

## Review of Data Mining Concept and its Techniques

Ashour A N Mostafa, Hamdi Elmburok Abdulghani Mahmoud

Senior Lecturer at Computer Technology Department, The Higher Institute of Science and  
Technology - Tobruk, Tobruk, Libya

Email: Ashour.Mostafa@thist.edu.ly, hamdi.sharei33@thist.edu.ly

To Link this Article: <http://dx.doi.org/10.6007/IJAROSS/v12-i6/13135>

DOI:10.6007/IJAROSS/v12-i6/13135

**Published Date:** 07 June 2022

### Abstract

Nowadays, some of the interesting roles of human life are data, information, and knowledge. Analyzing and modelling of big data have been required by data massive storehouses together with the rapid technologies growth to predict and analyze the future trends of information. Methodologies and techniques, which are employed into diverse information systems scope, are needed for detection of knowing in the databases. The technology which extracts advantageous information to discover knowledge is called Data Mining. Data mining, it has been defined as discovery of knowledge in data (KDD), it is the disclosure of modalities procedures and other valuable information from considerable sets of data. It has been a tremendous progress in machine learning, artificial agent systems, and decision-making in the expert systems. In the last decades, most of the techniques and applications has been surveyed via the researchers. Those techniques and applications are utilized in distinct areas in daily life like industrialization, education, engineering, commerce and business. Searching last years researches about the review of the most techniques and trends of data mining in multiple areas was the method which is followed in this paper. It has discovered in the learning field as diffusing data mining for educating activities, improvement quality of tasks into manufacturing field, text mining as a technique into research databases and so on. This study collectes a summary of information about the basci concept of Data Mining and its technques which other researchers may need to start their studies in Data Mining field.

**Keywords:** Knowledge Discovery in Database, Data Mining, Data Mining Techniques, Database Management Systems, Data Mining Processes.

### Introduction

The speedy upgrowth of information systems and software technologies lead measure of volumes of data and information to be raised. High storage capacity and large size are required for this. Databases store data which are related and meaningful to each other, as well as inconsistent and unimportant data. To get the data controlled by the administrator of data, applications are needed to manipulate and control those required data. A database

management system is an application which is used for maintaining and manipulating collections of databases. SQL is a Structured Query Language for manipulating and querying a specific required data. However, the main problem with manipulating data is finding knowledge in the database. To put it another way, mining of data methods are those mechanisms, paradigms, processes, function, or algorithms which are utilized to analyze massive amount of stored data inside the databases. The demand is to detect a modalities or principles which represent the data after analysis to discover the knowledge and unobserved relationships via variables which can make models of data.

Improvement of the data mining in assorted areas like machine learning, artificial intelligence, computing software and statistics have led the developers to improve and execute modern techniques methodologies of the data mining in the previous decades. Web mining, text mining, virtual education applications, manufacturing quality improvements, and databases of research publications are trend and area which could benefit from mining of data technologies that assist humans make-decisions. Some techniques and applications of data mining concepts are reviewed through this article. Firstly, the fundamentals of databases and their standards, as well as the advantages of Database Management Systems are described. subsequently, it clarifies the concept of data mining, data mining mechanism, the accomplishing data mining operation. After that, most of the classes which could be utilized by data mining in information world (Elmasri & Navathe, 2014).

### **Materials and Methods**

This paper represents a review article which is entirely based on the review and analysis of other authors' papers and articles to recognize the concepts and techniques of data mining. Therefore, in this essay, there is not a particular method and/or a framework that is used to gain the results. The material and reviewed articles give illustration for the data mining techniques and concepts.

#### **The Databases Concept**

The database systems are the master effectiveness in expansion of the systems. In the social life of us containing education, medicine, engineering, and business and so on, database Systems play a tremendous role. Databases can be explained as systematic combination of structured data or information, typically stored electronically into a computerized device. The beneficent procedure of impersonation which symbolizes data in an organized pattern is the database. It is being extended and spread because of the extension magnitude and information of organizations. For databases creating and administering, developers had built up a gathering of applications which authorize clients to dominate and operate the databases via using that is named a database management system, and abbreviated to (DBMS). DBMSs have processes to facilitate repair of data like definition, structure, doctrinaire and participation. The data types and the constraints must be determined to define the database and store it. This definition is dependent on the scheme or the index of the database. Construction of database means to save the data into the database through an intermediary storage. Database Processing indicates to stratify assignments on the data to backup, updates and queries. Data sharing awards the clients to access the database even if distantly. Finely, maintaining the data within the database is through accomplishing a security technique whether on the data itself or on the database systems (Elmasri & Navathe, 2014).

### Advantages of DBMS

Redundancy control, build up access limitations, supporting effective of search queries, restore and backup of data, interfaces of users, and associations of data are most of the advantages which DBMSs must contain.

The meaning of redundancy is to store the same data on various storages. If there are any changes applied on the data, the other data copies must be updated too. However, there are several issues like effort squandering for diversified stores, storage size and unproportionate of the files. Consequently, to ensure the consistency and storage space, developers and designers have developed normalization of data, moreover, they have controlled redundancy to get better the execution.

To assist the databases administrators to install the excellences and access dominance for clients to guarantee the safeness, restrictions of uncertified access have to be included like missions into the applications. The data search quickness data and information restore through queries have been decisive case in optimization and enforcement. DBMS have to guarantee techniques of search to strengthen the search and to explore the required data. It could be obtained by data catalogue and structure. In the event of a system deadlock, the DBMS must have a backup to restore. Graphical user interfaces must be included in DBMS which are designed by programming languages to enhance the cooperation with DBMS to ease the system exercising (Gillenson, 2008).

The entity-relationship model is the primary model that represents the complex relationships between data and related information, especially the data that needs to be updated, accessed, and controlled (Thalheim, 2013).

Systems of transactions processing and control of concurrency are the master emerging standard of the databases. Due to there are considerable databases and lots of Processes, especially, various clients who are making synchronous pursuances. The systems must include the rapid responses to queries and availability with integrity (Thalheim, 2013).

### Data Mining Concept

Over latest decades, companies have produced a major value of data which were symbolized as files and databases forms. For processing of data, Structured Query Language is that technology of databases which operates the data in supposing the users are conscious of the database schemes and the issue is this. Many procedures and relational tasks could be performed by the users to pick out rows and columns of data from tables. If there are massive information exporters from various files and database, extraction of information can be complicated through the traditional ways.

Hence, Mining of data is a linked connotation to transact together with senior values of data. It is an effective discovery of knowledge from large lots of data depending on the bases and models. Several features of data mining were gathered to RDBMS, nonetheless it is not well integrated with DBMS. Several of mining databases techniques like machine learning, neural networks and genetic algorithms are highlighted in this essay. Furthermore, kinds of applications of data mining with a group of commercial instruments will be shown (Kantardzic, 2011).

There is a technology, which assists to support decision-making with data, called data warehouse. Mining of data and data warehouse together can be utilized to support making of decisions. Data mining is stratified on functional databases which are volumes of terabytes and petabytes. In addition, it could be worked effectively if the collections of data have been summarized by the data warehouse. Furthermore, applications of data mining can bring out modern significant models (Freitas, 2002).

Likewise, data mining is known as a knowledge discovery in database (KDD). certainly, the database is the storage for data. If there are considerable magnitudes of data, The information from those data is needed to be extracted in a format to be symbolized as information. In fact, it is a difficult process to extract information from considerable databases. Data mining is technique of analyzing a lot of data and abstracting it to detect a model and expose the knowledge due to this knowledge which is obtained from information which is taken out from data. Though, statistics, machine learning, pattern recognition, and revolutionary systems have utilized data mining widely. These procedures of data mining indicate to a substantial area of decision-making (Reddy, 2011) (Chamizo et al., 2015). Data mining owns methods, paradigms, mechanisms, and algorithms which could be exercised to excavate modalities of beneficial information and knowledge. It dates back to the 1960s like an artificial intelligence area (Chamizo et al., 2015).

### **Data Mining Techniques**

It is the same connotation over multiple operations have started from data preparation till enforcement. It can be separated into descriptive and predictive According to DM missions or methods like clustering, classification, prediction, association rules and characterization. Descriptive data mining is via discovering the modality of knowledge.

Jobs of DM as *association* principles which are employed by the algorithms to explore the correlations between associated objects to a group with assigned components to another collection (Hand & Smyth, 2001). It aids to prognosticate the coming times actions depend on existing conduct, and to define the combinations which are convenient with others.

*Classification* utilizes model to impart how to assort classes of data. Ordinarily, it uses the supervised learning to construct the paradigm.

*Clustering* depending on the unsupervised learning, consequently, the classification conducts without any pre-practicing. Clustering anticipates the range to assign the similarity of objects which are appropriated to one group. Pair of methods of clustering partitioning and hierarchical (Han & Kamber, 2011).

### **Data Mining Processes**

The procedures of knowledge discovery in database or data mining could be abstracted like follows (Koksal & Testik, 2011):

- *Preparation of data*: to rummage the data which is stored in the database, firstly, the sources of the data must be defined which are needed to be minded. It could be impossible for several databases and data warehouses to be minded; therefore, the data is needed to be sort into classified shape, or decreased data set via sampling to ease the data mining mechanization.

- *Preprocessing of Data*: to consolidate the quality, some of the troubles of consistency, integration, accuracy, and redundancies are required to be settled. They can be accomplished via various stages as data cleaning, transformation, reduction and discretization. Cleaning means to address missed data and to take off duplicate data. Transformation, in the other word, conversion denotes to adjust the data into convenient shape. Reduction can be defined as to compress or to eliminate the needless attributes. The latest phase is to minimize the rates of data representation.
- *Data mining*: it could be described as discovering the patterns which represent the knowledge. It is descriptive and predictive. Predictive means to discover the future values employing some methods like (S-based, and DT-based algorithms). Moreover, ANN-based algorithm, all algorithms assist to predict behaviors.
- *Evaluation*: evaluation conducted after obtaining the results which are performed via various DM methods. It contrasts different outcomes to support to accomplish the decision.
- *Implementation*: yet the results, to execute the results by building a model or framework to produce the decision and define the best decision.

Mining of text is one of the prime areas of data mining which utilized to detect and extractor information using DM techniques like text analysis, clustering of data sets, and visualization (Nagarkar & Kumbhar, 2015).

#### **Trends of Data Mining techniques with its applications:**

According to Pena-Ayala (2014), data mining techniques or methods were categorized by several trends and applications whether in educational field or business or scientific computing as follows (Pena-Ayala, 2014; Liao & Hsiao, 2012):

- *Neural network*: it is known as artificial neural network too. It is a network of neurons utilized for classification. Some of applications are Bayesian, fuzzy, and back-propagation networks.
- *Algorithm architecture*: algorithms are restricted phases of written instructions which are executed to get a result. The best effects vary from algorithm to another based on the architecture. Some of applications are k-means, chi-square, Euclidean distance and support vector machines (SVM).
- *System architecture*: the analysis of the system is to design a model or framework conceptually, in which explains the dynamic flow of the work and enforcement. It contains hardware and software components to analyze design of the system. Some of applications are systems support of decision, cluster analysis, and decision trees.
- *Agent systems*: the concept of agent is independent structures which reads and supervises the environment revisions and learn then perform based on its database. Some applications are intelligent agents, multi-agent systems, and database systems.
- *Modelling*: models often created by quantitative methods to represent the data or the knowledge as XML modeling and meta-learning.

#### **Conclusions**

Several research of literature through last ten years about the techniques and trends applied via data mining have been displayed in this article. Diverse areas of information systems like educational, manufacturing, and research can execute the data mining techniques, methods, algorithms, or its applications to analyze the large amount of data which are stored into the

databases. It assists to protrude modern methodologies in decision-making and expert systems. It has been difficult to find out all techniques or applications for paper published over last five years, but most of the articles used in the review have been studied over the last decade.

#### Critical evaluation of the selected research paper

<u>Title/ Year/ Author</u>	<u>Key issue</u>	<u>Methodology/ theory/ model/ technique</u>	<u>Strength</u>	<u>Weakness</u>	<u>Comment</u>
<b>A review of data mining applications for quality improvement in manufacturing industry / 2011/ Gulser Koksai, et al.</b>	Review of data mining applications that are applied in quality improvement in manufacturing	The methodology is a literature review of DM applications from 1997 to 2007, and selection quality tasks using DM techniques and applications	Identifying quality tasks with DM functions and methods that are related to manufacturing field	The study was conducted from 1997 to 2007, and the current paper is published in 2011. So, there are 4 years as a gap	The gap in selecting best software that can be implemented in a specific corporation according to its strategy and field
<b>A Review on Data mining from Past to the Future/ 2011/ Venkatadri. M. and Lokanatha Reddy</b>	Review various trends of DM in the past, the current, and the future	It is a review with comparison of the DM trends	Help the researchers to focus on the various issues of DM	A weakness in description of DM techniques that used into DM applications	No mention of the classification algorithms of DM and evolutionary computing
<b>Data mining techniques and applications – A decade review from 2000 to 2011/ 2012/ Shu Liao, et al.</b>	It is a literature survey of DM applications and techniques over 10 years	The methodology is by searching in 5 databases to survey DM techniques within 3 areas which are knowledge types,	Reviewing of DM during 10 years to explain the techniques based on 9 categories DM trends that offer various methodologies because of the	Use other methodologies such as social sciences, psychology and cognitive science as alternative methodologies	Some articles has other methodologies in their applications but not indexed into DM techniques. Therefore, one of the limitation is

		analysis types, and architecture types	sources from different disciplines		the author's limited of knowledge. Also, no integration of methodologies of DM was being used
<b>Educational data mining for improving learning outcomes in teaching accounting within higher education/ 2015/ Julia Gonzalez, et al.</b>	Specifying better learning outcomes to accounting students using educational data mining (within higher education)	Using DM technique and virtual learning platform to build a model of learning activities among student, teacher and teacher's activities	Show differences between courses into the universities and online activities to achieve learning outcomes to help teachers to adjust their teaching ways	No incorporating student into the labor market. No designed activities in the teaching guide to enhance skills	The study implies on accounting students within economic faculty
<b>Text mining: An analysis of research published under the subject category 'Information Science Library Science' in Web of Science Database during 1999-2013/ 2015/ Shubhada Kumbhar</b>	Analyzing text mining literature in Web of Science (WoS) under subcategory ISLS by analyzing chronological growth of TM, major countries, institutions and departments	- Using references + citations as gathered data - Search into ISLS search with synonyms of TM keywords - Data analysis by MS Excel and HistCite	Search a literature review of TM during 13 years with refining the search to specify the bibliographic and citations to see the most journal published the literature and which area	Specifying the subcategory of WoS to search about TM literature and not the whole database	Determining WoS database and its ISLS subcategory is a weakness in finding out TM literature

### Acknowledgements

I would like to express my very great appreciation to staff of International Journal of Academic Research in Business and Social Sciences for their valuable and constructive suggestions to develop this research work. Their willingness to give Their time so generously has been very much appreciated.

## References

- Abu Saa, A., Al-Emran, M., & Shaalan, K. (2019). Factors affecting students' performance in higher education: a systematic review of predictive data mining techniques. *Technology, Knowledge and Learning*, 24(4), 567-598.
- Chamizo-Gonzalez, J., Cano-Montero, E. I., Urquia-Grande, E., & Munoz-Colomina, C. I. (2015). Educational data mining for improving learning outcomes in teaching accounting within higher education. *The International Journal of Information and Learning Technology*.
- Dogan, A., & Birant, D. (2021). Machine learning and data mining in manufacturing. *Expert Systems with Applications*, 166, 114060.
- Elmasri, R., & Navathe, S. (2014). *Fundamentals of database systems* (Vol. 7). Pearson.
- Freitas, A. A. (2002). *Data mining and knowledge discovery with evolutionary algorithms*. Springer Science & Business Media.
- Gillenson, M. L. (2008). *Fundamentals of database management systems*. John Wiley & Sons.
- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. Cambridge, MA: Bradford Book.
- Islam, M. S., Hasan, M. M., Wang, X., Germack, H. D., & Noor-E-Alam, M. (2018). A systematic review on healthcare analytics: application and theoretical perspective of data mining. In *Healthcare* (Vol. 6, No. 2, p. 54). MDPI.
- Issad, H. A., Aoudjit, R., & Rodrigues, J. J. (2019). A comprehensive review of Data Mining techniques in smart agriculture. *Engineering in Agriculture, Environment and Food*, 12(4), 511-525.
- Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms*. John Wiley & Sons.
- Koksal, G., Batmaz, I., & Testik, M. C. (2011). A review of data mining applications for quality improvement in manufacturing industry. *Expert systems with Applications*, 38(10), 13448-13467.
- Liao, S. H., Chu, P. H., & Hsiao, P. Y. (2012). Data mining techniques and applications—A decade review from 2000 to 2011. *Expert systems with applications*, 39(12), 11303-11311.
- Nagarkar, S. P., & Kumbhar, R. (2015). Text mining: an analysis of research published under the subject category 'Information Science Library Science' in Web of Science Database during 1999-2013. *Library Review*.
- Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications*, 41(4), 1432-1462.
- Reddy, D. L. C. (2011). A review on data mining from past to the future. *International Journal of Computer Applications*, 975, 8887.
- Rodrigues, M. W., Isotani, S., & Zarate, L. E. (2018). Educational Data Mining: A review of evaluation process in the e-learning. *Telematics and Informatics*, 35(6), 1701-1717.
- Thalheim, B. (2013). *Entity-relationship modeling: foundations of database technology*. Springer Science & Business Media.



