A Survey of Data Mining Applications and Techniques

Samiddha Mukherjee[1], Ravi Shaw[2], Nilanjan Haldar[3], Satyasaran Changdar[4]

[1,2,3,4] *Department of Information Technology, Institute of Engineering & Management, Kolkata, India*

Abstract--- In layman terms Data-mining can be related to human cognitive mind where based on previous knowledge and experience we can relate things happening around us or sometimes even predict the future. Data mining is a process of searching data from a pool of data like database, web-servers, cloud based servers etc. and provide a pattern or relationships among those data to produce desired information. This paper conducts a formal review of the concept of data-mining, the standard tasks involve in data-mining, its applications in day to day field, techniques and methodology.

Keywords--- KDD, Data-mining techniques, Predictive model, Descriptive model

I. INTRODUCTION

With the advent of the 21st century, generation of data is exponentially increasing. During the intervening decades, important innovations in computer systems have led to the introduction of new technologies [12], for well based education. Data mining allows a search for, valuable information, in large volumes of data [13] Data is ubiquitous, generating from social networking sites (e.g. Facebook, Twitter and Instagram), e-commerce websites, cloud services, data obtained from smartphones and sensors etc.. The explosive growth in databases has created a need to develop technologies that use information and knowledge intelligently. Therefore, DMT (Data-mining techniques) has become an increasingly important research area[14]. The relevance of this huge collection of data is highly subjective as according to the individual business interests.

Data Mining is a relatively new term in the field of informatics. Data mining is the technique of filtering relevant data according to one's business interests from the massive collection of data using different techniques and algorithms such as Association, Clustering and Classification etc. In data mining, comes a term, 'Knowledge Discovery in Database' or KDD which encompasses the collection, classification and relevant evaluation of data. KDD is an iterative process consisting of the following sequential steps as listed below:

1. *Data Integration and Cleaning:* It is the process of collecting data from multiple resources and hence filtering out irrelevant data according to the objective of the business enterprise.
2. *Data Selection:* In this we select the relevant data. This becomes the target data which is to be processed.
3. *Data Transformation:* In this step, the data is accumulated and put into standard formats for mining by methods such as association and clustering.
4. *Data Mining:* This is the most crucial step in which one of the several data mining techniques is selected as per

desired outputs. For example, if one is to figure out which group a particular data point belongs to, the method of clustering is used.
5. *Pattern Evaluation:* Pattern evaluation is the process in which a pattern is identified.
6. *Knowledge Representation:* This is the final stage in which the knowledge is represented using different visualization techniques. [1]
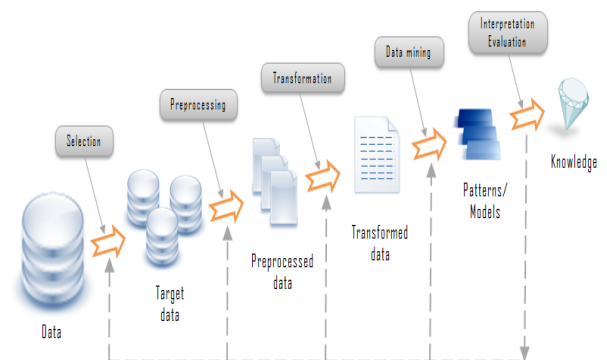

Fig-1[15]

Further in this paper we have discuss the various applications of Data-mining in different fields. The most useful and popular DMT(Data-mining Techniques) have been discussed in detail in this paper for understanding, which is followed by conclusion.

II. DATA-MINING APPLICATIONS

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it With the outgrowth of different kinds of organizations spanning over several continents the type of data is diversifying. The diversification of such data has led to the application of data mining in the following sectors.

1. *Business Sector:* In business world data mining is basically used for analyzing performance, profitability index, and customer feedback evaluation, and analysis of the stock values of existing organizations and their market trends to aid in future business decisions.
2. *Marketing and retailing sector:* Data mining provide accurate information regarding customer purchase trends, top selling products, so that the retail-store managers are able to identify their loyal customers, and providing discounts and arranging shelves according to customer requirements. [2]
3. *Bio-informatics:* Accumulation of medical records of the patient to develop a relationship between the disease

and the effectiveness of treatment, assessment of genomic and proteomic data in bio medical field. [3]

4. *Climatology:* Assessment of weather conditions over a period of time so as to predict future meteorological patterns for determining natural calamities like cyclone and also weather forecasting.

5. *Banking & Finance:* Assessment of individual banking records to generate different marketing strategies for a target customer segment, loan approval, stock forecasting, checking different kinds of fraud and money laundering. [4]

6. *Security and data integrity:* Data mining can be used to monitor different systems and raises alarm whenever any kind of security breach or intrusion is detected. It can help in identifying the reason for security problems in firewall.

7. *E-Commerce:* Data mining techniques are used in ecommerce to analyze customer search patterns to promote up sale and cross sale.

8. *Forensic & Criminal Investigation*: Data mining technique is used in forensic and criminal department to assess previous criminal records in order to identify the criminal as well as to determine the crime pattern ,sentiments of the accused and the accuse. [5]

9. *Government records:* Data mining technique is used in government record. For generating citizen specific data which can include anything starts form employment record till medical history, law enforcement, profiling tax cheaters.

10. *Cloud Computing:* Today cloud computing can be considered as one of the major source of every kind of data. Moreover cloud servers are fast, reliable, efficient and secure and reduce the cost of infrastructure of individual.[5] So using of KDD techniques and different kinds of data mining algorithms, one can even create different search patterns and applications for finding any information which remains hidden in unstructured data.

III. DATA MINING TECHNIQUES

Data Mining can be fundamentally divided into two broad groups: Predictive and Description.

1. **Predictive**: In this model, values for some existing variables in the data sets are known beforehand and using this knowledge the future values are predicted. The following are the popular predictive techniques used:

Fig-2[16]

1.1 *Classification:* In a database, the data which shows similar behavior or properties are categorized under some predefined classes. Example: In a library, the books are categorized according to the subjects. [6]

1.2 *Regression:* Using this technique the data items are mapped to the real valued prediction [7]. Regression is usually plotted on a graph to analyze the future trends based on the past and present data.
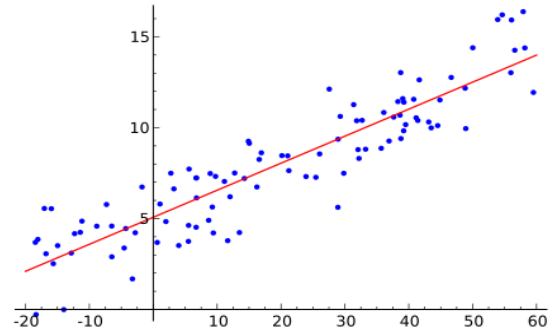

Fig-3 Plotting of Graph using regression [25]

Example: Using raw scientific data related to astronomical bodies the probability of collision is estimated and their trajectory is determined.

1.3 *Time-series Analysis:* Using statistical techniques, plotting of time dependent data points so as to assess future trend is Time series-Analysis.
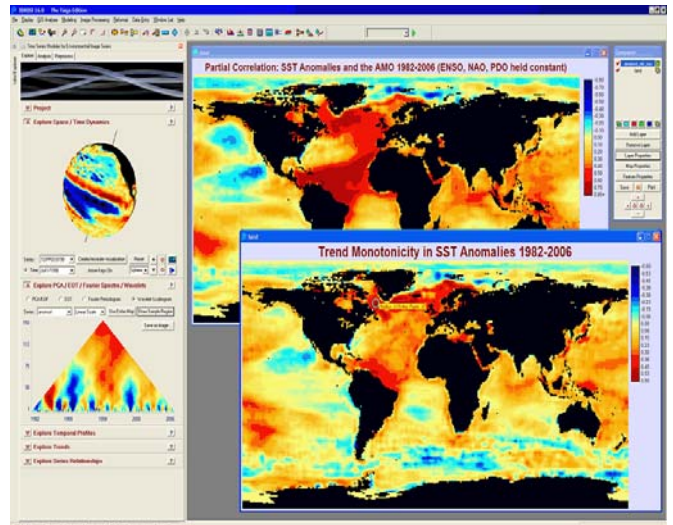

Fig-4[17]

Example: Analysis of weather conditions over a period of time helps predicting the future weather patterns and determining the arrival of weather-related natural calamities.

1.4 *Prediction:* Using this technique the relationship between dependent and independent variables are examined and future values are predicted [8]. The values plotted in the prediction model may be continuous or ordered.

1.5 *Artificial Neural Network:* This usually consists of sophisticated computer programs used to detect patterns and to build predictive models from the large set of databases. Neural networks grew out of the field of Artificial Intelligence and went on to include several statistical models.
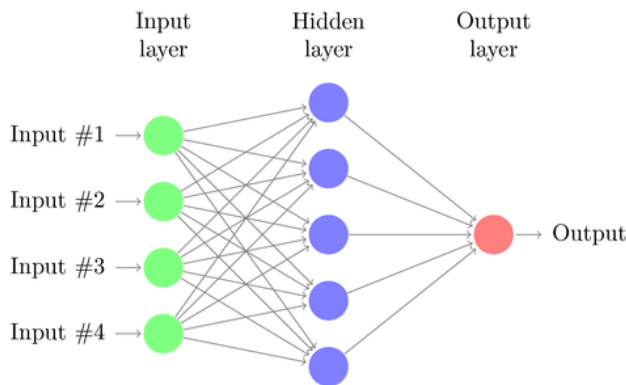
Fig-5 Neural Network [18]

1.6. *Decision Trees:* Also a predictive model, these are tree-like structures where each node denotes a test on an attribute value, each branch represents an outcome of the test and tree leaves represent classes or class distribution.
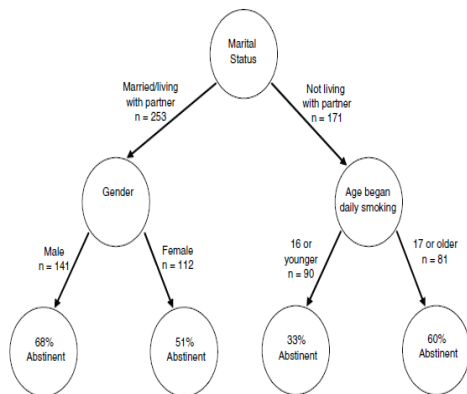

FIGURE 3. GUIDE decision tree predicting abstinence at the end of treatment excluding individuals who relapsed in the first week (N = 424; average abstinent rate = 54.3%).

Fig-6 Decision Tree [24]

2. Descriptive: In descriptive models, patterns are found giving the description of the data and subsequent presentation for user-interpretation. The following are the widely popular descriptive techniques:

2.1 *Clustering:* This technique is used to group data depending upon their similar characteristic resulting in the formation of clusters. The data points are so clustered that the intra-class similarities are maximized and inter-class similarities are minimized based on the data points attributes.
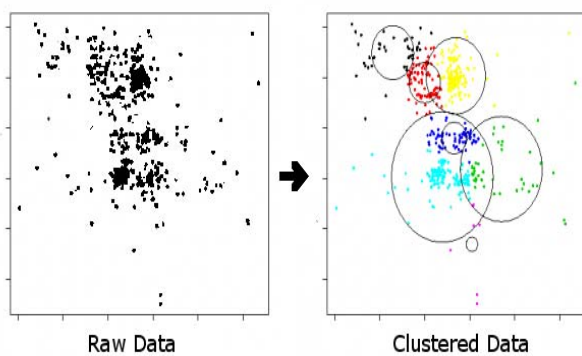

Fig-7 Clustering Process [19]

2.2 *Association:* Using association rules, one identifies the relationship between two or more items to find patterns. Association rule can be useful for retail management, marketing like cross-sell in e-commerce. The most used association rules include the Apriori algorithm and the fp-tree.
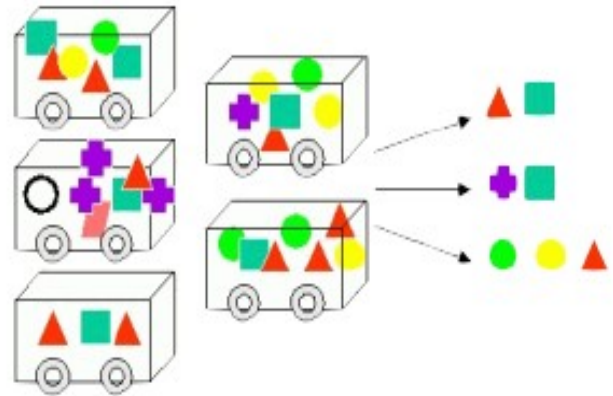

Fig-8 Association [20]

2.3 *Summarization:* This technique is used to form an abstraction of data. A set of relevant tasks is grouped and an overview is established. [10]

2.4 *Sequence- discovery:* The relationship amongst data is established and the events are organized sequentially as associated with the data. [11]

IV. CONCLUSION

In this paper we have discussed the concepts of kdd, data mining, and relevance of data mining in various sectors. We have also covered the different techniques of data mining – predictive model (classification, regression, time-series analysis, prediction, neural network and decision trees) and descriptive model (clustering, association, summarization, and sequence discovery). The commercial, educational, medical, scientific applications of data mining are increasing day by day and thus conceptualizing and developing new techniques to mine data is becoming an essential requirement of the 21st century information age.

REFERENCES

[1] Aarti Sharma et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 2023-2025 ISSN:0975-9646

[2] Delmater R and Hancock M, Data Mining explained-a manager's guide to customer-centric business intelligence (Digital Press, Boston) 2002.

[3] Aakanksha Bhatnagar, Shweta P. Jadye, Madan Mohan Nagar" Data Mining Techniques & Distinct Applications: A Literature Review" International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 9, November- 2012

[4] Industry Application of data mining, http://www.pearsonhighered.com/samplechapter/0130862711.pdf

[5] Usefulness and applications of data-mining in extracting information from different perspective by Jiban K Pal. (Annals of Library and Information Studies , Vol-58, March 2011, pp. 7-16)

[6] Data Mining: Tasks, Techniques and Appliactions by Yongjian Fu

[7] Ruxandra-Ștefania PETRE, "Data mining in Cloud Computing" Database Systems Journal vol. III, no. 3/2012

[8] Dr. M.H.Dunham, "Data Mining, Introductory and Advanced Topics", Prentice Hall, 2002.

[9] Nikita Jain, Vishal Srivastava "DATA MINING TECHNIQUES: A SURVEY PAPER" IJRET: International Journal of Research in Engineering and Technology, Volume: 02 Issue: 11 | Nov-2013**,**

[10] Data mining white paper, www.ikanow.com

[11] Dr. M.H.Dunham, "Data Mining, Introductory and Advanced Topics", Prentice Hall, 2002

[12] Ha, S., Bae, S., & Park, S. (2000). Web mining for distance education. In IEEE international conference on management of innovation and technology (pp. 715–719).

[13] Weiss, S. H., & Indurkhya, N. (1998). Predictive Data Mining: A Practical Guide. San Francisco, CA: Morgan Kaufmann Publishers.

[14] Fayyad, U., Djorgovski, S. G., & Weir, N. (1996). Automating the analysis and cataloging of sky surveys. In U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, & R. Uthurusamy (Eds.), Advances in Knowledge Discovery and Data Mining (pp. 471–494). Cambridge, MA: MIT Press

[15] http://tezywap.eteagle.co.uk/from-data-mining-to-knoledge-discover/

[16] http://www.dataminingcrm.com/

[17] http://www.clarklabs.org/products/Earth-Trends-Modeler.cfm

[18] http://stackoverflow.com/questions/27986677/detecting-users-emotions-by-analyzing-the-text

[19] http://stackoverflow.com/questions/1441319/whats-the-best-approach-to-recognize-patterns-in-data-and-whats-the-best-way

[20] http://gerardnico.com/wiki/_media/data_mining/data_mining_association.jpg

[21] http://www.oracle.com/technetwork/database/options/advanced-analytics/odm/odm-techniques-algorithms-097163.html

[22] http://www.thearling.com/text/dmtechniques/dmtechniques.htm

[23] http://www.zentut.com/data-mining/data-mining-techniques/

[24] http://4.bp.blogspot.com/-HgeUaINSvH4/T2lIdAXnOuI/AAAAAAAAAEY/0V-ezRRw3KQ/s1600/Figure+2+Decision+Tree.png

[25] http://datamining.xmu.edu.cn/main/~cloud/images/Linear_regression.png