

---

# A Simple Concept on Data Mining: Applications and Techniques

M. A. I. Navid, N. H. Niloy

Department of Science, Ruhea College, Rangpur Division, Bangladesh

## Email address

alveebd2017@gmail.com (MAI Navid), niloynh1997@gmail.com (NH Niloy)

## Citation

M. A. I. Navid, N. H. Niloy. A Simple Concept on Data Mining: Applications and Techniques. *International Journal of Information Engineering and Applications*. Vol. 1, No. 2, 2018, pp. 76-78.

**Received:** January 28, 2018; **Accepted:** February 18, 2018; **Published:** April 10, 2018

---

**Abstract:** Data Mining introduces in clear and simple ways how to use existing data mining methods to obtain effective solutions for a variety of management and engineering design problems. In this article concepts and techniques such as Neural Network, Decision Tree, Clustering, Association Rule, Clustering and many more techniques of Data Mining is reviewed. This paper focuses how different techniques of Data Mining are used in different applications for finding out patterns from the data taken from the data base.

**Keywords:** Data Mining, Data Mining Techniques, Database Management Systems, Data Mining Processes

---

## 1. Introduction

Data mining tools can include statistical models, mathematical algorithms, and machine learning methods such as neural networks or decision trees. Consequently, data mining consists of more than collecting and managing data, it also includes analysis and prediction. The rapid evolution of information systems and software technology has led to increase the amount volumes of data and information. This needs high capabilities of storage and big size. The databases store data that are interrelated and significant, and in addition inconsistency and insignificant data. To manipulate and control the required data, it needs an application that enables the administrator of those data to control it.

Database management systems are the application of maintenance and manipulation of collection of databases. Structured Query Language (SQL) is a structured language that manipulate and query specific required data, but the major issue of manipulation of data is discovery of knowledge in the database. In other meaning, data mining that is techniques, models, methods, tasks, or algorithms used to analyze vast amount of data stored into the databases. The requirement is discovery a pattern or rule represent the data after analysis to discover the knowledge and hidden relationships among variables may create data model.

Advancement of the data mining in various fields such as machine learning, artificial intelligence, soft computing and statistics has guided the researchers to develop and

implement new methodologies of techniques of the data mining in the last two decades. Web mining, text mining, educational virtual platforms, quality improvement in manufacturing and research publication databases are trends and areas can be benefited from data mining techniques to help the human to make decisions. This paper is a review of some techniques and applications of data mining concepts. First, it describes the fundamentals of databases and its standards, and advantages of DBMS. Afterwards, it explains the data-mining concept, data mining techniques or methods, how the processes of data mining are accomplished, then the most categories that can be used by data mining in the informational world [1].

## 2. Materials and Methods

The study was based on secondary data. The secondary data used were collected from different studies of data mining reports, different thesis, Books and other national and international journals.

## 3. Results and Discussions

The major impact in development of the systems is database systems. It plays roles in our social life including medicine, engineering, and business and so on. Database is a container of related data. To represent data in an organized manner, database is the optimized method of representation.

Due to the expansion size of organizations and its information, the database is being expanded and distributed. To create and administer the databases, researchers had created a collection of programs enable users to control and manipulate the databases using the so-called a database management system (DBMS). DBMS has processes to facilitate maintenance of data such as definition, construction, manipulation and sharing. To define the database, it must define the data types and the constraints to store it. It depends on the schema or the catalog of the database. Storing the data into the database over a storage medium means the construction of the database. Manipulation of the database means applying functions on the data to retrieve, update and query. Sharing grants the users to access the database even remotely. Last thing to preserve the data within the database is by implementing a protection mechanism whether on the database system or on the data itself [1].

Most of the benefits that DBMS must have are control of redundancy, setting up restrictions of access, supporting efficient search queries, backup and recovery, user interfaces, and data relationships.

Redundancy means storing the same data on multiple storage. If the data updated, the other copies should be updated. Nevertheless, there are some problems such as waste of effort for multiple store, storage space and inconsistent of the files. Therefore, designers and developers developed data normalization to guarantee the consistency and storage space, in addition, controlled redundancy to improve the performance.

Restrictions of unauthorized access must be added as functions into the application to help the database administrator to setting up privileges and access controls for users to ensure the security. The speed of search of data and retrieve information via queries have become critical issue in optimization and performance. DBMS should ensure search techniques to enhance the search and finding out the required data. It can be acquired through indexing and data structures. Any failure of the system, DBMS must have backup to recovery. To enhance the dealing with DBMS, it must have graphical user interface that designed by a programming language to facilitate the use of the system [2].

Entity relationship model is most important model that represent data and interrelated information specially the complex relationships among data to update, retrieve and control [3].

Transaction processing systems and concurrency control are the major emerging standard of the databases. Because we have a large database and hundreds of transactions, especially, multiple users are making concurrent execution. Systems must have fast response to queries and availability with integrity.

Last three decades, organizations generated a large amount of data that represented in the form of files and databases. To process the data, SQL is a database technology that manipulates the data in assuming the user is aware of the database schema and this is the problem. The user can

implement some operations and relational functions to choose rows and columns of data from tables. If we have huge sources of information from multiple files and database, information extraction would be difficult in the tradition way.

Therefore, data mining is a related concept to dealing with vast amounts of data. It is an efficient knowledge discovery from vast amount of data according to rules and patterns. Some data mining features added to RDBMS, but is not well integrated with DBMS. This report will highlight some techniques of mining databases such as machine learning, neural networks and genetic algorithms. In addition, it will show types of applications of data mining with a number of commercial tools [4].

There is a technology called data warehouse that helps to support decision making with data. Data mining can be utilized with data warehouse to help making of decisions. Data mining is applied on operational databases that are terabytes or petabytes of volumes, and can be operated efficiently if the data warehouse has summarized collection of data. Afterwards, data mining application can extract meaningful new patterns [5].

Data mining is also called a knowledge discovery in database (KDD). As known, data is stored in the database. If we have large volumes of data, we need to extract information from those data in a form to be represented as information. Extraction process of information from large databases is difficult task. Data mining is mechanism of analyzing lots of data and summarize it to discover a pattern and find out the knowledge because knowledge is induced from information that is extracted from data. Therefore, data mining has been used in various areas such as statistics, machine learning, pattern recognition, and revolutionary systems. This process of data mining refers to an important area of decision making [6, 7]. Data mining has methods, models, techniques, and algorithms that can be used to extract patterns of useful information and knowledge. It goes back to 1960's as a branch of artificial intelligence [7].

It is the same concept through multiple processes started from data preparation until implementation. Based on DM functions or methods such as clustering, classification, prediction, association rules and characterization, it can be divided into descriptive and predictive. Descriptive data mining is by finding out the pattern of knowledge.

Functions of DM such as *association rules* that are used algorithms to study the relationships between items associated to a group with items assigned to another group

[8] This helps to predict the future behavior based on current behavior, and to determine the groups that are compatible with others.

*Classification* uses model to learn how to classify classes of data. Usually, it uses supervised learning to build the model.

*Clustering* based on unsupervised learning, so, the classification conducts without any pre-training. Clustering uses distance to specify the similarity of objects that assigned to one group. Two methods of clustering partitioning and hierarchical [9].

The processes of knowledge discovery in database or data mining can be summarized as follows [10]:

**Data preparation:** to mine the data from the database, first it must determine the sources of the data that needed to be mined. It can be not possible of some databases and data warehouses to be mined, so, it needs to format the data into tabular format, or reduced data set using sampling to facilitate the data mining mechanism.

**Data preprocessing:** it required to solve some problems of consistency, integration, accuracy, and redundancies to enhance the quality. It can be accomplished by several steps such as data cleaning, transformation, reduction and discretization. Cleaning is to solve missing data and removing repeated data. Transformation is to setting the data into proper form. Reduction can be compression or eliminating unnecessary attributes. Last step is to reduce the levels of representation of data.

**Data mining:** it is to discover the pattern that represents the knowledge. It is descriptive and predictive. Predictive is to discover the future values using some methods such as (S-based, and DT-based algorithms). Also, ANN-based algorithm, all algorithms help to predict behaviors.

**Evaluation:** evaluation conducted after obtaining the results resulted by different DM methods. It compares different results to help to reach the decision.

**Implementation:** yet the results, to implement the results by building a model or framework to make the decision and determine the best decision.

**Text mining** is one of the major fields of data mining that used to discover and extract information using DM techniques such as text analysis, clustering of data sets, and visualization [11].

Data mining techniques/methods were classified by different trends and applications whether in educational field or business or scientific computing as follows [7, 12, 13]:

**Neural network:** or called as artificial neural network. It is a network of neurons used for classification. Some applications are Bayesian, fuzzy, and back-propagation networks.

**Algorithm architecture:** algorithms are finite steps of written instructions executed to get a result. The best results differ from algorithm to another based on the architecture. Some applications are k-means, chi-square, Euclidean distance and support vector machines (SVM).

**System architecture:** the analysis of the system is to draw a model or framework conceptually, in which explains the dynamic flow of the work and execution. It includes software and hardware components to analyze design of the system. Some applications are decision support systems, cluster analysis, and decision tree.

**Agent systems:** agent concept is autonomous entities that reads and observes environment changes and learn then act 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 such as XML modeling and meta-learning.

## 4. Conclusion

In conclusion we can say that different areas of information systems such as educational, manufacturing, and research can be implemented the data mining techniques, methods, algorithms, or its applications to analyze the vast amount of data stored into the databases. It helps to emerge new methodologies in decision-making and expert systems.

---

## References

- [1] Elmasri, R., & Navathe, S. B. (2014). *Fundamentals of database systems*. Pearson.
- [2] Gillenson, M. L. (2008). *Fundamentals of database management systems*. John Wiley & Sons.
- [3] Thalheim, B. (2000). *Entity-relationship modeling-Fundamentals of database technology*.
- [4] Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms*. John Wiley & Sons.
- [5] Freitas, A. A. (2013). *Data mining and knowledge discovery with evolutionary algorithms*. Springer Science & Business Media.
- [6] Venkatadri, M., & Lokanatha, C. R. (2011). A review on data mining from past to the future. *International Journal of Computer Applications*, 15(7), 19-22.
- [7] Chamizo-Gonzalez, J., Cano-Montero, E. I., Urquia-Grande, E., & Muñoz-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*, 32(5), 272-285.
- [8] Hand, H., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. Cambridge, Mass: MIT Press.
- [9] Han, J., & Kamber, M. (2006). *Data mining: Concepts and techniques*. San Francisco: Morgan Kaufmann Publishers.
- [10] Köksal, G., Batmaz, İ., & Testik, M. C. (2011). A review of data mining applications for quality improvement in manufacturing industry. *Expert systems with Applications*, 38(10), 13448-13467.
- [11] 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*, 64(3), 248-262.
- [12] 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.
- [13] 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.