APPLICATION AND EFFICACY OF DATA MINING TECHNIQUES

Erappa G

Research Scholar,Opjs university, jhunjhunu Rajasthan./ CSE department.

Dr Abdul Majid

prof & HOD in CSE Dept / Dr Smce Bangalore.

Dr Yashpal pal Singh

Prof in CSE Dept, Opjs university, Jhunjhunu Rajasthan.

ABSTRACT

Data mining is a technique that involves evaluating vast volumes of data from a variety of perspectives in order to extract information that is helpful. The practice of automatically or advantageously extracting examples relevant to information that is kept or collected in huge databases, information warehouses, the Internet, or other massive data vaults or streams is referred to as data mining. This method is also commonly referred to as learning discovery from data (KDD). Information mining operations such as classification, grouping, regression, and summarization are discussed in this book, along with other information mining procedures. In addition, this study covers a variety of applications that are related to data mining.

Keywords: Data mining, data mining application, data mining technique

INTRODUCTION

With the breakthroughs that have been made in information technology, researchers from a wide range of fields now have access to enormous volumes of data that are stored in databases. An approach to the storage and management of data has been developed as a result of research into knowledge mining. This approach is intended to promote better basic decision-making.

Fig 1. Data Mining-An overview

DATA MINING PROCESS

Knowledge discovery in databases is another name for data mining, which refers to the act of finding relevant

information within enormous datasets. Data mining is sometimes known as "data mining." The use of data mining techniques to enormous datasets reveals connections and patterns that were not previously visible, which in turn helps to influence decision-making decision-making. The term "knowledge discovery in data" (abbreviated KDD) is frequently used to refer to the process of data mining.

In order to extract or discover knowledge, data mining is a procedure that is broken down into seven steps, as shown in Figure 2.
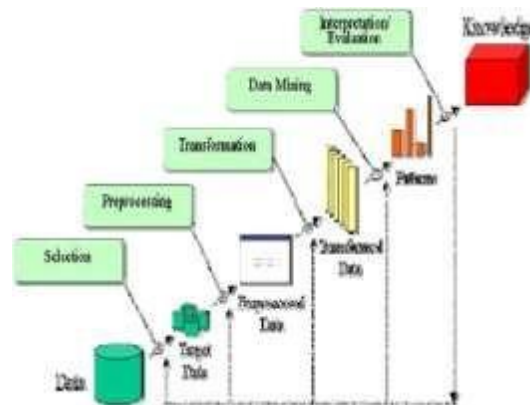

Fig 2. Data Mining Process

1. **Data cleaning:** At this point in the process, we remove noisy data and data that is irrelevant from the raw data that has been acquired.

2. **Data integration:** During this stage of the process, we remove data that is noisy and data that is irrelevant from the raw data that has been acquired.

3. **Data Selection:** In this step, data that is pertinent to the analytical task is retrieved from the database in the form of pre-processed data.

4. **Data transformation:** These processes, which include summarizing and aggregating the data, are used to consolidate the data into standard formats that are suitable for mining.

5. **Data Mining:** In order to extract data patterns or rules, this step involves the application of a variety of intelligent strategies and different instruments.

6. **Pattern evaluation:** During this stage, it is imperative to find tree patterns that are representative of knowledge.

7. **Knowledge representation:** In this final stage, approaches for visualization and knowledge representation are utilized in order to assist users in comprehending and interpreting the knowledge or result obtained by data mining.

METHODOLOGIES OF DATA MINING

Knowledge discovery from databases is made feasible by a wide range of information mining techniques that have recently been developed and are widely utilized. Some of these approaches include association,

classification, clustering, prediction, and sequential pattern, amongst others.

- **Association:** The application of a couple of sets of association rules to parts of the structure of a website is what ARM does in order to determine the associations that exist between the various modules of a web engineering application. It is the process of identifying all association rules that have support and confidence values that are more than or equal to a user-specified minsup and minconf, respectively. This is what is known as association rule mining.

  The following is an illustration of such a rule that was extracted from the transactional database of All Electronics: **buys(X; "computer"))buys (X; "software") [support = 1%; confidence = 50%]** X is a variable that stands for a customer in this circumstance. A confidence level of fifty percent, also known as a certainty level, indicates that if a consumer purchases a computer, there is a fifty percent chance that she will also purchase software. There was a one percent support, which indicates that one percent of all the transactions that were analyzed demonstrated that the computer and software were acquired simultaneously using rules of association that are only one-dimensional. In the absence of the predicate notation, the rule described above can be stated straightforwardly as "computer." software [1%, 50%]".

- **Classification:** is the data mining technique that is utilized the most frequently. This technique makes use of a collection of pre-classified samples in order to construct a model that is capable of classifying the entire population of records. This method usually makes use of classification algorithms that are based on decision trees or neural networks. Classification tasks are characterized by supervised learning, categories dependent variables, and the assignment of new data to one of a set of well-defined classes. Common characteristics of classification tasks include these qualities. There are a wide variety of applications that make use of classification technique, including consumer segmentation, modeling enterprises, credit analysis, and many others. Classify countries according to their population, for instance, or categorize bicycles according to their mileage.

  Among the several classification methods are the following:

  Regression,

  Distance

  Decision

  Rules

  Neural networks

- **Clustering:** It is examined in the study by G.P. and MARTY, among others, how Clustering is a valuable approach that can be used to detect different types of information by taking into consideration a variety of examples. This allows one to discover where the ranges and similarities are in agreement. Through the examination of one or more qualities or classes, it is possible to group different pieces of data together in order to create a judgment regarding the structure. Clustering, at its most fundamental level, is the process of recognizing a cluster of correlated outcomes by using one or more attributes as

the basis for the identification. Clustering can be used in either direction. One can make the assumption that there is a cluster at a particular site, and then use our identification criteria to determine whether or not this assumption is accurate.
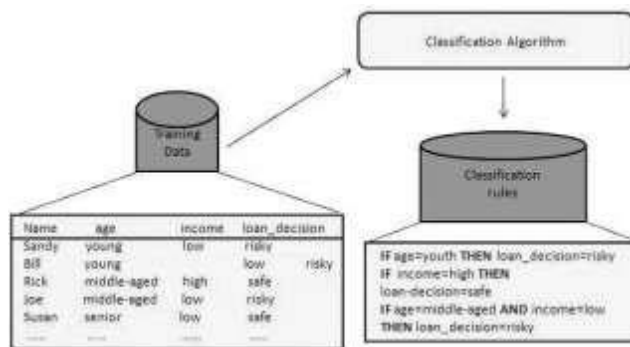

Fig 3. Clustering

Prediction: Categorical labels, which are discontinuous and unordered, are predicted by the classification, which uses continuous-valued functions as prediction models. In other words, rather than class labels, prediction is used to make predictions about numerical data values that are either unavailable or missing. Prediction, on the other hand, can relate to both the prediction of numerical values and the prediction of class labels.

Example: Although there are numerous ways available, the statistical methodology known as regression analysis is the one that is utilized for the purpose of numerical prediction the majority of the time. The process of prediction also includes the detection of distribution trends based on the data that is already available.

Applications of prediction: An approval of credit Identifying potential customers Medical diagnostics Evaluation of the efficacy of each treatment
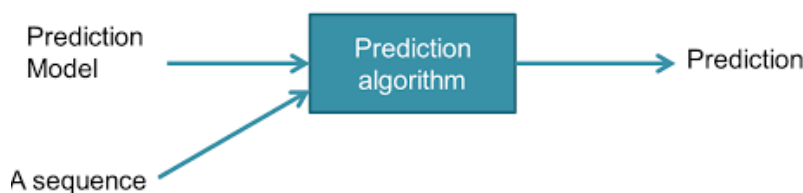

Fig 4. Prediction

- **Sequential pattern:** It is a subfield of data mining that focuses on discovering patterns that are statistically applicable between data visualizations in which the values are presented in a particular arrangement. In most cases, it is presumed that the characteristics are discrete, and hence, time arrangement mining is closely connected to it. However, it is typically considered to be an alternative movement. In the realm of structured data mining, sequential pattern mining stands out as a unique example.

OBJECTIVES

1. To study data mining techniques.

2. To study efficacy of data mining techniques.

DATA MINING APPLICATIONS

Data mining has become an essential activity for improving knowledge management and, more generally, for gaining understanding to drive decision making as a result of the phenomenal growth of the data that is accessible on the internet and stored in large enterprise databases. One of the most important sources of this knowledge comes from the capability of recognizing hidden examples and linkages within the data.

Data mining applications for Finance

The ability to recognize the appropriate data at the appropriate time can mean the difference between gaining or losing a significant amount of money. This is because internet transactions generate a vast amount of data, which can be used to make significant financial decisions.

- Increasing customer loyalty can be accomplished by collecting and analyzing information regarding client behavior.

- Make it possible for financial institutions to anticipate the behavior of their customers and to provide them with the appropriate services and solutions.

- The identification of hidden connections between different financial indicators is necessary in order to differentiate between suspicious activities that have a high potential risk.

- Enhance the level of due diligence in order to expedite cautions and strengthen the continual basic leadership.

- identification of fraudulent or non-fraudulent behaviors through the collection of recorded information and the transformation of that information into substantial and valuable insights.

Data mining applications for healthcare

Numerous records are produced by the pharmaceutical industry, the most of which are not exploited to their full potential. In addition to lowering costs, data mining can improve healthcare systems:

- Make information available to the government, administrative bodies, and competitors that can help fuel a competitive edge.

- Support to the R&D procedure and the go-to-showcase approach with fast access to data at each period of the improvement operation.

- In order to identify new treatments or to ensure that patients receive appropriate and beneficial care, it is necessary to investigate the links that exist between diseases and the efficacy of medications.

- Provide assistance to providers of the safety net for medical services in identifying instances of misrepresentation and mishandling.

Data mining applications for Intelligence

The process of data mining involves the investigation of information and the differentiation of how to arrive at a conclusion from among the numerous components of the various data. This is a fundamental perspective for anyone working in government:

- Uncover hidden information that has been linked to unlawful tax avoidance, the trafficking of opioids, corporate deception, scare mongering, and other similar activities.

- Improve the location of interruptions by placing a strong emphasis on the identification of unexpected occurrences and identifying suspicious behavior from the very beginning.

- Convert content-based crime reports into word processing documents that can be used to assist in the process of coordinating the investigation of wrongdoing.

Data mining applications for Telecommunication

The huge amounts of call, client, and system information that are generated and stored by media communications firms necessitate the use of data mining in order to extract hidden information and differentiate important information in order to better understand customers and identify instances of extortion:

- Make a competitive offer and reduce the level of client agitation by gaining an awareness of the characteristics of statistics and predicting the behavior of customers.

- Increasing customer loyalty and boosting productivity can be accomplished through the provision of customizable services.

- Provide assistance to the client division method via the development of appropriate showcasing efforts and valuation systems.

Data mining applications for Energy

In the oil and gas business, a comprehensive and clear picture of the process can be obtained by combining the enormous amount of unstructured data with the typical organized information. A significant amount of assistance is provided by data mining to the upstream oil and gas industry:

- Identify precarious indications of potentially destabilizing occurrences and differentiate between examples, associations, and relationships that have not yet been established in order to engage in offering, supporting, and sharing techniques.

- In order to facilitate specialized critical thinking, permit more informed basic leadership, and enable prompt notice of approaching specialized leaps forward, it is necessary to structure the identification of essential information and rapidly distil it.

- Improve center procedures in upstream, midstream, and downstream by utilising a variety of sources to conduct research and acquire information utilising expertise.

- It is necessary to extract, in a step-by-step manner, the relevant information from an infinite number of data streams pertaining to your core business areas.

CONCLUSION

Data mining is a genuine and focused on result from enormous and greatly developing data all over the world, as suggested by the ways of data mining that were recorded above. It is found out that this is a capable and basic procedure for performing modification of data that is data mining. This article discusses the possibilities of data mining, the process of KDD, and a variety of approaches, including association, classification, clustering, prediction, and sequential pattern, among others. In addition to that, we discussed a few of our experiences with the applications of data mining.

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, 2021

3. BABU, G.P. and MARTY, M.N. 2014. 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, 2016.

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. 2016.

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.

8. Jiawei Han and Micheline Kamber (2016), Data Mining Concepts and Techniques, published by Morgan Kauffman, 2nd ed.

9. Dr. Gary Parker, vol 7, 2014, Data Mining: Modules in emerging fields, CD-ROM. Crisp-DM 1.0 Step by step Data Mining guide from http://www.crisp-dm.org/CRISPWP-0800.pdf.

10. Han Jiawei, Micheline Kamber, Data Mining: Concepts and Technique. Morgan Kaufmann Publishers,2020

11. W.M.R. Tissera, R.I. Athauda, H. C. Fernando "Discovery of Strongly Related Subjects inthe Undergraduate Syllabi using Data Mining", IEEE International Conference on Information Acquisition, 2016

12.　Sri Lanka Institute of Information Technology, http://www.sliit.lk/, retrieved on 28/02/2016

13.　Sun Hongjie, "Research on Student Learning Result System based on Data Mining", IJCSNSInternational Journal of Computer Science and Network Security, Vol.10, No. 4, April 2020

14.　Wayne Smith, "Applying Data Mining to Scheduling Courses at a University", Communications of the Association for Information Systems, Vol. 16, Article 23, 2015

15.　Firdhous, M. F. M. (2020). Automating Legal Research through Data Mining. International Journal of Advanced Computer Science and Applications - IJACSA, 1(6), 9-16.

16.　The Result Oriented Process for Students Based On Distributed Data Mining. (2018). International Journal of Advanced Computer Science and Applications - IJACSA, 1(5), 22- 25.

17.　Jadhav, R. J. (2021). Churn Prediction in Telecommunication Using Data Mining Technology. International Journal of Advanced Computer Science and Applications - IJACSA, 2(2), 17-19.

18.　Shaeela Ayesha, Tasleem Mustafa, Ahsan Raza Sattar, M. Inayat Khan, "Data Mining Modelfor Higher Education System", Europen Journal of Scientific Research, Vol.43, No.1, pp.24- 29, 2019

19.　K. H. Rashan, Anushka Peiris, "Data Mining Applications in the Education Sector", MSIT, Carnegie Mellon University, retrieved on 28/01/2016