



**INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH
TECHNOLOGY**

**A SURVEY ON DATA MINING TECHNIQUES, THEIR APPLICATION AND FUTURE
SCOPE**

Yash Jain*, Shreyank Chaubey, Harsh Agrawal

ABSTRACT

Data mining techniques are rapidly growing and used by many researchers in past few years. The main advantage of using data mining is to extract the knowledge from a huge amount of data which stores in variety of data repository. Knowledge is suggesting the significance through direct or indirect. In this paper, we have included the survey of various data mining techniques. We will also discuss the application areas and its future scope.

KEYWORDS: Mining of data, Techniques of Data mining, Applications of Data mining, association and clustering.

INTRODUCTION

The data mining database may be a logical rather than a physical subset of data warehouse, presented that the data warehouse data base management system can support the additional resource demands of data mining. If it is not the case, then you will use with a separate data mining database. Data mining refers to extracting or mining the knowledge from huge amount of dataset which are restored in different repositories. Now a day's it is possible for any organizations to build up huge amount of data using data collection and storage technology at lower cost. Utilizing this stored data, in categorize to extract functional and useful information, is the overall mission of the standard activity termed as data mining. In other words, we can say that Data mining is the process of exploration and analysis, by regular means, of large quantities of data in a way to discover noteworthy patterns and rules. If we talk about the real world than the huge amount of data are available in the field of education sector, medical sector, industry and many other areas. With the help of such amount of data, we can generate some sort of information and knowledge which helps us in decision making. For example, you can find out Alumni student in any university, with the help of shopping database, we can identify how many items sold out. Data can be examined/analyzed, summarized, understand and meet to challenges.[1] Data mining is a powerful concept/ model. With the help of this we can not only analyzes the data but also process to discovery interesting pattern from the huge amount of data. Data can be stored anywhere in various databases such as data warehouse, World Wide Web, external sources .Interesting pattern that we are generating by applying some of the data mining techniques are easy to understand, unidentified, applicable, possibly useful. In other words, we can say that the data mining is a type of sorting technique which is actually used to extract hidden patterns from large data repositories. The main goal of data mining is to efficiently retrieve of data, information, knowledge Discovery from the databases; we need to recognize unseen patterns and those patterns which are previously not considered/ explored, to reduce the level of complexity, time saving, etc [2].

Various algorithms and techniques like Classification technique, Clustering technique, use of Regression technique, use of Artificial Intelligence technique, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., are meant for knowledge discovery from databases [5]. The main objective of this paper learns about the data mining.

DATA MINING TECHNIQUES

- a) **Classification:** Classification based on categorical (i.e. distinct, unordered). This technique based on the supervised (controlled) learning (i.e. for the given input, the desired output is already known to us) .It can be classifying the data based on the training set and values. We can achieve the goals by using a decision tree, neural network and classification rule (IF- Then).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.

- b) **Regression:** Regression is used to map a data item to a real valued prediction variable [8]. In other words, regression can be adapted for prediction. When we use regression techniques target values are known. For example, you can predict the student behavior based on university marks.
- c) **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 [9]. For example stock market.
- d) **Prediction:** It is one of a data mining techniques that discover the relationship between independent variables and the relationship between dependent and independent variables [4]. Prediction model based on continuous or ordered value.
- e) **Clustering:** The goal of clustering is to organize data by finding some 'sensible' grouping of the data items. Clustering is unsupervised learning because it doesn't use predefined category labels associated with data items. Clustering algorithms are engineered to find structure in the current data, not to categorize future data. With the help of clustering, we can make the collection of similar data values. Dissimilar data values are another cluster. It is way finding similarities between data according to their characteristic. This technique based on the unsupervised learning (i.e. we do not know the expected output for the given input). For example, image processing, pattern recognition, city planning.
- f) **Summarization:** Summarization is abstraction of data. It is collection of appropriate task and provides an overview of data. For example, long distance race can be summarized total minutes, seconds and height. Association Rule: Association is the most popular data mining techniques and fined most frequent item set. Association strives to discover patterns in data which are based upon relationships between items in the same transaction. This method of data mining is utilized within the market based analysis in order to identify a set, or sets of products that consumers often purchase at the same time [6].
- g) **Sequence Discovery:** Uncovers relationships among data [8]. It is set of object each associated with its own time period of events. For example, natural disaster and analysis of DNA sequence.

DATA MINING APPLICATIONS

Various field adapted data mining technologies because of fast access of data and valuable information from a large set of data. We can perform Data mining techniques in different application areas that include marketing, telecommunication, scam/fraud detection, and finance sector, and education sector, medical sector and so on. Some of the generally used applications listed below:

- ❖ **Data Mining in Education Sector:** We are applying data mining in education sector then new emerging field called "Education Data Mining". Using these term enhances the performance of student, drop out student, student behavior, which subject selected in the course. Data mining in higher education is a recent research Use of Data Mining in Various Fields.
- ❖ **Data Mining in Banking and Finance:** Data mining has been used extensively in the banking and financial markets [11]. 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.
- ❖ **Data Mining in Market Basket Analysis:** These methodologies based on shopping database. The ultimate goal of market basket analysis is finding the products that customers frequently purchase together. The stores can use this information by putting these products in close proximity of each other and making them more visible and accessible for customers at the time of shopping [12].
- ❖ **Data Mining in Earthquake Prediction:** Predict the earthquake from the satellite maps. Earthquake is the sudden movement of the Earth's crust caused by the abrupt release of stress accumulated along a geologic fault in the interior. There are two basic categories of earthquake predictions: forecasts (months to years in advance) and short-term predictions (hours or days in advance) [13].
- ❖ **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 [4].
- ❖ **Data Mining in Telecommunication:** The telecommunications field implement data mining technology because of telecommunication industry have the large amounts of data and have a huge number of customers, and speedily varying and highly cutthroat environment. Telecommunication companies' employs data mining technique to improve their efforts in the field of marketing, fraud detection, and improved management of telecommunication networks [4].

- ❖ **Data Mining in Agriculture:** Data mining than emerging in agriculture field for crop yield analysis with respect to four parameters namely year, precipitation, construction and area of propagating. 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) [14].
- ❖ **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 [15]. Cloud computing uses the Internet services that rely on clouds of servers to handle tasks. The data mining technique in Cloud Computing to perform efficient, reliable and secure services for their users.

CONCLUSION

This paper provides a general idea of data mining, data techniques and data mining in various fields. The main objectives of data mining techniques are to discover the knowledge from active data. These applications use classification, Prediction, clustering, Association techniques and so on. Hopefully in future work we review various classifications and clustering algorithm and its significance's. If the conception of computer algorithms being based on the evolutionary of the organism is surprising, the extensiveness with which these methodologies are applied in so many areas is no less than astonishing. At present data mining is a new and important area of research and ANN itself is a very suitable for solving the problems of data mining because its characteristics of good robustness, self-organizing adaptive, parallel processing, distributed storage and high degree of fault tolerance. The commercial, educational and scientific applications are increasingly dependent on these methodologies.

REFERENCES

- [1] Xingquan Zhu, Ian Davidson, "Knowledge Discovery and Data Mining: Challenges and Realities", ISBN 978- 1-59904-252, Hershey, New York, 2007.
- [2] Joseph, Zernik, "Data Mining as a Civic Duty – Online Public Prisoners Registration Systems", International Journal on Social Media: Monitoring, Measurement, Mining, vol. - 1, no.-1, pp. 84-96, September 2010.
- [3] Zhao, Kaidi and Liu, Bing, Tirpark, Thomas M. and Weimin, Xiao,"A Visual Data Mining Framework for Convenient Identification of Useful Knowledge", ICDM '05 Proceedings of the Fifth IEEE International Conference on Data Mining, vol.-1, no.-1,pp.- 530- 537,Dec 2005.
- [4] R. Andrews, J. Diederich, A. B. Tickle," A survey and critique of techniques for extracting rules from trained artificial neural networks", Knowledge-Based Systems,vol.- 8,no.-6, pp.-378-389,1995.
- [5] Lior Rokach and Oded Maimon,"Data Mining with Decision Trees: Theory and Applications(Series in Machine Perception and Artificial Intelligence)", ISBN: 981-2771-719, World Scientific Publishing Company, , 2008.
- [6] Venkatadri.M and Lokanatha C. Reddy ,"A comparative study on decision tree classification algorithm in data mining" , International Journal Of Computer Applications In Engineering ,Technology And Sciences (IJCAETS), Vol.- 2 ,no.- 2 , pp. 24- 29 , Sept 2010.
- [7] Ankita Agarwal,"Secret Key Encryption algorithm using genetic algorithm", vol.-2, no.-4, ISSN: 2277 128X, IJARCSSE, pp. 57-61, April 2012.
- [8] Li Lin, Longbing Cao, Jiaqi Wang, Chengqi Zhang, "The Applications of Genetic Algorithms in Stock Market Data Mining Optimisation", Proceedings of Fifth International Conference on Data Mining, Text Mining and their Business Applications,pp- 593-604,sept 2005.
- [9] Fu Xiuju and Lipo Wang "Rule Extraction from an RBF Classifier Based on Class-Dependent Features ", ISNN'05 Proceedings of the Second international conference on Advances in Neural Networks ,vol.-1,pp.- 682-687,2005.
- [10]H. Johan, B. Bart and V. Jan, "Using Rule Extraction to Improve the Comprehensibility of Predictive Models". In Open Access publication from Katholieke Universiteit Leuven, pp.1-56, 2006
- [11]M. Craven and J. Shavlik, "Learning rules using ANN ", Proceeding of 10th International Conference on Machine Learning, pp.-73-80, July 1993. BIOGRAPHIES