Xournals

Academic Journal of Computer Sciences

ISSN UA | Volume 01 | Issue 01 | June-2018

Data Mining Techniques: A Study

Devendra Sharma[1] and Aman Sharma[2]

Available online at: www.xournals.com

Received 13th January 2018 | Revised 23th April 2018 | Accepted 24th May 2018

Abstract:

The process of data mining is the extraction of the useful data, patterns and trends from a data of large amount that uses techniques such as clustering, classification, association and regression. There are wide variety of application of data mining that uses various tools that supports different algorithms. This paper provides various data mining techniques and can also applied in the educational sector, marketing, detection of frauds, the telecommunication and manufacturing. As data mining is the notion of all methods and techniques that allows the analysis of large set of data to extract and discover unknown structures and their relations out of large details.

Keywords: *data mining techniques, Fraud detection,Clustering*

Authors:

1. *MB Khalsa College, Indore, Madhya Pradesh, INDIA*
2. *Maharaja Ranjit Singh College of Professional Sciences, Indore, Madhya Pradesh, INDIA..*

Introduction

Data mining simply refers to the extraction and mining of the knowledge from the large amount of data and it is also called as the "Knowledge mining from data" or called as "Knowledge mining". The technology for storing and collection of data has made this possible for the different organizations to accumulate large amounts of data at lower cost. This stored data is exploited for making the extraction useful and information should be actionable which is considered as the goal of the generic activity that is referred to as data mining.

Data mining is the exploration and analysis process, by means of automatic or semi-automatic one having large quantities of data that helps in discovering meaningful rules and patterns.

This discipline is the subfield of computer science that basically involves the computational process of data sets in large quantity and then transform it into a structure that should be understandable for further usage (Jain and Srivastava, 2013).

The Information Development Technology has produced databases in large amount in various areas or regions. The database and information technology research has given rise to an approach that stores and manipulates the previously existed data for the decision making process. This process of data mining is considered as the logical process which is further used to search by large amount of data that helps in the searching of the useful data. The main objective of this process is to find the patterns that are unknown previously. There are three steps involved,

- Exploration
- Identification of pattern
- Deployment

In Exploration step of data exploration, clean the data and then transform it into another form. After this determine the important variables and the nature of data.]

After the exploration of data, the second step is refining and defining the specific variables and to form pattern identification. For the purpose of best prediction, identify and choose the patterns.

Deployment in which the patterns are deployed for providing the desired outcome (Ramageri, 301).

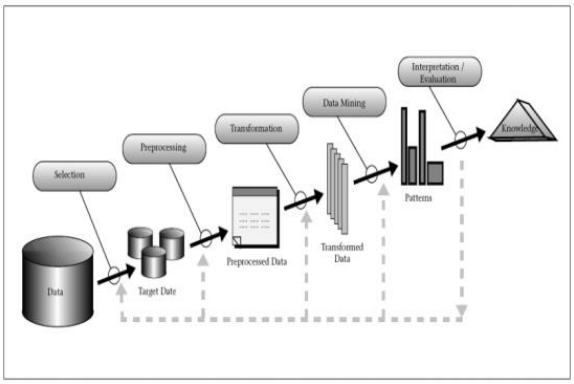

Figure 1 – Process of Knowledge Discovery

Functions of Data Mining

Fayyda et. al. in 1996 has explained the main functi0ons for the data mining

Classification process is the model findings that helps in the analysis and classification of data item into various classes that are predefined.

- Regression is data item mapping to a variable that is real valued prediction.
- Clustering is the finite set identification of the categories or clusters that describes the data.
- Dependency Modelling is a model finding that explains the major dependencies between the variables.
- Anamoly Detection is the discovery of the most important changes in the data.
- Summarization is the findings of the data subset along with a compact description.

There are two main goals of data mining that are need to be fulfilled which are prediction and description. Prediction involves the variables usage in sets of data in order to estimate the unknown values of other relevant variables whereas Description includes the findings of human patterns which are understandable and have data trends (Silwattananusarn and Tuamsuk, 2012).

Techniques of Data Mining

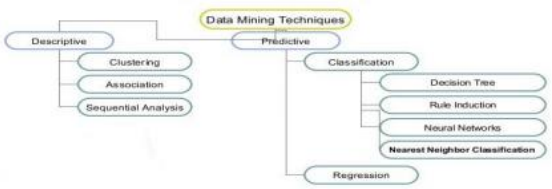

Figure 2- Descriptive and predictive data mining techniques

Classification approach

It is a supervised method of learning and the classification of data is a two-step process in which the first step is building a model by data analysis from the training data which have a set of characteristics or variables. On the data of data training, classification algorithm is applied to create the model whereas in the second classification step, the data for test is used to authenticate or validate the model accuracy. The use of the classification techniques is to classify the records of data into one among a set of predefined class. They work by a model construction for training dataset that consists of records having known class labels.

Clustering Approach

This approach includes the object's group findings such that the object present in one group will be considered as similar to one another and to the different from the objects in another group. This can be considered as the unsupervised and supervised learning techniques. In this, the structure is found in a collection of data that is unlabeled. The analysis of clustering has been used largely in many kinds of applications such as image pattern of business intelligence recognition web search biology and security. To organize large number of customers into groups where a customer present in a group share some of the similar characteristics, business intelligence clustering is used. This process facilitates the strategic development of business for the management of customer relationship (Dogra 2015).

According to Ramagari, 301, there are five types of clustering methods,

- Partitioning Methods
- Hierarchical Agglomerative methods
- Density based methods
- Grid based methods
- Model based methods

Predication

The technique of Regression can also be used for Predication and then to model the relationship that exists in between the one or more variables which are independe3nt and dependent. Independent variables are those attributes which known or existed already and response variables what the researcher wants to predict. So various complex techniques such as logistic regression, decision tree may be considered as necessary to predict the future values. Sometimes the same types of model can used for both regression and classification.

Regression model types

- Linear Regression
- Multivariate Linear Regression
- Nonlinear Regression
- Multivariate Nonlinear Regression

Association Rules

This rule is used in businesses to make certain kinds of decisions such as design of catalogue , cross marketing and behavior analysis of customer shopping. This rule of algorithm –produces confidence rules having values less than one.

Types of Association rules

- Multilevel association rule
- Multidimensional association rule
- Quantitative association rule (Ramageri , 301; Mukherjee et al , 2015)

Neural Networks

This is the set of input or output connected units and during the phase of learning, the learning of networks is occurs by the weight adjustment that helps in predicting the correct class labels of the input tuples.

Types of neural Networks

- Back Propagation

Data mining which is a new technology has wide applications such as in FBTO Dutch Insurance Company, provident financials home credit division and standard life mutual financial services.

Literature Review

Hilage and Kulkarni 2011, examined the results after applying the rule of association in minimg technique, rule induction technique and Apriori algorithm. They have applied the technique to the database of shopping mall. The analysis of Market Basket is performed by the data mining techniques and buying behavior is determined.

Gulati and Sharma 2012, discussed about the educational mining of data and how it is utilized to improve the functional activities related to education business by students, teachers and the manner in which class are arranged. The educational data mining helps teachers, students and managements to organize the classes for teaching and their schedule to improve the performance of students.

Silwattananusarn and Tuamsuk 2012, has reviewed the applications of data mining for knowledge management from 2007 to 2012.This paper have discussed about the four topics which are knowledge types or knowledge datasets, knowledge resources, tasks of data mining and data mining techniques and applications.

Jain and Srivastava 2013, discusses the concept of data mining in a summarized way and its importance toward its methodologies. The data mining approaches that is based on the Neural Network and Genetic Algorithm is researched in detail and the major technologies , the ways through which the data mining on Neural Network and Genetic Algorithm are also surveyed.

Smita and Sharma 2014, surveyed about the use of data mining in various fields. These data mining techniques includes the association, correlation, clustering and neural network. This paper also conducts a data mining applications review such as in the education sector, marketing, and detection of frauds, telecommunication and manufacturing.

Dogra and Wala 2015, reviewed the data mining techniques and algorithms. In this paper, data mining is described as the notion of all the methods and techniques that allows the analysis of large data sets to extract and discover the previous set unknown standards. The classification and clustering techniques are studied on the basis of algorithm process that is used to estimate previously unknown objects.

Gera and Goel 2015, illustrated the techniques of data mining, their methods and algorithms. This paper also discusses about the tools that are available and the supporting algorithms. Various tools has been compared to enable the user to use various tools on the basis of their requirements and applications. They have summarized the different validation indices for the proper validation.

According to the **A, Lakshmi and Srinivasa 2017**, the data analysis methods are used to pattern analysis of data in various fields and on the basis of analysis, suggestions are provided to decision making authorities. In this paper, the authors carried out the research by applying a mathematical model and a tool for the determination of effective course conduction strategy. The accuracy that is predicted in this research is more as it is measured by Rsquare value which is near to one.

Ramageri, explained all the techniques of data mining and their applications and this paper also discusses about the algorithms used in the data mining. Some of the organizations have adapted technology of data mining to improve their businesses and to provide better results.

Conclusion

The importance of data mining is basically among the pattern determination, knowledge discovery and in different business areas. All the algorithms and techniques of data mining such as classification, clustering , predicament, association and neural networks helps in finding the patterns and to decide the future trends in business to grow. This approach has wide application in every industry where the data is being generated because of which data mining is considered one of the main frontiers in database and information systems.

References:

A, Parkavi, et al. “Predicting Effective Course Conduction Strategy Using Datamining Techniques.” Academic Journals, vol. 12, no. 24, Dec. 2017, pp. 1188–1198.

Dogra, Ashish Kumar, and Tanuj Wala. “A Review Paper on Data Mining Techniques and Algorithms.” International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), vol. 4, no. 5, May 2015, pp. 1976–1979.

Gera, Mansi, and Shivani Goel. “Data Mining - Techniques, Methods and Algorithms: A Review on Tools and Their Validity.” International Journal of Computer Applications, vol. 113, ser. 18, Mar. 2015, pp. 22–29. 18.

Gulati, Pooja, and Archana Sharma. “Educational Data Mining for Improving Educational Quality.” IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS, vol. 2, ser. 3, June 2012, pp. 648–650. 3.

Hilage, Tejaswini Abhijit, and R V Kulkarni. “Application of Data Mining Techniques to a Selected Business Organization with Special Reference to Buying Behavior.” International Journal of Database Management Systems (IJDMS), vol. 3, ser. 4, Nov. 2011, pp. 169–181. 4.

Jain, Nikita, and Vishal Srivastava. “DATA MINING TECHNIQUES: A SURVEY PAPER.” IJRET: International Journal of Research in Engineering and Technology, vol. 2, no. 11, Nov. 2013, pp. 116–119.

Mohata, Pranit B, and Sheetal Dhande. “Web Data Mining Techniques and Implementation for Handling Big Data.” International Journal of Computer Science and Mobile Computing, vol. 4, no. 4, Apr. 2015, pp. 330–334.

Ramageri, Bharati M. “DATA MINING TECHNIQUES AND APPLICATIONS.” Indian Journal of Computer Science and Engineering, vol. 1, no. 4, pp. 301–305.

Silwattananusarn, Tipawan, and Kulthida Tuamsuk. “Data Mining and Its Applications for Knowledge Management: A Literature Review from 2007 to 2012.” International Journal of Data Mining & Knowledge Management Process, vol. 2, no. 5, 2012, pp. 13–24., doi:10.5121/ijdkp.2012.2502.

Smita, and Priti Sharma. “Use of Data Mining in Various Field: A Survey Paper.” IOSR Journal of Computer Engineering (IOSR-JCE), vol. 16, no. 3, 2014, pp. 18–21.