. Gupta, S. D and J. Paul, perceived usefulness of online customer reviews: A review mining approach using machine learning & exploratory data analysis, *Journal of Business Research*, **150,** 147-164 (2022).

[33] K. Kundu and M. Sharma, M, Data Mining and Techniques, *Emerging trends in big data, IOT and cyber security*, ***33*** ,(2020).

[34] T.C. Huang, C.C Liu, C. C and D.C. Chang, An empirical investigation of factors influencing the adoption of data mining tools. *International Journal of Information Management*, **32**(3), 257-270 (2012).

[35] M. Wook, Z.M. Yusof, Z and M.Z.A. Nazri, Data mining technology adoption in institutions of higher learning: a conceptual framework incorporating technology readiness index model and technology acceptance model 3, *Journal of Applied Sciences*, **14**(18), 2129-2138 (2014).

[36] M.L. Hart, K. Davies, E.J. Barker-Goldie, A. Theron, A, Issues affecting the adoption of data mining in South Africa, *South African Computer Journal*, **2002**(29), 41-48 (2002).

[37] A.M.T. Al-Odat, *The Adoption of data mining technology within accounting information systems in publicly listed companies in Jordan.* PHD thesis, University of Tasmania, Australia, (2013).

[38] H. Talal and S. Saeed, A study on adoption of data mining techniques to analyze academic performance, *ICIC Express Letters, Part B: Applications*, **10**(8), 681-687(2019).

[39] S. Nandhini, P. Khanna, S. Jain, R. Pal, An integrated approach to encompass Big data cloud adoption using data mining, *Journal of Computational and Theoretical Nanoscience*, **17**(8), 3383-3388(2020).

[40] F.Z. Maksood, G. Achuthan, Analysis of data mining techniques and its applications, *International Journal of Computer Applications*, **140**(3), 6-14(2016).

[41] A. Massaro, V. Vitti, A. Galiano and A, Morelli, Business intelligence improved by data mining algorithms and big data systems: an overview of different tools applied in industrial research, *Computer Science and Information Technology*, **7**(1), 1-21 (2019).

[42] H. Oliff and Y. Liu, Towards industry 4.0 utilizing data-mining techniques: a case study on quality improvement, *Procedia Cirp*, *63*, 167-172(2017).

[43] M.R. Kraft, K.C. Desouza and I. Androwich, *Data mining in healthcare information systems: case study of a veterans' administration spinal cord injury population,* In Proceedings of the 36th Annual Hawaii International

Conference on System Sciences, IEEE., 9 (2003, January).

[44] M. Dagaeva, A. Garaeva, I. Anikin, A. Makhmutova, and R. Minnikhanov, Big spatio-temporal data mining for emergency management information systems, *IET Intelligent Transport Systems*, **13**(11), 1649-1657(2019).

[45] L. Duan, W.N. Street and E. Xu, Healthcare information systems: data mining methods in the creation of a clinical recommender system, *Enterprise Information Systems*, **5**(2), 169-181 (2011).

[46] J.F. Hair, W,C. Black, B.J. Babin and R.E. Anderson, *Multivariate data analysis*, Upper Saddle River, NJ:Pearson Prentice Hall, 7th ed., (2010).

[47] M. Sarstedt, C.M Ringle, J.H. Cheah, H. Ting, O.I. Moisescu, L. Radomir, Structural model robustness checks in PLS-SEM, *Tourism Economics*, **26**(4), 531-554 (2020).

[48] J. Daoud, Multicollinearity and regression analysis. *Journal of Physics: Conference Series,* **949(**1), Article ID: 012009. https://doi.org/10.1088/1742-6596/949/1/012009. IOP Publishing (2017).

[49] A. Sarhan, Data Mining in Internet of Things Systems: A Literature Review. *Journal of Engineering Research*, ***6***(5), 252-263 (2023)

[50] O. Firas, A combination of SEMMA & CRISP-DM models for effectively handling big data using formal concept analysis based knowledge discovery: A data mining approach. *World Journal of Advanced Engineering Technology and Sciences*, ***8***(1), 009-014 (2023).

[51] P. Sunhare, R. R. Chowdhary & M. K. Chattopadhyay, Internet of things and data mining: An application oriented survey. *Journal of King Saud University-Computer and Information Sciences*, ***34***(6), 3569-3590 (2022).