# Data Mining: An Emerging Approach towards Business Decision Making

*Puja Dutta, **Prasenjit Kundu

_____

**Abstract**

Data mining is a field of computer science. The process of extracting information to identify patterns, trends, and useful data that would allow the business to take the data- driven decision from huge sets of data is called Data Mining. Data mining utilizes complex mathematical algorithms for data segments and evaluates the probability of future events. Data Mining is also called Knowledge Discovery of Data (KDD). It finds out specific data from a long data set. The paper discusses about is data mining, process, techniques, algorithms, applications which are adopting by business enterprises round the globe for better decision making.

**Keywords**: Data Mining, Data Driven Decision, KDD, Business Enterprise

_____

_* Puja Dutta, MBA Student, Institute of Engineering & Management, Kolkata, India._

_Email: pujadutta832@gmail.com_

_** Prasenjit Kundu_

_Associate Professor, Institute of Engineering & Management, Kolkata, India._

_Email: prasenjit.kundu@iemcal.com_

**Introduction**

The process of digging through data to discover hidden connections and predict future trends has a long history (Agrawal, Imielinski & Swami, 1993). Sometimes referred to as "knowledge discovery in databases," the term "data mining" wasn't coined until the 1990s. Data Mining is carried using various techniques such as clustering, association, and sequential pattern analysis & decision tree. Data mining is a process of extraction of useful information and patterns from huge data. It is also called as knowledge discovery process, knowledge mining from data, knowledge extraction or data /pattern analysis.

**Data Mining** is a process of finding potentially useful patterns from huge data sets. It is a multi-disciplinary skill that uses machine learning, statistics, and AI to extract information to evaluate future events probability (Parker, 2004).

Three steps involved are

Exploration

Pattern identification

Deployment

**1.     Process:**

**Business understanding**: Understand the project objectives and requirements from a business perspective, and then convert this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives.

**Data understanding**: Start by collecting data, then get familiar with the data, to identify data quality problems, to discover first insights into the data, or to detect interesting subsets to form hypotheses about hidden information.

**Data preparation**: Data preparation is the process of cleaning and transforming raw data prior to processing and analysis. It is an important step prior to processing and often involves reformatting data, making corrections to data and the combining of data sets to enrich data.

**Modelling**: various modelling techniques are applied, and usually more types of algorithms are used, with different setup parameters (often with some met approach for optimization of

parameters). Because methods have different formats of inputs and other needs, the previous step of data preparation could be repeated in a small feedback loop.

**Evaluation**: Evaluating the performance of a data mining technique is a fundamental aspect of machine learning. Evaluation measures can differ from model to model, but the most widely used data mining techniques are classification, clustering, and regression.

**Deployment**: The last step in the data mining process is to deploy the models to a production environment. Deployment is important because it makes the models available to users so that you can perform any of the following tasks: Use the models to create predictions and make business decisions.

## 2.    Algorithms and Techniques

Various algorithms and techniques like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., are used for knowledge discovery from databases (Berry & Linoff, 1997).

### *Data cleaning and preparation:*
Data cleaning and preparation is a vital part of the data mining process.

### *Tracking patterns:*
Tracking patterns is a fundamental data mining technique.

### *Classification:*
Classification data mining techniques involve analyzing the various attributes associated with different types of data.

### *Association:*
Association is a data mining technique related to statistics. It indicates that certain data (or events found in data) are linked to other data or data-driven events.

### Outlier detection:

Outlier detection determines any anomalies in datasets.

### Clustering:

Clustering is an analytics technique that relies on visual approaches to understanding data.

### Regression:

Regression techniques are useful for identifying the nature of the relationship between variables in a dataset.

### Prediction:

Prediction is a very powerful aspect of data mining that represents one of four branches of analytics.

### Sequential patterns:

This data mining technique focuses on uncovering a series of events that takes place in sequence.

### Decision trees:

Decision trees are a specific type of predictive model that lets organizations effectively mine data.

### Statistical techniques:

Statistical techniques are at the core of most analytics involved in the data mining process.

### Visualization:

Data visualizations are another important element of data mining. They grant users insight into data based on sensory perceptions that people can see.

### Neural networks:

A neural network is a specific type of machine learning model that is often used with AI and deep learning.

***Data warehousing:***

Data warehousing is an important part of the data mining process. Traditionally, data warehousing involved storing structured data in relational database management systems.

***Long-term memory processing:***

Long term memory processing refers to the ability to analyze data over extended periods of time.

***Machine learning and artificial intelligence:***

Machine learning and artificial intelligence (AI) represent some of the most advanced developments in data mining.

## 3.      Applications

Data mining being a promising emergent technology, already started adapting by multiple industries worldwide as a part of their marketing strategies formulation specially for statistics, pattern recognition, and other important analysis and insights finding. Data mining gaining popularity among many businesses because it provides identification of potential customers, their sentiments (Ghosh, Mukherjee & Kundu, 2021) and insisting for smart marketing decisions making (ORALHAN, Kumru & ORALHAN, 2016).

Financial Analysis. The banking and finance industry relies on high-quality, reliable data.
Telecommunication Industry. Expanding and growing at a fast pace, especially with the advent of the internet.
Intrusion Detection.
Retail Industry.
Higher Education.
Energy Industry.
Spatial Data Mining.
Biological Data Analysis.
The predictive capacity of data mining has changed the design of business strategies. Now, you can understand the present to anticipate the future. These are some examples of data mining in current industry.

*Marketing:* Data mining is used to explore increasingly large databases and to improve market segmentation. By analyzing the relationships between parameters such as customer age, gender, tastes, etc., it is possible to guess their behavior in order to direct personalized loyalty campaigns. Data mining in marketing also predicts which users are likely to unsubscribe from a service, what interests them based on their searches, or what a mailing list should include to achieve a higher response rate.

*Retail:* Supermarkets, for example, use joint purchasing patterns to identify product associations and decide how to place them in the aisles and on the shelves. Data mining also detects which offers are most valued by customers or increase sales at the checkout queue.

*Banking:* Banks use data mining to better understand market risks. It is commonly applied to credit ratings and to intelligent anti-fraud systems to analyze transactions, card transactions, purchasing patterns and customer financial data. Data mining also allows banks to learn more about our online preferences or habits to optimize the return on their marketing campaigns, study the performance of sales channels or manage regulatory compliance obligations.

*Medicine:* Data mining enables more accurate diagnostics. Having all of the patient's information, such as medical records, physical examinations, and treatment patterns, allows more effective treatments to be prescribed. It also enables more effective, efficient and cost-effective management of health resources by identifying risks, predicting illnesses in certain segments of the population or forecasting the length of hospital admission. Detecting fraud and irregularities, and strengthening ties with patients with an enhanced knowledge of their needs are also advantages of using data mining in medicine.

*Television and radio:* There are networks that apply real time data mining to measure their online television (IPTV) and radio audiences. These systems collect and analyses, on the fly, anonymous information from channel views, broadcasts and programming. Data mining allows networks to make personalized recommendations to radio listeners and TV viewers, as well as get to know their interests and activities in real time and better understand their behavior (Pawar & Naranje, 2016). Networks also gain valuable knowledge for their advertisers, who use this data to target their potential customers more accurately.

*CRM:* Customer Relationship Management is all about acquiring and retaining customers, also improving customers' loyalty and implementing customer focused strategies. To maintain a proper relationship with a customer a business need to collect data and analyze the information. This is where data mining plays its part. With data mining technologies the collected data can be used for analysis. Instead of being confused where to focus to retain customer, the seekers for the solution get filtered results.

*Fraud Detection:* Billions of dollars have been lost to the action of frauds. Traditional methods of fraud detection are time consuming and complex. Data mining aids in providing meaningful patterns and turning data into information. Any information that is valid and useful is knowledge. A perfect fraud detection system should protect information of all the users. A supervised method includes collection of sample records. These records are classified fraudulent or non-fraudulent. A model is built using this data and the algorithm is made to identify whether the record is fraudulent or not.

*Intrusion Detection:* Any action that will compromise the integrity and confidentiality of a resource is an intrusion. The defensive measures to avoid an intrusion includes user authentication, avoid programming errors, and information protection. Data mining can help improve intrusion detection by adding a level of focus to anomaly detection. It helps an analyst to distinguish an activity from common everyday network activity. Data mining also helps extract data which is more relevant to the problem.

*Lie Detection:* Apprehending a criminal is easy whereas bringing out the truth from him is difficult. Law enforcement can use mining techniques to investigate crimes, monitor communication of suspected terrorists. This filed includes text mining also. This process seeks to find meaningful patterns in data which is usually unstructured text. The data samples collected from previous investigations are compared and a model for lie detection is created. With this model processes can be created according to the necessity.

*Customer Segmentation:* Traditional market research may help us to segment customers but data mining goes in deep and increases market effectiveness. Data mining aids in aligning the customers into a distinct segment and can tailor the needs according to the customers. Market is always about retaining the customers. Data mining allows to find a segment of customers

based on vulnerability and the business could offer them with special offers and enhance satisfaction.

## 4.      Conclusion

Data mining has wide application domain almost in every industry where the data is generated that's why data mining is considered one of the most important frontiers in database and information systems and one of the most promising interdisciplinary developments in Information Technology. Semantics to improve taxonomies and reasoning about data and text. The need to develop new tools to incorporate ad hoc data as seamlessly and effectively as possible.

Finally, it is noted that data can also go beyond numerical and text data. One important type of data is images. The mining of images is quite established in some fields such as medical science (e.g., in the detection and interpretation of patterns in MRI scans) and biometrics/security (e.g., in the identification of faces as well as objects posing threats). The power of data mining will be greatly enhanced if the mining of all types of data (numerical, text, web/internet and image) can be integrated.

## References

1. Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining Associations between Stes of Items in Massive Databases. In ACM SIGMOD (Vol. 93).

2. Parker, Dr. Gary (2004). Data Mining: Modules in emerging fields, Vol 7, CD-ROM.

3. M. J. A. Berry, and G. S. Linoff (1997). Data Mining Techniques. New York: Wiley.

4. Ghosh, S., Mukherjee, W., & Kundu, P. (2021). Opinion Mining: A Silent Warrior for Business Forecast, PENSEE Journal, Volume 51, Issue 03, Manuscript ID: PNS-0321-224, [ ISSN No: 00314773]

5. ORALHAN, B., Kumru, U. Y. A. R., & ORALHAN, Z. (2016). Customer satisfaction using data mining approach. International Journal of Intelligent Systems and Applications in Engineering, 63-66.

6. Pawar, S., & Naranje, S. (2016). 'A Study on Factors Influencing on Buying Behaviour of Customers'. Research journal 2015- Institute of Science, Poona College of Computer Sciences ISSN2394-1774 Issue II