

A Survey on Data Mining Methods and its Applications

R.Deepa¹, S.Vaishnavi²

Assistant Professor, Department of Information Technology,
PSG College of Arts and Science, Coimbatore, Tamil Nadu, India¹

Research Scholar, Department of Information Technology,
PSG College of Arts and Science, Coimbatore, Tamil Nadu, India²

ABSTRACT:- Data mining is the process used to convert raw data into useful data using software. It is also the process of selecting, searching, analyzing database for patterns and recent trends in big amount of data. It describes hidden patterns and relationships helpful in decision making. Data mining is approach for extracting of information from vast database to produce a new instruction that provides to define further tract in knowledge discovery in database. This paper explains the survey of data mining methods such as clustering, regression, classification and summarization. This paper also tells about applications and methods in data mining.

1. INTRODUCTION

Data mining technique that combines data analysis methods with advance algorithm for processing large amount of data. Data mining has the option for exploring and analyzing new type of data and old type of data in new way. Data mining analyses and summarizes the knowledge or data from different perspectives into effective information. Data mining is relating phase of knowledge discovery database (KDD). It abstracts from the hidden information similar and needful knowledge. Data mining is the technically way that finds patterns and correlation among number of field in huge database. This type of research that manipulates and stores decision making process and future prediction.

2. DATAMINING PROCESS AND STEPS

Data mining is the process in which provides a technique to find vital values and convert information into knowledge from hidden information we can abstract similar wanted knowledge. It identifies source information it takes data points and analyzes the needful and extracting related knowledge, that find the key values for extracted data and reporting the results. So in data mining is known KDD (knowledge Discovery in Database). The process of KDD is to significance the identify goals and understandable pattern in data. In Data Mining, knowledge extraction seven steps occurs in Fig 1.

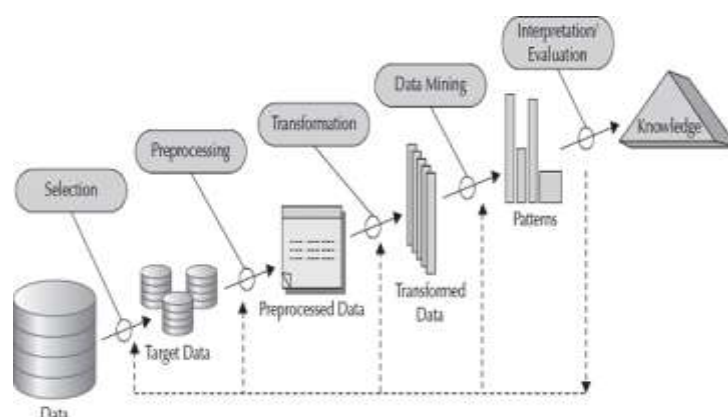


Figure 1 Data Mining Process

1. **Data cleaning:** This step is to remove noise data and inconsistent or irrelevant data from volume of data.
2. **Data integration:** First process is to collect the data, in this step we are integral from all the different sources. We combine n number of data sources into single data store called as target data.
3. **Data selection:** Data relevant is to analysis the task room the retrieved database from numerous resources.

4. **Data transformation:** The data is converted into required form of mining as we transform to appropriate for mining and techniques used to accomplish the functions are done such as smoothing, aggregation, normalization.
5. **Data mining:** This approach and tools used in this step that discovers data pattern and principles.
6. **Pattern evaluation:** This step, Patterns represents knowledge that are identified established on given estimate.
7. **Knowledge representation:** This is final step represents the knowledge technique which is used for the users helps to perceive and view the knowledge in data mining and outcome.

The aim of knowledge discovers and data mining process is to discover the patterns that are not known among the large set of data to interpret needful information and knowledge.

3. DATAMINING METHODS

Data mining process extracts the information for large volume of data sets and converts into predictable for future use. So it helps in achieving defined objectives. The data mining aims normally creates predictive and descriptive model [7]. A Predictive model turns worthy data giving sufficient information this uses data to determine headed future output of an event occurs it has many types of techniques of statistical modeling, machine learning, data mining and game theory that analysis current and historical facts to do predictive future. A Descriptive model concentrates data and past analysis events for instance how to future approach. Descriptive looks at historical performances by mining past data and looks for past success and failure. The Predictive and Descriptive models are represented fig 2.

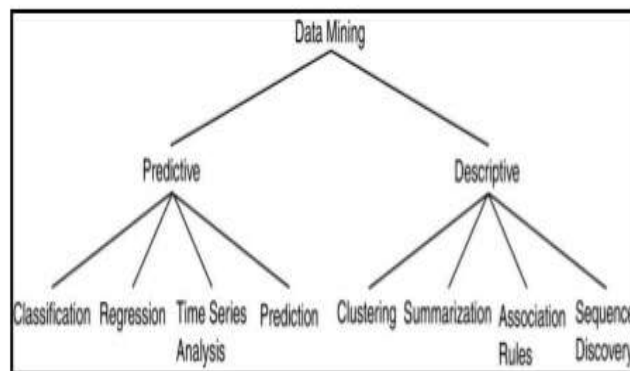


Fig 2. Predictive and Descriptive Models

1. **Classification:** Classification assigns new records to (predefined classes) advance establishment of classes, it uses to predictive the future outcome that is true or false describes range high or low. This is most familiar technique classification performs and assumes so data knowledge used to classify neural network, logistic regression, decision tree algorithm. Classification technique is use in segmenting of customer, business modeling and various applications. E.g. prediction of future employee records.
2. **Regression:** It is an another predictive data modeling type it known as supervised learning technique. This is analyses is dependent on another attributes mainly that presents in same item regression knows the target value. Regression models are computed by many statistical measures and the difference between predicted and expected values. E.g. prediction of house value and number of rooms, size etc.
3. **Time series data analysis:** Time series database is one of the popular data type it is used in many fields such as meteorological, business and medical. The recent change of this analysis is collecting in spot patterns. It can be analyzed to identify correlations. For E.g. Stock market that predict future values and defines similar patterns over time.
4. **Prediction:** Prediction calculates the rough value. The future is to predict a future than the current state. It defines the relationship between the dependent and independent variables. E.g sales and profit. It includes the fore warning of disasters occurs naturally(flooding, snowstorm).
5. **Clustering:** Clustering is the collection of similar group of objects in data. Another form of cluster is dissimilar of objects. Clustering is used in many applications mainly in research marketing, recognition of pattern and image

processing. Clustering application is to further discover marketing to help marketers and use to develop targeted marketing functions. We can find closed and scattered regions in object space. For effective means of classification approach is used of distinguishing groups. This method is expensive, so it can be used as preprocessing approach. Used in for E.g astronomy, galaxies, studies about earthquake detection.

6. **Summarization:** Data summarization summaries calculated data that includes earlier stage and obtained stage. In case to make denied evaluation data that is generalized in nature. The data in data warehouse is very high amount of space. Data summarization is large multi dimensional datasets as the case of data warehouse is very challenging work. Summarization simplifies methods of tabulating the mean and standard deviation which is applied in data analysis, variation and generating report.
7. **Association:** Association technique is used to find frequent patterns, correlates from datasets. Association rule is mostly used to express sales. Association rule mining is used to analyzing customer behavior and predict the customer behavior in supermarket. This rule given in a set of transaction that finds rule activate us to estimate the future occurrence of selected item based on the occurrence of other items in the current process of transaction. This plays the major role in shopping analysis of data, clustering of product and designs of catalog.
8. **Sequence Discovery:** it finds the relation between one or more data. The set of object that is connected with its own graphical representation of time period, which marks that important event. For E.g. scientific experimentation and natural disaster and analysis the sequence of data.

1. Tabular Column of Methods Used in Data Mining

Classification	Neural Network
	Nearest –Neighbor Classifier
	Decision Tree- oblique decision tree
Regression	Linear Regression
	Logistic Regression
	Polynomial Regression
	Stepwise Regression
	Ridge Regression
Time Series Data Analysis	Methods used in forecasting
	Exponential Smoothing
	Method V-Croston

Prediction	Artificial Intelligence
	Social Media Monitoring
	Customer Support and Chatbot
Clustering	Connectivity
	Centroid
	Distribution
	Density
Summarization	Cluster Based Method
	Graph Theoretic Approach
	Latent Semantic Analysis
	Automatic Text Summarization
Association	Rule Based Machine Learning
	Frequent Item set Generation
	Rule Generation
	Computational Complexity

Sequence Discovery	Mining Sequence Patterns
	Finding Sequence Patterns
	Constraint Based Sequence Pattern Mining

4. DATAMINING APPLICATIONS

Data mining application is used by many types of large and small scale companies and industries with stronger concentration in financial and retail communication and groups into access data which has lower end level of hierarchically structured database their transactional data and determine the rate of customer priority and arranging of product. In data mining a retailer can use point out with records of customer sales and purchase to develop and promote specifically to the customers.

1. **Healthcare and Medical:** Data mining handles great and popular to higher the improvement in health systems it uses the data analytics that identifies good practices that improves care and reduce cost. A researcher uses data mining to detect and survey different kind of diseases and takes future measurements.
2. **Education:** Education data mining that mines knowledge discovering from originating the data it helps in future outcomes. To computerize the syllabus which is needy for practical. It provides E-learning it fetches from database. Data mining accurate the result of student database. EDM learns to rise the technology to teach in different methods.
3. **Market Basket Analysis:** Data mining provides hidden patterns in business that helps in planning and releasing of marketing in cost efficient that allows the seller and buyer to understand the business the makes profitable change and customer satisfaction. It uses transaction of credit and debit card and hidden correlation.
4. **Finance Banking:** Banking uses to identify the customers to analyze the purchase, retain and stock trading rules. The data mining acquires in targeting and maintaining the profitable customers. It helps and improves in customer relationship management.
5. **Agriculture:** Agriculture in data mining emerges the technology in for crop yielding bases of occurrence of rainfall, production and seasonable changes. Data mining techniques such as k means and Artificial Neural Network.
6. **Cloud Computing:** Data mining that implements and utilizes programming, application of software managing of data storage .the huge volume of business data can be stored in information center in low estimation. It enables and performs that is efficient, reliable and secured and low cost efficiency for users.
7. **Transportation:** Data mining determines and distributes the schedule from warehouse and analyses the patterns.
8. **Engineering in Manufacturing:** Data mining tools is very useful in discovery patterns in manufacturing process. DM works in system level to design and extract the relationship of architecture product and product portfolio. It predicts the development of span and cost of the product.
9. **Lie Detection:** Lie detection is easy to bring the truth. Mining Technique used to detect investigate the crimes it includes text mining too. The sample data which collected from previous investigators that compares model of lie detection which is created. Law enforcement can use mining techniques for investigating of crime and monitoring suspected terrorists.
10. **E- Commerce:** E-commerce Company's uses data mining to check the range of selling customers that delivers number of products according to customer view. It used to sell and shows which preferred and update the trends. It helps to buy and sell faster and delivers on time. It checks for low operational and best quality.
11. **Bio Informatics:** this technology used to manage biological information bio informatics has the science of storing, extracting, analyzing, interpreting and utilizing the information of sequence and models. It is uses recent trend of biology genomics and research biomedical. This application includes discovery of structural patterns analysis of genetic networks.
12. **Financial Data Analysis:** This facility is systematic data analysis and mining in banking and industries in finance is reliable and systematic it designs and builds of data warehouse for multi dimensional data analysis and mining it helps customer in loan payment and credit policy.

5. CONCLUSION

Data mining is powerful and useful technique that manipulates data performance and gives exact and target output from large amount of data that grows world wide it helps the experts to validate, analysis of various application results. Data mining is purposely used for present and future domain. It provides high security, maintenance in business portfolio. This paper discusses about data mining ideas, KKD process and various techniques. We discussed some vision of data mining application.

REFERENCES

1. Aarti Sharma et al, "Application of Data Mining – A Survey Paper ",International Journal of Computer Science and Information technologies, Vol,5(2), 2014.
2. Smita, Priti and Sharma, "Use of Data Mining in Various Field: A Survey Paper" IOSR Journal of Computer Engineering,8727 Volume 16, Issue 3, Ver.V(May –June 2014)
3. Data Mining: Concepts and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management System's) 3rd Edition.
4. 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.
5. J. Han and M.Kamber, "Data Mining, Concepts and Techniques ", Morgan kauffmann, 2000.
6. Nikitha Jain, Vishal Srivastava "DATA MINING TECHNIQUES: A SURVEY PAPER" IJRET: International Journal of Research in Engineering and Technology, Volume:02 Issue: 11| Nov-2013.
7. Pradnya P.SondWale,"Overview of Predictive and Descriptive Data Mining Techniques" IJARCSSE, Volume 5, Issue 4, April 2015.
8. Prof.Dr.Wofgang Karl Hardle,"Time Series Data Mining Techiques" IJARCSSSE, Volume 5, Issue 4, April 2015.