DATA MINING & ITS TECHNIQUES

Yogesh Bhatia, Sanjeev Verma
Student, Department of Information Technology,
Dronacharya College of Engineering, Gurgaon, Hr, India

Abstract- **In this paper the concept or usage of data mining is summarized & its importance among different kinds of methodologies are illustrated.It also includes various data mining techniques such as clustering,association rules,decision tree,knowledge discovery through neural networks and genetic algorithim,fuzzy set theory which helps for better manipulation of data and further helps pharmaceutical firms to manage their inventories & compete on lower costs while improving the quality of drug discovery and delivery methods. The paper presents how Data Mining finds and extracts the patterns from large amount of data to find observable patterns.**

I. INTRODUCTION

Data mining is about processing data and identifying patterns and trends in that
information so that you can decide or judge.Data mining is the computational process of discovering the different patterns in large data sets which involves different methods to process the data.The overall goal of data mining is to extract information or data from a data set and transform it into an understandable structure for further use. It involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures.The collection of data or information and storage technology has made it possible for various organizations toaqquire large amount of data at a lower cost.The explotion of this stored data to extract useful information is the overall objective of data mining.Data mining provides various methods and techniques like clustering,decision tree,knowledge discovery through neural networks & genetic algorithim,fuzzy techniques which allow extraction from large data collections unknown relationships among the data items that are useful for decision making.Traditional dataanalysis methods involves manual work & interpretation of data which is time consuming,highly subjective and expensive.Data Mining which is also called as knowledge discovery in large data enables firms and organizations to make accurate decisions by accumulating,analyzing and accessing the corporate data.

II. DATA MINING TECHNIQUES

ASSOCIATION RULE: Association is one of the most familiar and straightforward data miningtechnique. Here,you can create a straightforward correlation between two or additional things,typically of constant kind to spot patterns. An example of association rule would be If a customer buys a dozen eggs,then he is 80% likely to also purchase milk.In data mining, association rules are useful to analyse & predict the customer behavior. They play an important role in shopping basket data analysis, product clustering, catalog design and store layout.

Clustering: Cluster is a group of objects that belong to the same class. In other words the similar object are grouped in one cluster and dissimilar are grouped in other cluster. Clustering is the method by which same records are grouped together. Usually this is done to give the end user a high level view of what is going on in the database. During cluster analysis, we first partition the data set into groups based on data similarity and then assign the label to the groups. The main advantage of Clustering over classification is that it is adaptable to changes and help single out useful features that distinguished different groups. Clustering Analysis is widely used in many applications such as market research, pattern recognition, data analysis, and image processing. Clustering is also used in outlier detection applications such as detection of credit card fraud.

Neural network: When data mining algorithms are talked about these days most of the time people are

talking about either decision trees or neural network as this technique is widely & effectively used.This technique uses nonlinear predictive models which allow learning through training across a wide range of diverse problems. Here, the computer system is trained to think, respond and take decisions like humans. However, lots of training is given to the system and only processed information is fed which gradually makes the system efficient to mine and predict different patterns from a database.Neural network models are quite complex to use and deploy . Neural networks are used in a wide variety of applications. They have been used in all facets of business from detecting the fraudulent use of credit cards and credit risk prediction to increasing the hit rate of targeted mailings

Genetic algorithim: The idea of Genetic Algorithm is derived from natural evolution. In this technique concepts of genetics, combinations, natural selection and mutation is used.The basic idea behind this technique is that we can build a better solution if we are able to combine the good parts of another solution. Genetic algorithms promote "survival of the fittest"using heuristic functions. There are two approaches to apply
genetic algorithms –either directly as a classifier or as an optimization tool for resetting the parameters in other classifiers.Genetic algorithms are useful in finding an optimal set of feature weights that improve classification accuracy.

Fuzzy set theory: Fuzzy Set Theory is also known as Possibility Theory. This theory was proposed by Lotfi Zadeh in 1965. This approach is an alternative Two Vlue logic.It allow us to work at high level of abstraction. It also provide various means for dealing with imprecise measurement of data. Applications of fuzzy technology can be found in artificial intelligence, computer science, control engineering, decision theory, expert systems, logic, management science, operations research, robotics, and others.

Decision tree: A decision tree is a predictive model that, as its name implies, can be viewed as a tree. This technique has tree shape like structures which represents sets of decisions generating rules for data set classification. The top node is called as root which is the starting node which is partitioned into two or more
nodes depending on the results of the test.This technique is fast which allow results to be presented as rules with little or no pre-processing of business datadecision tree has various advantages like as it does not require any domain knowledge, easy to assimilate by human,simple to learn.

III. CONCLUSION

In this paper we have concluded that data mining is a wide range application field almost in every industry that's why data mining is considered to be one of the most important frontiers in database and information systems.It helps in finding the patterns, forecasting, discoveryof knowledge etc indifferent business domains.Different data mining techniques such as clustering,decision trees,association rule,genetic algorithim,neural networks etc helps in finding the patterns which decides the future trends in businesses to grow.

REFERENCES

[1] http://www.springer.com/mathematics/book/978-0-7923-7435-0
[2] file:///C:/Users/HCL/Downloads/Documents/7vol3no4.pdf
[3] file:///C:/Users/HCL/Downloads/Documents/V2I10-0156.pdf
[4] file:///C:/Users/HCL/Downloads/Documents/RJCSE-Y12-TJ-CF-D098.pdf
[5] file:///C:/Users/HCL/Downloads/Documents/ijret_110211019_2.pdf
[6] file:///C:/Users/HCL/Downloads/Documents/IJARCCE3E__s_ranshul_a_survey.pdf