A Review On:Data Mining Tools and Application

Nikhil Bhongade[1]
M.E. Scholar, Dept. of CS&E
Prof. Ram Meghe Institute of Technology & Research
Badnera, India

Shrikant Akarte[2]
Assistant Professor, Dept. of CS&E
Prof. Ram Meghe Institute of Technology & Research
Badnera, India

Gaurav Sawale[3]
Assistant Professor, Dept. of CS&E
Prof. Ram Meghe Institute of Technology & Research
Badnera, India

Abstract: Data Ming is the process of extracting knowledge, patterns and trends from large amount of data by using various data mining technique like clustering, classification association and regression. In this paper we discuss the data mining technique and various application of data mining.

Key words: Data mining Techniques, Data mining applications.

I.   INTRODUCTION

Data mining is the process of extracting useful information. Basically it is the process of discovering hidden patterns and information from the existing data. In data mining, one needs to primarily concentrate on cleansing the data so as to make it feasible for further processing. The process of cleansing the data is also called as noise elimination or noise reduction or feature elimination [1].Data mining is also called knowledge discovery from data, knowledge mining, and data/pattern analysis. Data mining is process where find the useful information from previously unknown.Onece this information is find that is used is decision making in various business application.

II.   KNOWLEDGE DISCOVERY IN DATABASES PROCESS

Knowledge Discovery in Databases Process (KDD) is process to extract the knowledge or useful information from lower databases. This widely used data mining technique is a process that includes data preparation and selection, data cleansing, incorporating prior knowledge on data sets and interpreting accurate solutions from the observed results. Steps in the KDD process are shown in the following figure.
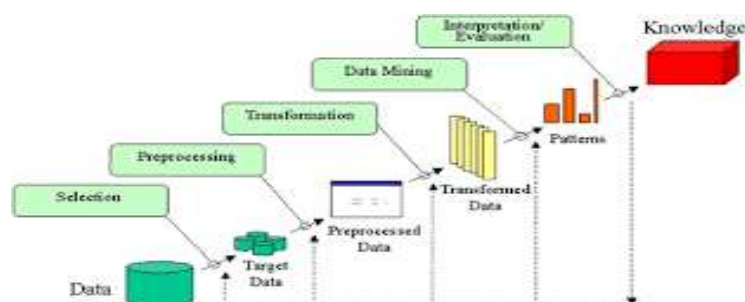

Fig. 1. Knowledge Discovery in Databases Process

In the above figure:

- Data come from heterogeneous resources and combine in single unit data called target data.
- After that, Data preprocessed and transformed into standard format.
- The output of the data in the form of patterns or rule by applying the data mining algorithm.
- Then those patterns and rules are interpreted to new or useful knowledge or information.

International Journal of Science Technology Management and Research
Volume 2, Issue 5, May 2017
www.ijstmr.com

The goal of the KDD and Data Mining is to find out the patterns that are hidden among the large datasets of data and called them knowledge or patterns.

Major KDD application areas include marketing, fraud detection, telecommunication and manufacturing.

III.   DATA MINING TECHNIQUES

There are number of data mining technique are developed and used in the various application. Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., are used for knowledge discovery from databases.

➢   **Classification**

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. Fraud detection and credit risk applications are particularly well suited to this type of analysis. This approach frequently employs decision tree or neural network-based classification algorithms. The data classification process involves learning and classification. In Learning the training data are analyzed by classification algorithm. In classification test data are used to estimate the accuracy of the classification rules. If the accuracy is acceptable the rules can be applied to the new data tuples. For a fraud detection application, this would include complete records of both fraudulent and valid activities determined on a record-by-record basis. The classifier-training algorithm uses these pre-classified examples to determine the set of parameters required for proper discrimination. The algorithm then encodes these parameters into a model called a classifier [2].

➢   **Clustering**

Clustering is "the process of organizing objects into groups whose members are similar in properties".By using clustering techniques we can further identify dense and sparse regions in object space and can discover overall distribution pattern and correlations among data attributes. For example, to form group of customers based on purchasing patterns, to categories genes with similar functionality.

Types of clustering methods
· Partitioning Methods
· Hierarchical Agglomerative (divisive) methods
· Density based methods
· Grid-based methods
· Model-based methods

➢   **Regression**

Regression analysis can be used to model the relationship between one or more independent variables and dependent variables. In data mining independent variables are attributes already known and response variables are what we want to predict. Unfortunately, many real-world problems are not simply prediction.

Types of regression methods
· Linear Regression
· Multivariate Linear Regression
· Nonlinear Regression
· Multivariate Nonlinear Regression

➢   **Association rule**

Association and correlation is usually to find frequent item set findings among large data sets. This type of finding helps businesses to make certain decisions, such as catalogue design, cross marketing and customer shopping behavior analysis. Association Rule algorithms need to be able to generate rules with confidence values less than one. However the number of possible Association Rules for a given dataset is generally very large and a high proportion of the rules are usually of little (if any) value [2].

Types of association rule

· Multilevel association rule

· Multidimensional association rule

· Quantitative association rule

IV.   APPLICATIONS IN DATA MINING

There are large scopes for application of data mining in different types of area as follows [3]:

1) In Medical Science:
In medical science there is large scope for application of data mining. Diagnosis of diesis, health care, patient profiling and history generation etc. are the few examples. Mammography is the method used in breast cancer detection. Radiologists face lot of difficulties in detection of tumors that's why CAM(Computer Aided Methods) could help to the medical staff .

2) In the Web Education:
In the 21st century the beginners are using the data mining techniques which is one of the best learning method in this era. This makes it possible to increase the awareness of learners.Web Education which will rapidly growth in the application of data mining methods to educational chats which is both feasible and can be improvement in learning environments in the 21st century .

3) A malicious Executable is Threat:
A malicious executable is threat to system's security, it damage a system or obtaining sensitive information without the user's permission. The data mining methods used to accurately detect malicious executable before they run.

4) Sports data Mining:
The data mining and its technique is used for an application of Sports center. Data mining is not only use in the business purposes but also it used in the sports .In the world, a huge number of games are available where each and every day the national and international games are to be scheduled, where a huge number of data's are to be maintained .

V. TRENDS IN DATA MINING

Table 1: Data Mining Trends Comparative Statements [5]

Data Mining Trends	Algorithms/Techniques Employed	Data Formats	Computing Resources
Past	Statistical, Machine Learning Techniques	Numerical data and structured data stored inTraditional databases	Evolution of 4G PL and Various related techniques
Current	Statistical, Machine Learning, Artificial Intelligence, Pattern Reorganization Techniques	Heterogeneous data formats includes structured, semi structured and unstructured data	High speed networks, High end storage devices and Parallel, Distributed computing etc…
Future	Soft Computing techniques like Fuzzy logic, Neural Networks and Genetic Programming	Complex data objects includes high dimensional, high speed data streams, sequence, noise in the time series, graph, Multi instance objects.	Multi-agent technologies and Cloud Computing

International Journal of Science Technology Management and Research
Volume 2, Issue 5, May 2017
www.ijstmr.com

CONCLUSION

Data mining has very importance in the finding the patterns and discovery of knowledge from large databases. In this paper we summarize the differ technology of data mining i.e., Classification, Clustering, Regression etc., By using this technique organisation can take decision for future trade of businesses. Data mining has wide application field almost in every industry where the data is generated that's why data mining is considered one of the most important frontiers in database and information systems and one of the most promising interdisciplinary developments in Information Technology also.

REFERENCE

1.   Phridvi Raj MSB., Guru Rao CV (2013) Data mining – past, present and future – a typical survey on data streams. INTER-ENG Procedia Technology 12:255 – 263

2.   Bharati M. Ramageri "Data Mining Techniques and Applications," Indian Journal of Computer Science and Engineering,Vol. 1 No. 4 301-305.

3.   Robert P. Schumaker ,Osama K. Solieman ,Hsinchun Chen ,Springer. Content Technology and its Applications Volume 4, Number 9, December 2010.

4.   Anmol Kumar et al." data Mining: Various Issues and Challenges for Future A Short discussion on Data Mining issues for future work" International Journal of Emerging Technology and Advanced Engineering, (ISSN 2250-2459(Online), Volume 4, Special Issue 1, February 2014)

5.   Sangeeta Goele, Nisha Chanana, "Data Mining Trend In Past, Current And Future," International Journal of Computing & Business Research, in Proc. I-Society 2012, 2012.

6.   PhridviRaj MSB., GuruRao CV (2013) Data mining – past, present and future – a typical survey on data streams. INTER-ENG Procedia Technology 12:255 – 263.

7.   Demšar J, Zupan B (2013) Orange: Data Mining Fruitful and Fun - A Historical Perspective. Informatica 37:55–60

8.   Jain AK, Murty MN, Flynn PJ (1999) Data Clustering: A Review. ACM Computing Surveys, 31:264-323

9.   Baker RID, Yacef K (2009) The State of Educational Data Mining:A Review and Future Visions. JEDM - Journal of Educational Data Mining, 1: 3-16

10.   http://www.slideshare.net/Annie05/sequential-pattern-discovery-presentation

11.   http://dataminingtools.net/wiki/introduction_to_data_mining.php

12.   http://www.dataminingtechniques.net

13.   http://www.slideshare.net/huongcokho/data-mining-concepts