International Journal of Advance Research in Science and Engineering
Vol. No.5, Issue No. 01, January 2016
www.ijarse.com
IJARSE
ISSN 2319 - 8354

A STUDY ON DATA MINING INVESTIGATING ITS METHODS, APPROACHES AND APPLICATIONS

Mrs. Jyoti Nawade[1], Dr. Balaji D[2], Mr. Pravin Nawade[3]

[1]Lecturer, JSPM'S Bhivrabai Sawant Polytechnic, Pune (India)

[2]Assistant Professor, [3]Assistant Administrative Officer, SIBM-H, Hyderabad (India)

ABSTRACT

Due to rapid computerization and digitalization in current decade's large amounts of data are now available in every sector of business and Industry. This data may provide us plenty resource of knowledge and help us in making correct decisions. Data Mining has adopted its techniques from various areas like difference Machine Learning Process, statistics, database system, Visualization, Neural Networks, and Rough sets etc. The various methods such as Artificial Neural Networks, Decision Trees, Genetic Algorithms, Nearest Neighbor Method and Rule induction are described. Business Applications, Database Marketing and Profitable Applications are explained in detail, with its various appendixes of data mining terms.

Keywords: Data Mining, Decisions, Database, Networks, Knowledge Discovery

I. INTRODUCTION

Due to rapid computerization and digitalization in current decade'slarge amounts of data are now available in every sector of business and Industry. This data may provide us plenty resource of knowledge and help us in making correct decisions. The primary social and economic value of modern societies is knowledge. For this reason, mastering Information technology for handling information, recorded in Datawarehouses or the Web, is essential to the development of individuals and society.When we shop at store where such system is already adopted, the cash counter person scans bar code of items and stores information of purchased items into a database system, this help to store keeper to analysis sales data for his record and in shopping transaction database. Data mining tools predict future trends and behaviors, allowing businesses to make proactive. Most of the companies already collect and refine massive quantities of data. Data mining techniques can be implemented rapidly on existing software and hardware platforms to enhance the value of existing information resources. To understand make use of database and meet the challenges, data mining is proposed as a multidisciplinary approach. Data Mining is core part of Knowledge Discovery in database. Knowledge Discovery Database consists of following steps, as shown in Fig 1.

Fig 1: STEPS IN KNOWLEDGE DISCOVERY DATABASE

II. TECHNIQUES OF DATA MINING

Data Mining has adopted its techniques from various areas like difference Machine Learning Process, statistics, database system, Visualization, Neural Networks, and Rough sets etc.

- **Machine Learning Process:** It is a process that search best model that suits to testing data. The searching space in Machine learning methods is a reasoning space of n attributes instead of vector space of n dimensions.

  The most common method used for the data mining includes decision tree induction. Decision tree is a classification of tree which determines an objects class by following path the root to leaf. Decision trees are induced from the training set and classification rules can be extracted from the decision trees. From the decision tree we can conclude, for example, a medium size vehicle will have medium mileage.

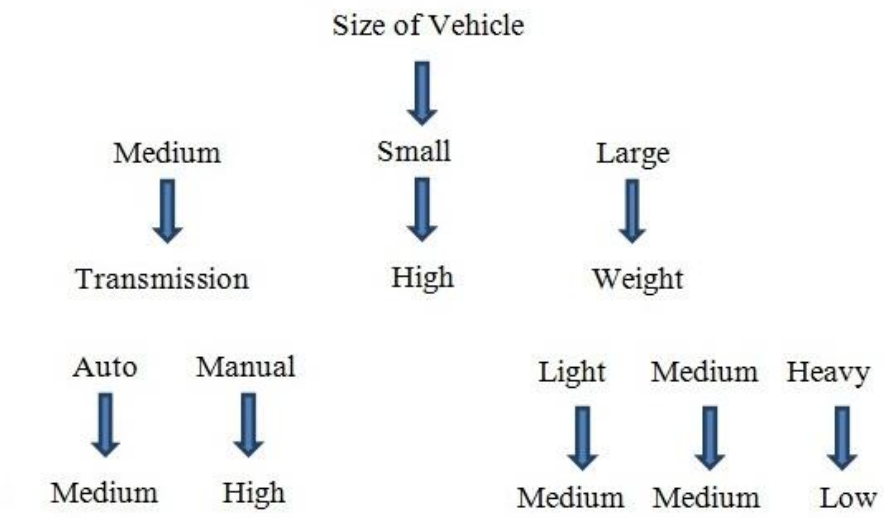

Fig 2: DECISION TREE - MEDIUM SIZE VEHICLE WILL HAVE MEDIUM MILEAGE

- **Database Approach:** Database heuristics are used to feat the characteristics of the data in hand. The attribute oriented induction, the iterative database scanning for common items sets. The attribute oriented inductions are primary in nature in this approach low level data are generalize into high level concepts using conceptual hierarchies. The attribute method focused on patterns with unusual possibilities by adding different attributes to partners. Refer following fig 3, which depicts conceptual hierarchy for students.

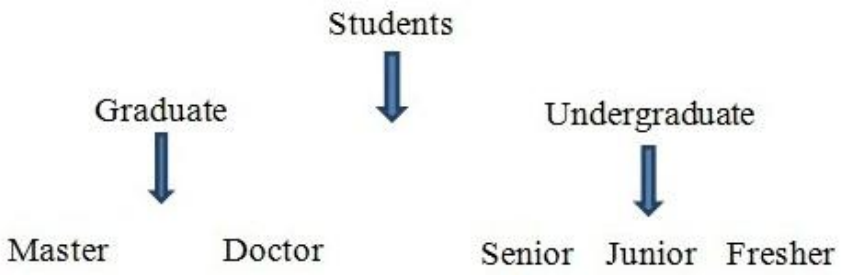

Fig 3: CONCEPTUAL HIERARCHY FOR STUDENTS

- **Visualization Approach:** Visualization is a interested set of approach in data mining, in this technic data transformed in visual substances such as different lines, dots and specific areas and display into various dimensional spaces. To understand the result of and further examination summarized data is usually presented as chart graphs etc.
- **Neural Network Approach:** Neural Network is a set of neurons. A neuron is computing device which function of inputs which can be outputs for other neurons. A neural network trained to model the relationship between a set of input attributes and output attributes by adjusting connection and the functional parameters.

- **Rough Sets Approach:** A rough sets are fuzzy in nature. Sets of object can be arranged to form group of rough sets which may be used for classification and examples.
- **Statistics Approach:** Built from a set of training data. In this kind of approach many statistical tools are used for data mining for example Correlation analysis, cluster analysis, Bayesian Networks, Regression analysis etc.

III. METHODS OF DATA MINING

The most commonly used Methods in data mining are:

- **Artificial Neural Networks**: It consist learning through training and resemble biological neural networks in structure.
- **Decision Trees**: It consist Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset.
- **Genetic Algorithms**: It consist genetic combination, mutation, and natural selection in a design based on the concepts of evolution.
- **Nearest Neighbor Method**: A technique that classifies each record in a dataset based on a combination of the classes of the record.
- **Rule induction**: It consist extraction of useful if-then rules from data based on statistical significance.

IV. DATA MINING APPLICATIONS

In current decades we can see so many applications are available related to Data Mining. Classification of database is as follows.

- **Business Applications:** In such kind of applications data mining is used in database marketing, retail data analysis, different kind of credit approvals and for stock selections.
  - Database Marketing: is a very popular business application of data mining. By preserving data in this application users can be used for effective marketing.
  - Retail Database: contains customer shopping history which can help in making sales campaign.
  - By effectively using data mining techniques user can build models which can be used for performance of stock.
- **Science Applications:** Data mining techniques currently using in various sectors like medical science, geology, astronomy and many more.
  - The use of Data mining techniques increased rapidly in various sectors such as banking sector to find out defaulters and money laundering related activities, by revenue department to find out tax fraud cases, health care department and many more sectors.
  - Data mining is a process of removing required decoration from large database. Data mining is an effective solution for big organization to find out problems rapidly for making corrections within the time limits.
- **Profitable Applications:** Due to heavy computerization in recent decades

- So many companies have deployed successful applications of data mining. The use of technology is applicable to any company looking to leverage a large data warehouse to better manage their customer relationships. Basically there are two factors are responsible for success with data mining are: well-integrated data warehouse and a well-defined understanding of the business process within which data mining is to be applied. Some successful Profitable applications include: A credit card company can influence its vast warehouse of customer transaction data to identify customers most likely to be interested in a new credit product. A Medicine company can analyze its sales and their results to improve to determine which marketing activities will have the greatest impact in the next few months. Each of these examples has a clear common ground. They leverage the knowledge about customers implicit in a data warehouse to reduce costs and improve the value of customer relationships.

V. VARIOUS APPENDIXES OF DATA MINING TERMS

- **CHAID:** Chi Square Automatic Interaction Detection. Provides a set of rules that can apply to a new dataset to forecast which records will have a given outcome.
- **Anomalous Data:** Result from errors or that represents unusual events.
- **Analytical model**: A structure and process for analyzing a dataset.
- **Artificial Neural Networks:** Non-linear predictive models that learn through training and resemble biological neural networks in structure.
- **Data classification:** The process of dividing a dataset into mutually exclusive groups such that the members of each group are as "close" as possible to one another, and different groups are as "far" as possible from one another, where distance is measured with respect to specific variable(s) you are trying to predict.
- **Clustering:** The process of dividing a dataset into mutually exclusive groups such that the members of each group are as "close" as possible to one another, and different groups are as "far" as possible from one another, where distance is measured with respect to all available variables.
- **Data Mining:** The extraction of hidden predictive information from large databases.
- **Data Navigation:** The process of viewing different dimensions, slices, and levels of detail of a multidimensional database.
- **Genetic Algorithms:** Optimization techniques that use process such as genetic combination, mutation, and natural selection in a design based on the concepts of natural evolution.
- **Outlier:** A data item whose value falls outside the bounds enclosing most of the other corresponding values in the sample. May indicate anomalous data.

VI. CONCLUSION

In this paper we briefly reviewed the various steps in data mining, investigating its methods, approaches and applications. This review would be helpful to researchers to focus on the various issues of data mining. The domain

experts play important role in the different stages of data mining. The decisions at different stages are influenced by the factors like domain and data details, aim of the data mining, and the context parameters. The domain specific applications are aimed to extract specific knowledge. The domain experts by considering the user's requirements and other context parameters guide the system. The results yields from the domain specific applications are more accurate and useful. Therefore it is conclude that the domain specific applications are more specific for data mining. From above study it seems very difficult to design and develop a data mining system, which can work dynamically for any domain.

REFERENCES

[1]. Balaji D, Sripathi K & B.R. Londhe, 'ECC Condition Enhances Organizational Excellence', International Journal of Advanced Technology in Engineering and Science, Vol. No.3, Special Issue. No. 01, ISSN: 2348 - 7550, September 2015, pp. 501-508.

[2]. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R... "CRISP-DM 1.0 : Step-by-step data mining guide, NCR Systems Engineering Copenhagen (USA and

[3]. Denmark), DaimlerChrysler AG (Germany), SPSS Inc. (USA) and OHRA Verzekeringenen Bank Group B.V (The Netherlands), 2000".

[4]. Dunham, M. H., Sridhar S., "Data Mining: Introductory and Advanced Topics", Pearson Education, New Delhi, ISBN: 81-7758-785-4, 1st Edition, 2006

[5]. Fayyad, U., Piatetsky-Shapiro, G., and Smyth P., "From Data Mining to Knowledge Discovery in Databases," AI Magazine, American Association for Artificial Intelligence, 1996.

[6]. Introduction to Data Mining and Knowledge Discovery, Third Edition ISBN: 1-892095-02-5, Two Crows Corporation, 10500 Falls Road, Potomac, MD 20854 (U.S.A.), 1999.

[7]. Jyoti Nawade, Balaji D, Pravin Nawade, 'Exploration of Cyber Security Awareness and its Serious Implications', International Journal of Advance Research in Science and Engineering, Vol. No.4, Special Issue. No. 01, ISSN: 2319-8354, November 2015, pp.40-48.

[8]. Larose, D. T., "Discovering Knowledge in Data: An Introduction to Data Mining", ISBN 0-471-66657-2, ohn Wiley & Sons, Inc, 2005.

[9]. Venkatesh. J and Balaji. D, 'Emotional Intelligence Enhances Unique Leadership', EXCEL International Journal of Multidisciplinary Management Studies, Vol. 1; Issue: 3; December 2011, ISSN: 2249-8834, A Double Blind Reviewed Refereed Multi-Disciplinary Journal, pp. 195-201.