A study of Data Mining Techniques and its Applications

Jagannath Mahajan
Assistant Professor
Seva Sadan Mahavidyalaya, Burhanpur .

Anish Patel
Assistant Professor
Seva Sadan Mahavidyalaya, Burhanpur .

Anil Choudhary
Assistant Professor
Seva Sadan Mahavidyalaya, Burhanpur

Abstract- With the help of data mining an important and valuable knowledge is extracted from the large massive collection of data. Data mining is very interesting and recent field having various techniques and applications. It is used in finding the patterns to decide future trends in medical field. The paper discusses some of the data mining techniques, applications, algorithms and various areas where data mining technology is widely used to improve their businesses and found best results.
Keywords - Data mining, Information, Data mining techniques, Data mining application, Prediction.

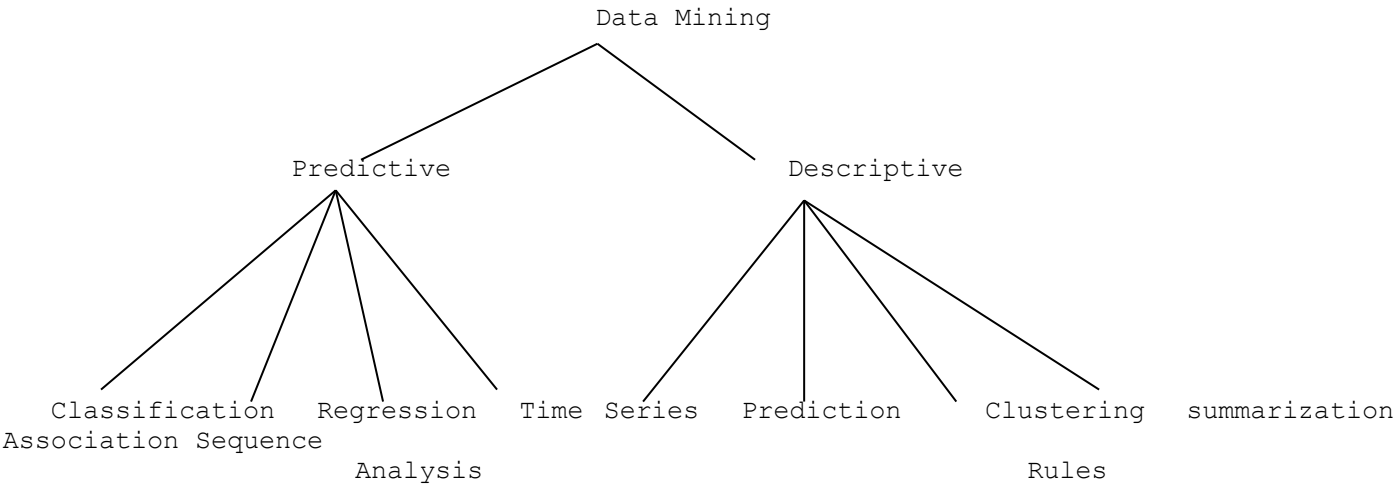
I - Introduction

Data mining is becoming an important tool to convert the data into information. It is commonly used in a wide series of profiling practices, such as marketing, fraud detection and scientific discovery. Data mining is the method of extracting patterns from data. It can be used to uncover patterns in data but is often carried out only on sample of data. The mining process will be ineffective if the samples are not good representation of the larger body of the data. The discovery of a particular pattern in a particular set of data does not necessarily mean that pattern is found elsewhere in the larger data from which that sample was drawn. An important part of the method is the verification and validation of patterns on other samples of data.
Data mining means collecting relevant information from unstructured data. So it is able to help achieve specific objectives. Data mining is a powerful concept for data analysis and process of discovery interesting patterns from the huge amount of data, data stored in various databases such as data warehouse, worldwide web and external sources. Data mining is a type of sorting technique which is actually used to extract hidden patterns from large databases. The goals of data mining are fast retrieval of data or information, knowledge Discovery from the databases, to identify hidden patterns and those patterns which are previously not explored, to reduce the level of complexity, time saving etc. [1]
Data mining includes followings:

- Extract, Transform, and load transaction data onto the data warehouse system.
- Store and manage the data in a multidimensional database system.
- Provide data access to business analysts and information technology professionals.
- Analyze the data by application software.
- Present the data in useful format, such as graph or table.
  There are various techniques and algorithms like clustering; Regression, Artificial Intelligence, Classification, Neural Network, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc. are meant for knowledge discovery from databases [2]. In this paper learns about the data mining techniques and applications.

II - Techniques of Data Mining

Data mining is the process which retrieves useful patterns from large amount of data. This paper discusses some of the data mining techniques, algorithms and some of the organizations which have adapted data mining technology to improve their businesses and found best results. The purpose of a data mining effort is normally either to create a descriptive or predictive model. A descriptive model presents, in concise form, the main characteristics of the data set. The purpose of a predictive model is to allow the data miner to predict an unknown (often future) value of a specific variable, the target variable [3]. The goal of predictive and descriptive model can be achieved using a variety of data mining techniques as shown in figure 1 [4]



1. Classification: Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. Fraud detection and credit risk applications are particularly well suited to this type of analysis. For example we can apply the classification rule on the past record of the student who left for university and evaluate them. Using these techniques we can easily identify the performance of the student.
2. Regression: Regression is used to map a data item to real valued prediction variable [4] In other words regression can be adapted for prediction. In the regression techniques target value are known. For example you can predict the child behavior based on family history.
3. Time Series Analysis: Time series analysis is the process of using statistical techniques to model and explain a time-dependent series of data points. Time series forecasting is a method of using a model to generate predictions (forecasts) for future events based on known past events [5]. For example stock market.
4. Prediction: It is one of a data mining techniques that discover the relationship between independent variables and the relationship between dependent and independent variables [6].Prediction model based on continuous or ordered value.
5. Clustering: Clustering means the act of partitioning an unlabeled dataset into groups of similar objects. The goal of clustering is to group sets of objects into classes such that similar objects are placed in the same cluster while dissimilar objects are in separate clusters. Clustering is used as a data processing technique in many different area including artificial intelligence ,bioinformatics, biology ,computer vision, city planning, data mining, data compression, earthquake studies, image analysis, medicine, statistics and web mining outlier detection i s very important research problem in data mining . Clustering algorithms are used for detecting the outliers efficiently.
6. Summarization: Summarization is abstraction of data. It is set of relevant task and gives an overview of data. For example, long distance race can be summarized total minutes, seconds and height
7. Association: It is one of the most popular data mining techniques. In this technique we mine frequent patterns lead to discovery of interesting association and correlations within data. Example: Association technique is used in marketing analysis to identify items which are frequently purchased within the same transactions.[7]
8. Sequence Discovery: Uncovers relationships among data [4]. It is set of object each associated with its own timeline of events. For example, scientific experiment, natural disaster and analysis of DNA sequence.

III - Data Mining Applications

In this section, we have discussed on applications of data mining. A wide range of companies and various field adapted data mining technologies because of fast access of data and valuable information from a large amount of data. Data mining application area includes healthcare, cloud computing, agriculture, bioinformatics, E-commerce, Banking and finance, education and so on. Some of them are listed below.

1. Data Mining in Healthcare
   Data mining applications in health can have tremendous potential and usefulness [8]. However, the success of healthcare data mining hinges on the availability of clean healthcare data. In this respect, it is critical that the healthcare industry look into how data can be better captured, stored, prepared and mined. Possible directions include the standardization of clinical vocabulary and the

sharing of data across organizations to enhance the benefits of healthcare data mining applications.

2. Data Mining in Cloud computing
   Data Mining techniques are used in cloud computing. The implementation of data mining techniques through Cloud computing will allow the users to retrieve meaningful information from virtually integrated data warehouse that reduces the costs of infrastructure and storage [9]. The cloud in cloud computing refers to the servers, storage, databases, networking, software, analytical provided by the cloud provider. The cloud provider is basically the Internet or data provider.
3. Data Mining in agriculture
   Data Mining in Agriculture: Data mining than emerging in agriculture field for crop yield analysis a with respect to four parameters namely year, rainfall, production and area of sowing. Yield prediction is a very important agricultural problem that remains to be solved based on the available data. The yield prediction problem can be solved by employing Data Mining techniques such as K Means, K nearest neighbor (KNN), Artificial Neural Network and support vector machine (SVM) [10].
4. Data Mining in Bioinformatics
   Bioinformatics generated a large amount of biological data. The importance of this new field of inquiry will grow as we continue to generate and integrate large quantities of genomic, proteomic, and other data [11].
5. Data Mining in E-commerce
   Many E-commerce companies are using data mining business Intelligence to offer cross-sells through their websites. One of the most famous of these is, of course amazon. They use sophisticated mining techniques to drive their people who viewed that product.
6. Data Mining in Banking and Finance
   Data Mining in Banking and Finance: Data mining has been used extensively in the banking and financial markets [12]. In the banking field, data mining is used to predict credit card fraud, to estimate risk, to analyze the trend and profitability. In the financial markets, data mining technique such as neural networks used in stock forecasting, price prediction and so on.
7. Data Mining in Education
   There is newly emerging field called Educational Data Mining. The goals of EDM are identified as predicting students future learning behavior, studying. We use data mining by as institution to take accurate decisions and also predict the results of the student. With the results, the institution can focus on what to teach and how to teach. Learning pattern of the students can be captured and used to develop techniques to teach them.

IV- Conclusion

This Paper provides a brief idea of data mining, Data mining techniques and its applications. The main objectives of data mining techniques are to extract valuable information from data. In this paper we conclude most useful applications. These applications use classification, Prediction, clustering, association techniques and so on.

References

[1] Er. Rimmy chuchra "Use of Data mining Techniques for the evaluation of students performance: A Case study " International journal of computer science and management research Vol 1 Issue 3 october 2012.

[2] Brijesh Kumar Baradwaj, Saurabh Pal " Mining Educational Data to Analyze Students Performance " (IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 2 ,No. 6, 2011.

[3] Nikita Jain, vishal Srivastava "DATA MINING TECHNIQUES: A SURVEY PAPER" International Journal of Research in Engineering and Technology, Vol. 2, Issue 11,Nov-2013.

[4] Dr. M.H. Dunham , "Data Mining ,Introductory and Advanced Topics", Prentice Hall ,2002.

[5] Time Series Analysis and Forecasting with Weka, http://wiki.pentaho.com/display/DATAMINING/

[6] Aakanksha Bhatnagar, Shweta P. Jadye, Madan Mohan Nagar" Data Mining Techniques & Distinct Applications: A Literature Review" International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 9, November- 2012

[7]Aarti Sharma,Rahul Sharma,Vivek Kr. Sharma,Vishal Shrivatava" Application of Data Mining – A Survey Paper" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014

[8] HIAN CHYE KOH , School of Business, SIM University, Singapore

[9] Ruxandra-Ştefania PETRE, "Data mining in Cloud Computing" Database Systems Journal vol. III, no. 3/2012
[10] D Ramesh , B Vishnu Vardhan, "Data Mining Techniques and Applications to Agricultural Yield Data" International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 9, September 2013
[11] Aakanksha Bhatnagar, Shweta P. Jadye, Madan Mohan Nagar" Data Mining Techniques & Distinct Applications: A Literature Review" International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 9, November- 2012
[12] Industry Application of data mining , http://www.pearsonhighered.com/samplechapter/0130862711.pdf
[13] Han and Kamber, "Data Mining and Cocepts."