

# Data Mining Techniques and Applications

Mrs. Patil Vrushali<sup>1</sup> Mrs. Madhuri Bidwe<sup>2</sup>

<sup>1,2</sup>DYPCOE, India

**Abstract**— Data mining is the process to find useful patterns from very large amount of data. In this paper we are discussing few of the data mining techniques, methods and algorithms and some of the organizations which have adapted data mining techniques and methods to improve their businesses and are highly successful in doing so. Data mining DM is also known as Knowledge Discovery in Database aka (KDD). It is also defined as the process which includes extracting the interesting, interpretable and useful information from the raw data. There are different sources that generate raw data in very large amount. This is the main reason the applications of data mining are increasing rapidly. This paper reviews data mining techniques and its applications such as educational data mining (EDM), finance, commerce, life sciences and medical etc. We group existing approaches to determine how the data mining can be used in different fields. Our categorization specifically focuses on the research that has been published over the period 2007-2017, this paper conducts a formal review of the concept of data-mining, the standard tasks involve in data-mining, its applications and techniques in day to day fields.

**Keywords:** Data mining Techniques; Data mining algorithms; Data mining applications, Educational data mining, social network analysis

## I. INTRODUCTION

Data mining techniques (DMT) are used to transform raw data to useful information or knowledge. Data itself is nothing, but to process it, is very useful and interesting. There are many advance technologies that use data as useful information intelligently. For example, Knowledge Discovery in Database (KDD) is the process of required output extraction in different formats from raw data. KDD is also defined as the process to view useful patterns in data. A generic and most common diagram of data mining or KDD is shown in Fig.1.

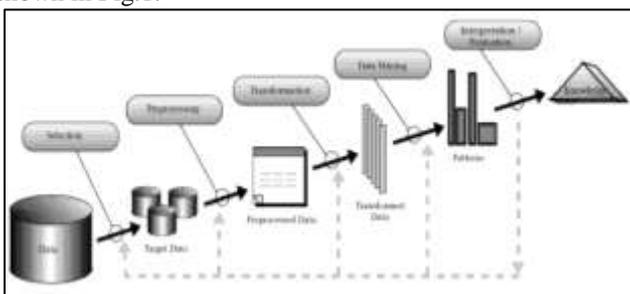


Fig. 1:

Data Mining is a relatively new term in informatics. In simple terms Data mining is the technique of filtering and retrieving relevant data according to business interests from the enormous collection of data using different techniques and algorithms and methods such as Association, Clustering and Classification. In data mining, Knowledge Discovery in Database aka KDD which performs collection, classification and relevant evaluation of data. KDD is an iterative process which consists of the following sequential steps as listed below:

- 1) **Data Integration and Cleaning:** It is the process of gathering data from multiple resources and filtering out irrelevant and not needed data according to the objective of the business enterprise or the demand of the business.
- 2) **Data Selection:** In this we select the relevant data. This becomes the target data which is to be processed.
- 3) **Data Transformation:** In this step, the data is accumulated after sorting and put into some standard formats for mining by methods such as association and clustering.
- 4) **Data Mining:** This is the most crucial step involved in data mining, several data mining techniques are selected as per desired outputs. For example, if one is to find out which group a particular data belongs to, then the method of clustering can be used.
- 5) **Pattern Evaluation:** Pattern evaluation is the process which is used to identify patterns.
- 6) **Knowledge Representation:** This is the final stage in which the knowledge is represented using different visualization techniques.

### A. Data Mining Applications

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it with the outgrowth of different kinds of organizations spanning over several continents the type of data is diversifying. The diversification of such data has led to the application of data mining in the following sectors.

- 1) **Business Sector:** In major businesses, data mining is basically used for analyzing performance, profitability index, finding useful information and customer feedback evaluation, and analysis of the stock values of existing businesses and their market trends to aid in future business decisions making choices.
- 2) **Marketing and retailing sector:** Data mining provide accurate information extraction regarding customer purchase trends, top selling products, customers interest while browsing and purchasing , so that the retail-store managers are able to identify their loyal customers, interested customers and providing discounts and arranging shelves and maintaining according to customer requirements.
- 3) **Bio-informatics:** Gathering and preparing of medical records of the patients to develop a relationship between the disease and the effectiveness of treatment, so that process of curing can be made easy and also assessment of genomic and proteomic data in bio medical fields.
- 4) **Weather Climatology:** Assessment of weather conditions over a period of time to predict future weathers meteorological patterns for determining natural disasters like cyclone and also weather forecasting.
- 5) **Banking & Finance Sectors:** Assessment of individual and multi banking records to generate different marketing strategies for target customer segment, loan approval, stock forecasting, checking various kinds of fraud and money laundering.

## B. Data Mining Techniques

Data mining is a vast field and it has a large number of applications, so it has become interesting subject to research. Data mining techniques are classified into characterization, generalization and association [15]. There are different measures to use data mining techniques as the use of data mining is tricky but helpful if properly used [16]. Some data mining methods are classified and briefly discussed below:

- 1) Clustering Data can be in physical or digital form which is stored as big data. Diverse repositories are used to store such data. A data set exceeding the computational limit of software can be termed as big data. In clustering, groups of different objects and their classes are made on the basis of their different aspects like location; connection etc. For example, schools can be grouped on the basis of their similarities or differences. Similarly, students can be clustered on the basis of their behavior. The purpose of clustering is to search data points that are naturally grouped together.
- 2) Prediction often depends on previous knowledge and experience. It is the focus on a single aspect of data with respect to some other aspect of data, called predictor variable. Prediction is used to predict some unknown result on the basis of previous experience or history.
- 3) Relationship Mining Relation mining also known as relational data mining is commonly used for relational database. In relationship mining, a relationship is discovered among different variable within a data set. In database, relational data mining algorithm search for pattern among different patterns. Relationship between variables must satisfy two things: interestingness and significance [17].
- 4) Outlier Detections generally, if the new observation is different from the existing one compared, it is named as outlier. Outlier detection compares different values with smallest or largest values in a data set and finds the deviation among values.
- 5) Text Mining this data mining technique described as the text data in data mining is specific with text data. Text data include documents, emails, messages, and html files. Text mining can be classified as document processing, document summarization, indexing, topic clustering and mapping.

## II. APPLICATIONS OF DATA MINING METHODS

There are many applications of data mining methods. Some of them are discussed below:

### A. Statistics

In the data mining, user of applications is the main subject. Some tools use form usage statistics that are Access Watch, Web Stat, and Analog. One example of usage statistics is measuring the number of visits. If data produce a relational database, then SQL provide many functions such as sample size and mode. All the techniques convert large data into specific visual display. Commonly, large data are described as charts, graphs and 3D representation. These visualized data can be about assignments, exams, courses and marks. Instructors can get information about their students and distance classes.

## B. Web data mining

Web data mining is also an application of DM. Here, information is filtered from data obtained from web. Web data include web structure, web content and web usage. The main purpose of web data mining is to facilitate users with information they seek.

## III. CONCLUSION

Data being the core entity in every field needs to be managed in efficient way. Data mining helps a lot in this regard. The main issue faced today, is data privacy and data security. In case of global data sharing, privacy becomes more important, especially for web. Therefore, our future work includes the data privacy and security by applying a specific security algorithm that would not harm the data efficiency. It help's in finding the patterns to decide upon the future trends in businesses to grow. Data mining has wide range of applications domain almost in every industry and in every businesses where the data is generated, as data is generated then there is the need of data mining so as to use data effectively and is considered one of the most important frontiers in database and information systems DB systems and one of the most promising pivotal points in interdisciplinary developments in Information Technology.

## REFERENCES

- [1] Jiawei Han and Micheline Kamber (2006), Data Mining Concepts and Techniques, published by Morgan Kauffman, 2nd ed.
- [2] Dr. Gary Parker, vol 7, 2004, Data Mining: Modules in emerging fields, CD-ROM.
- [3] Crisp-DM 1.0 Step by step Data Mining guide from <http://www.crisp-dm.org/CRISPWP-0800.pdf>.
- [4] Customer Successes in your industry from [http://www.spss.com/success/?source=homepage&hpzone=nav\\_bar](http://www.spss.com/success/?source=homepage&hpzone=nav_bar).
- [5] <https://www.allbusiness.com/Technology/computer-software-data-management/633425-1.html>, last retrieved on 15th Aug 2010.
- [6] <http://www.kdnuggets.com/>.
- [7] A. K. Choudhary, J. A. Harding, and M. K. Tiwari, "Data mining in manufacturing: A review based on the kind of knowledge," J. Intell. Manuf., vol. 20, no. 5, pp. 501–521, 2009.
- [8] C.-L. Huang, M.-C. Chen, and C.-J. Wang, "Credit scoring with a data mining approach based on support vector machines," Expert Syst. Appl., vol. 33, no. 4, pp. 847–856, 2007.
- [9] J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan, "Web usage mining," ACM SIGKDD Explor. Newsl., vol. 1, no. 2, pp. 12–23, 2000.
- [10] Data mining white paper, [www.ikanow.com](http://www.ikanow.com)
- [11] Dr. M.H.Dunham, "Data Mining, Introductory and Advanced Topics", Prentice Hall, 2002
- [12] Ha, S., Bae, S., & Park, S. (2000). Web mining for distance education. In IEEE international conference on management of innovation and technology (pp. 715–719).