A Study on Techniques of Data Mining and its Applications

[1]Mr P.Boopathi [2]Mr L.Ramesh
[1]Assistant Professor [2]Research Scholar
[1]Department of Computer Applications
[1]STC,Pollachi,Coimbatore,India
[2]PG & Research Department of Computer Science
[2]Government Arts College,Coimbatore,India

Abstract: Data mining is the way toward breaking down information from various perspectives and abridgi ng it into helpful information. "Data mining, likewise prominently alluded to as learning revelation from information (KDD), is the computerized or advantageous extraction of examples speaking to information verifiable put away or caught in substantial databases, information stockrooms, the Web, other gigantic data vaults or information streams.". This paper gives a review on different information mining procedures, for example, classification, clustering, regression, summarization and so on.. This paper too talks about a portion of the Data mining applications

Index Terms- knowledge discovery in data, methodologies of data mining, data mining application,

I. INTRODUCTION

Improvement of Information technology has created vast measure of data-base and immense measure of data in different research fields. To look into in knowledge mining has offer ascent to store data and control already put away data for advance basic decision making process.

Fig 1. Data Mining-An overview

II. DATA MINING PROCESS

Data mining is also known as Knowledge Discovery in Database, refers to finding or “mining" knowledge from large amounts of data. Data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationships helpful in decision making. So, many people use the term “knowledge discovery in data” or KDD for data mining [1].

In Data mining, Knowledge extraction or discovery is done in seven sequential steps as in Fig 2.

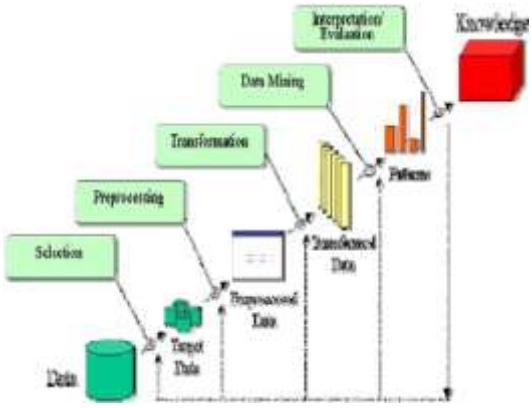

Fig 2. Data Mining Process

1) **Data cleaning:** we remove noise data and irrelevant data from collected raw data at this progression.

2) **Data integration**: At this progression, we combine multiple data sources into single data store called target data.
3) **Data Selection:** Here, data relevant to analysis task are retrieved from data base as pre-processed data.
4) **Data transformation:** Here, data is consolidating into standard formats appropriate for mining by summarizing and aggregated operations.
5) **Data Mining:** At this step, various smart techniques and tools are applied in order to extract data pattern or rules.
6) **Pattern evaluation:** At this step, strictly identify tree patterns representing knowledge.
7) **Knowledge representation:** This is the last stage in which, visualization and knowledge representation techniques are used to help users to understand and interpret the data mining knowledge or result.

III. METHODOLOGIES OF DATA MINING

There are different significant information mining methods that have been created and utilized as a part of data mining ventures as of late including **association, classification, clustering, prediction and Sequential pattern etc.,** are used for knowledge discovery from database.

- **Association:** In ARM a couple set of association rules are applied on elements of web site structure to identify relationships among various modules of web engineering application. Association rule mining is finding all association rules with support and confidence values that are greater than or equal a user-specified minsup and minconf respectively.

  An example of such a rule, mined from the *All Electronics* transactional database, is

  ***buys*(*X*; "*computer*"))*buys* (*X*; "*software*") [*support* = 1%; *confidence* = 50%]**where *X* is a variable representing a customer. A confidence, or certainty, of 50% means that if a customer buys a computer, there is a 50% chance that she will buy software as well. A 1% support means that 1% of all of the transactions under analysis showed that computer and software were purchased together.as single-dimensional association rules. Dropping the predicate notation, the above rule can be written simply as **"*computer*)*software*[1%, 50%]".**

- **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 [2]. This approach frequently employs decision tree or neural network-based classification algorithms. The common characteristics of classification tasks are as supervised learning, categories dependent variable and assigning new data to one of a set of well-defined classes. Classification technique is used in customer segmentation, modeling businesses, credit analysis, and many other applications. E.g., classify countries based on population, or classify bikes based on mileage.

  Some of the Classification techniques are:

  Regression,
  Distance
  Decision
  Rules
  Neural networks

- **Clustering :**G.P and MARTY et.al[3] examines in the paper ,how Clustering technique is useful to identify different information by considering various examples and one can see where the similarities and ranges agree. By examining one or

more attributes or classes, you can group individual pieces of data together to form a structure opinion. At a simple level, clustering is using one or more attributes as your basis for identifying a cluster of correlating results. Clustering can work both ways. You can assume that there is a cluster at certain point and then use our identification criteria to see if you are correct[4][5].

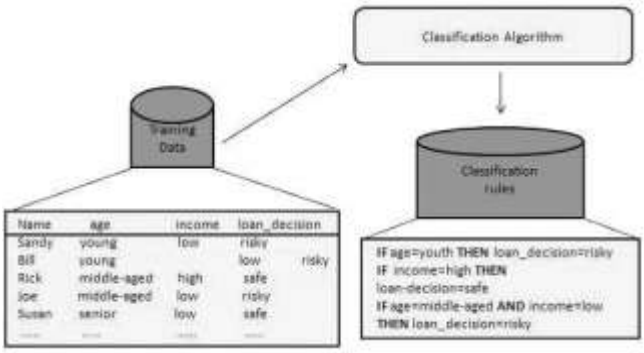

Fig 3. Clustering

- **Prediction:** The classification predicts categorical (discrete, unordered) labels, prediction models continuous-valued functions. That is, prediction is used to predict missing or unavailable numerical data values rather than class labels. But, the term prediction may refer to both numeric prediction and class label prediction.

  Example: Regression analysis is a statistical methodology that is most often used for numeric prediction, although other methods exist as well. Prediction also encompasses the identification of distribution trends based on the available data.

Applications of prediction:
Credit approval
Target marketing
Medical diagnosis
Treatment effectiveness analysis

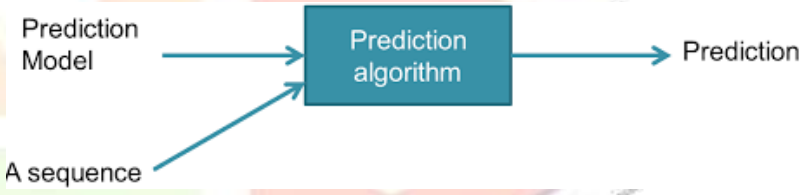

Fig 4. Prediction

- **Sequential pattern :** It is a subject of data mining worried about finding statistically applicable patterns between data illustrations where the values are conveyed in an arrangement. It is generally assumed that the qualities are discrete, and subsequently time arrangement mining is firmly related, however normally thought about an alternate movement. Sequential pattern mining is an exceptional instance of structured data mining.

IV. DATA MINING APPLICATIONS

The sensational development of the data accessible on the web and put away in big business databases has made data mining a basic undertaking for upgrading knowledge management and, generally, for gain understanding to drive decision making. A noteworthy source of this insight devices from the ability to distinguish concealed examples and connections in data.

- **Data mining applications for Finance**

A tremendous measure of data is produced in online exchanges, so the capacity to recognize the correct data at the opportune time can mean the contrast between picking up or losing a large number of dollars:

- ✓ **Increase customer loyalty** by gathering and breaking down client conduct information
- ✓ Enable banks to foresee client conduct and dispatch **relevant services and products**
- ✓ F**Discover hidden correlations between various financial indicators** to distinguish suspicious exercises with a high potential hazard

- ✓ **Improve due diligence** to speed cautions and bolster continuous basic leadership
- ✓ **dentify fraudulent or non-fraudulent actions**by gathering recorded information and transforming it into substantial and valuable data.

❖ **Data mining applications for Healthcare**

The pharmaceutical business creates a lot of records that are frequently underutilized. Data mining can enhance heath syatems and diminish costs:

- ✓ Give government, administrative and contender **information that can fuel competitive advantage**.
- ✓ **Support to the R&D procedure** and the go-to-showcase methodology with fast access to data at each period of the improvement procedure.
- ✓ **Find the connections amongst ailments and the viability of medicines** to distinguish new medications, or to guarantee that patients get proper, auspicious care.
- ✓ **Support medicinal services** safety net providers in identifying misrepresentation and mishandle.

❖ **Data mining applications for Intelligence**

Data mining investigates information and obviously distinguishes how to draw an obvious conclusion among various data components. This is a basic viewpoint for government offices:

- ✓ **Uncover concealed data** identified with illegal tax avoidance, opiates trafficking, corporate misrepresentation, fear mongering, and so forth.
- ✓ **Enhance interruption location with a high spotlight** on oddity discovery and recognize suspicious action from a the very beginning.
- ✓ **Change over content based crime reports** into word handling documents that can be utilized to help the wrongdoing coordinating procedure.

❖ **Data mining applications for Telecommunication**

The vast volumes of call, client and system information created and put away by media communications organizations require data mining to extricate concealed knowledge and distinguish valuable information to better comprehend clients and recognize extortion:

- ✓ **Pick up a competitive** and diminish client agitate by understanding statistic qualities and anticipating client conduct.
- ✓ **Increase client devotion** and enhance productivity by giving modified administrations.
- ✓ **Support client division technique** by creating proper showcasing efforts and valuing systems.

❖ **Data mining applications for Energy**

In the Oil and Gas industry, the extensive measure of unstructured data coordinated with conventional organized information offers a clear and full image of the procedure. Data mining offers strong help for the upstream oil and gas industry:

- ✓ **Catch weak signals** of possibly undermining occasions and distinguish already unidentified examples, associations and relations to engage offering, supporting and exchanging methodologies.
- ✓ **Structure identification of imperative information**, and rapidly distil it to support specialized critical thinking, enable more educated basic leadership and empower quick notice of imminent specialized leaps forward.
- ✓ **Enhance center procedures** in upstream, midstream and downstream with investigation and knowledge capacities utilizing an assortment of sources.
- ✓ **Extract**—progressively—the **pertinent knowledge** from a boundless measure of data streams with respect to your center business areas.

V.CONCLUSION

As indicated by the methods of data mining recorded above, it is found out that this a capable and basic procedure for performing manipulation of data that is data mining gives legitimate and focused on result from large and vastly developing data around the world. This paper talks about the possibility of data mining, the procedure of KDD, diverse methods, for example, association, classification, clustering, prediction and Sequential pattern. We also talked about a few experiences of the data mining applications.

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] 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

[3] BABU, G.P. and MARTY, M.N. 1994. Clustering with evolution strategies Pattern Recognition, 27, 2, 321-329

[4] Q. Yang and X. Wu, "10 Challenging Problems in Data Mining Research," Int'l J. Information Technology and Decision Making, vol. 5, no. 4, pp. 597-604, 2006.

[5] G.P.C. Fung, J.X. Yu, H. Lu, and P.S. Yu, "Text Classification without Negative Examples Revisit," IEEE Trans. Knowledge and Data Eng., vol. 18, no. 1, pp. 6-20, Jan. 2006.

[6] Deepashri.K.S , Ashwini Kamath "Survey on Techniques of Data Mining and its Applications", International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 (Volume-6, Issue-2),2017

[7] Sunil Kumar Patel, Kshipra Soni "A Survey On Data Mining Techniques and Applications" International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 5 Issue: 5 678 – 682,2017.