RESEARCH ARTICLE   OPEN ACCESS

A Review On Data Mining Techniques

Parth Gupta*, Garvit Maheshwari*, Pawan Sen**, Naveen Kumar Tiwari**

*B.Tech Student, Department of CSE, Arya Institute of Engineering & Technology, Jaipur
** Assistant Professor, Department of CSE, Arya Institute of Engineering & Technology, Jaipur

ABSTRACT
methodical changes and experiments are done to make the technologies advance. this advancement led to the exposure of a different wide range of new fields. some of these fields are science, business, education, health, and fitness produce a high amount of daily data. the technique which is used to control and extract valuable and useful information is data mining. This paper provides an overview of what is data mining and how data mining is used in different fields using various techniques.
Keywords — Mining, MI, Data, Challenges, data processing, Evaluation.

I. INTRODUCTION

The process of finding hidden data and predicting future information and events has a long history. Although data mining is currently considered a new technology all over the world, it has been widely known since the early 90s. This is the most useful technique used to find specific and hidden information in large amounts of data. It also turns raw data into useful information for solving problems.Data mining is like data science, performed by humans under specific circumstances, using specific data sets for specific purposes.

Data mining is also known as KDD (knowledge discovery data). KDD is a method in which data is given as input and useful information is taken as an output. the KDD consist of five methods selection, preprocessing, transformation, data mining, and evaluation. all these methods are used to change the raw data into meaningful information. data mining tasks are divided into two categories descriptive and predictive. descriptive data mining finds data patterns and useful data from available information and on another side, predictive data mining predicts the information that is not specifically available.

Data mining is the process of searchingnew patterns and relationships between largeinformation to predict the future.Using a variety of strategies, we can use this information to increase profit, decrease expenses, improve customer relationships, reduce the chances of risks and improve future facilities. Other than its applications in business it can be used in various sectors such as digital marketing, health care, social media, banks and insurance respectively.

II. DATA MINING STEPS

In the field of business, data is most important to understand what customer needs from companies and what changes they can make along with the time to fulfill the needs of the customer if an industry can predict what customer wants so they can make a profit it can only possible when they can predict the future needs of customers it can do with the help of data mining.

Here are the some data mining steps:

Figure 1 : Overview of Data Mining Steps

Identify the problem:
Initially identifYING the problem is important .we need to understand that clean data is the most important aspect of the business the unwanted and dirty data may lead to plan failure and increase the maintenance expenses.to solve all these problems we have to set some goals and get the required data. this is the basic step to identifying what kind of data is required or whether it is profitable or not in the business.
Identifying required data
After finding the problem collect all the information which is needed to understand the solution and remove the unwanted data and take out the quality data. analysis of data also provides the future use of data and the requirement of data preprocessing. this step helps to extract useful information.
Preparing and pre-processing
In preprocessing preparing useful required data is the main focus. many companies use this technique to remove the irrelevant and unwanted data to format the raw data and change it into the required data. so in the end we have many data sets and we can combine them into processed data. addition of some external information also makes the data more useful.
Modelling

This is the most important part of the data mining process in this section certain algorithms are used to transform the data according to the business goals to get the optimized data results and generate the final data sets, with the help of different algorithms and tools we improve the productivity and create a data model to test the results

Training and testing

This section is important for the evaluation of data .in training and testing we determine whether the models are correct or not. this process is performed multiple times until the results are not satisfying or in optimized form. the models are generally sets of information if the information is correct then companies can take advantage of this information to make a profit.

Verify and deploy

Check the final model is ready with useful information and that the resulting model is ready for deployment. many companies use these models for decision making and to predict the growth of the business if they find anything useful in this process then they will make the final process to execute the plan.

III.   DATA MINING TECHNIQUES

Data mining includes the usage of subtle facts evaluation equipment to discover formerly unknown, legitimate styles and relationships in huge facts sets. These equipment can encompass mathematical models, gadget mastering strategies, and mathematical algorithms, which includes neural networks or selection trees.

Thus, facts mining includes evaluation and prediction. Depending at the form of techniques and technology springing up from gadget mastering, facts management, and mathematics, facts mining specialists have furnished their offerings to higher recognize a way to system and draw conclusions approximately huge quantities of facts. a number of important facts mining techniques had been evolved and implemented, consisting of association, classification, clustering, prediction, sequential styles, and regression.
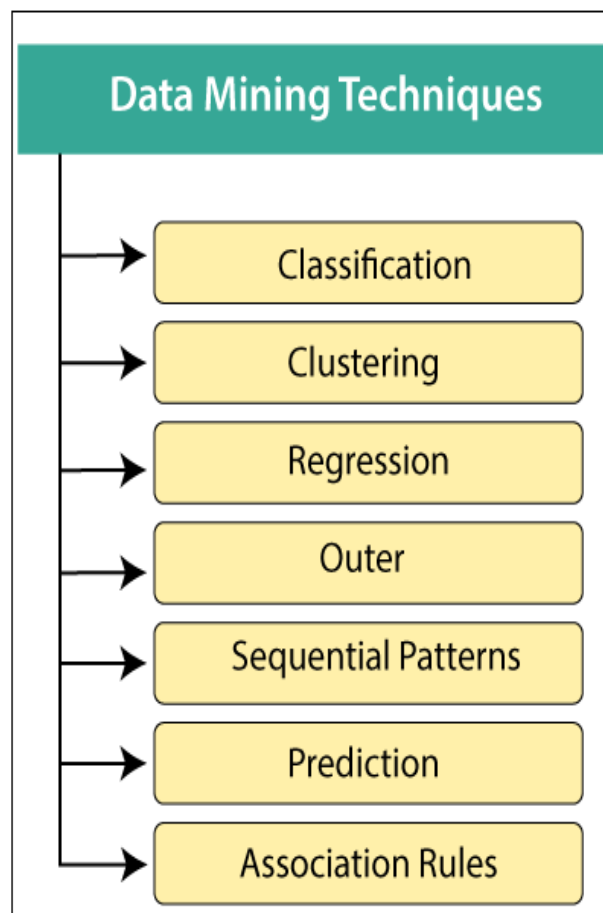

Figure 2 : Classification of Data Mining Techniques

Classification Analysis:

In the classification, WE use algorithms and tools to find out how the data can be determined and divided into the different categories. thisprocess's main target is to divide the data into different sets.this process is used to get the optimized and important information which is useful in future.

Clustering Analysis:

Cluster evaluation is a method by which you {WE}can check the correlation between the different patterns of data. It is the simplest way to check the similarities between the two sets and the outcome of this analysis help to improve the business profile.it also helps to manage the data in the different sets according to the similarities.

Regression Analysis:

Regression is a method used to check the probability of occurring an event .it is the technique to predict the useful information .by this method the possibilities to identify the presence of different factors in between the two data sets are increased.it is the best way to increase the use of optimization information.

Association Rule Learning:

The association rule method is used to discover the hidden link between the two different extraordinary types of sets of information. this process helped to predict the behavior of a person by the use of some unique facts and specific data by

applying some algorithm. this technique has several kinds of applications but the two most common uses are in the field of clinical and sales.

Anomaly or Outlier Detection:

In this technique, the set of data items is under observation from which we detect the different groups of patterns and behavior. after detection, they provide some useful and valuable information. outliers are deviating from the conventional relationship between a piece of information and a mixture of data all these things are hypothetical which is a hundred percent practically not possible. it also suggests that something uncommon has taken place and wants greater attention.

IV.   DATA MINING APPLICATION

We can use Data mining in the process of finding confusing, patterns and relationships between large data sets to predict results. Using a variety of strategies, we can use this information to increase revenue, reduce costs, improve customer relationships, reduce risks and more

Education: In today's world, everyone wants innovation and also the everyday demand for new things is also increasing, everybody wants a solution to growing their needs .data mining helps the students to decide on their careers and also helps in higher education by predicting the behavior and interests of the students. data mining helps to make new changes and selections in an exceedingly simpler manner.

Fraud Detection: the abnormal data, activity, and events are detected by software the use of data mining helps in detecting the behavior of a person which helps the companies to detect the fraud mostly known in the monetary and financial organizations many companies use these varieties of techniques to remove the fake and fraud user data from their database.

Counter Terrorism: Some advanced approaches like IDM (investigative data mining) and complex algorithms are used to extract the hidden information and data, these information are indicating an organized crime activity. the main roles in the investigation are distinctive past criminal records, networks, and organizational links. data mining is also used to find where to deploy forces and what action taken according to the situation.

Financial Analysis: Data mining strategies may be very beneficial to carry out the economic analysis efficiently and accurately. it allows commercial enterprises to increase their efficiency decrease expenses and develop at a quicker pace it also helps the monetary organization to predict the payment of loans and rating of payment cards.

Spatial Data Mining: Geographic Information Systems (GIS) and some different roaming programs use facts mining to defend vital statistics and recognize its impacts. This new exercise consists of the discharge of Geographical, environment, and astronomy facts, together with photographs from space. In general, neighborhood facts mining can screen capabilities which includes topology and distance.

Research Analysis: History shows that we have seen changes in research. Data mining helps with data processing and data integration. Researchers may find any similar data on a website that may bring any changes to the study. The identification of any sequence that occurs in the relationship between any activities is unknown. Data visualization and visual extraction of data gives us a clear view of the data.

Retail: Major customer websites have hidden customer insights that can help you improve relationships, improve marketing campaigns and sales forecasts. With more accurate data models, marketing companies can offer more targeted campaigns and get offers that make a big impact on the customer.

Manufacturing: Data processing tools may be useful to find patterns during an advanced production process. data processing can be utilized in a system-level style to extract relationships between product structure and client desires data. It also can be accustomed predict development time, costs, and alternative activities.

V. CONCLUSION

To conclude, data mining is a very powerful and useful method and technology generate decision-making information. The Future development of it is expected to make data mining more powerful and more useful. Data mining is the extraction of information that can predict future trends and behavior which helps to bring amazing information to businesses on the problems they face and helps identify new opportunities and also helps in security of data and its privacy.

REFERENCES

[1]   Rahul Misra and Ramkrishan Sahay, "A Review on Student Performance Predication Using Data Mining Approach", International Journal of Recent Research and Review, vol. X, no. 4, pp. 45-47, December 2017.

[2]   P. Ristoski and H. Paulheim, "Web Semantics : Science , Services and Agents on the World Wide Web Semantic Web in data mining and knowledge discovery : A comprehensive survey," vol. 36, pp. 1–22, 2016.

[3]   K. Choudhary, J. A. Harding, and M. K. Tiwari, "Data mining in manufacturing: A review based on the kind of knowledge," J. Intell. Manuf., vol. 20, no. 5, pp. 501–521, 2009.

[4]   H. Arora, R. Agarwal, P. Sharma, G. Shankar and D. Arora, "Image Security Utilizing Hybrid Model of Steganography and Asymmetric Cryptography Methods," 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), pp. 277-280, 2023.

[5]   C.-L. Huang, M.-C. Chen, and C.-J. Wang, "Credit scoring with a data mining approach based on support vector machines," Expert Syst. Appl., vol. 33, no. 4, pp. 847–856, 2007.

[6]   Dr. Himanshu Arora, Gaurav Kumar Soni and Deepti Arora, "Analysis and Performance Overview of RSA Algorithm", International Journal of Emerging

Technology and Advanced Engineering, vol. 8, no. 4, pp. 10-12, 2018.

[7] Arpita Tiwari, Gori Shankar and Dr. Bharat Bhusan Jain, "Digital Image and Text Data Security Improvement Using The Combination of Stenography and Embedding Techniques", Design Engineering, no. 7, pp. 8592-8599, 2021.

[8] S. Mishra, D. Singh, D. Pant and A. Rawat, "Secure Data Communication Using Information Hiding and Encryption Algorithms", IEEE 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS), pp. 1448-1452, 2022.

[9] Shweta Saraswat, Bright Keswani and Vrishit Sarasawat " The role of Artificial Intelligence In Healthcare: Applications and Challenges after COVID-19" IJTRS Apr. 2023.

[10] Rahul Misra and Ramkrishan Sahay, "Evaluat ion of Student Performance Prediction Models with Two-Class Using Data Mining Approach", International Journal of Recent Research and Review, vol. XI, no. 1, pp. 71-79, March 2018.

[11] Kiran Ahuja, Harsh Sekhawat, Shilpi Mishra and Pradeep Jha, "Machine Learning in Artificial Intelligence: Towards a Common Understanding", Turkish Online Journal of Qualitative Inquiry (TOJQI), vol. 12(8), pp. 1143-1152, 2021.

[12] Himanshu Arora, Shilpi Mishra and Manish Dubey, "Development of the Framework for the Solution of the Security Problems in Data Transmission Involving Advanced Asymmetric Algorithm", International Journal of Emerging Technology and Advanced Engineering, vol. 8, no. 4, pp. 18-20, April 2018.

[13] P. Jha, R. Baranwal, Monika and N. K. Tiwari, "Protection of User's Data in IOT," 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS), pp. 1292-1297, 2022.