



Data Mining Techniques: A Review of Literature

¹Prof. S. B. Bele, ²Vaishnavi G. Shinde, ³Yash S. Kalaskar, ⁴Dipali V. Mendhe

¹Assistant Professor, Department of MCA, Vidya Bharati Mahavidyalaya, Amaravati, India

^{2,3,4} Student, Department of MCA, Vidya Bharati Mahavidyalaya, Amaravati, India

Date of Submission: 20-10-2023

Date of Acceptance: 04-11-2023

Abstract

Data mining is very important and useful. It involves using various algorithms to extract valuable knowledge and insights. It allows us to obtain diverse information from datasets and use it to perform various tasks.

Keywords

Data Mining Techniques, Classification, Predication, Knowledge base.

I. Introduction

Information and knowledge from complex engineering design to scientific research can be quickly and effectively applied. Human life has become increasingly characterized by data, which is essential in creating data-driven solutions. The collection of data from numerous sources, stored in large and diverse warehouses, forms the basis of this process.

The data collection thus exceeds the human aptitude for analysis without a powerful analysis tool, as an effect these warehouses become 'data vaults', that are not often visited.

Data mining is simple process of sorting the large amount of data sets to identify patterns and relationship that can help solve our problems through data analysis.

Data mining is the process of analysing and extracting valuable insights and meaningful interpretations from large amounts of data, such as customer feedback and product reviews. By using techniques like natural language processing and sentiment analysis, businesses can gain valuable insights into customer preferences and improve their products.

II. Review of literature

Based on your messages, the optimization technique specifically applied to cancer patients and supports treatment decisions using cluster analysis. This approach can help strengthen validity. To put it simply, the technique involves using a method on the first tap and creating a new configuration by

finding differences in the existing rules. The K-means algorithm is then used to improve the visualizations of the results. These various social media platforms have made complex tasks easier. Social media data itself is vast in quantity, dynamic, and intricate.

It particularly creates a rich and dynamic set for social media communication data and meta-data, which has not been systematically treated in the literature on data and text mining. Literary review refers to the survey of various scholarly sources (such as books, journal articles, and research papers) related to a specific topic or research question. It is often written as part of a thesis, dissertation, or research paper to contextualize your work within existing knowledge.

III. What is Data Mining

Data mining is a process of extracting valuable insights from data.

Analysing samples from a large dataset that incorporates methods from machine learning, statistics, and database systems. This is a field of computer science, and has various applications in different industries. Data science that aims to retrieve information from data (using intelligent methods) for specific purposes and transform the information into a usable structure for further applications.

Data mining is very important! By using it, you can explore your data to uncover hidden insights and improve your business relationships, customer behaviour, and performance. Data mining can help your business and decision-making. This is a multi-disciplinary skill that uses machine learning, statistics, and AI to gather information for evaluating the likelihood of future events.

Types of Data Mining Techniques:

A. Tracking patterns:

Tracking patterns is a fundamental data mining technique. It involves identifying to extract trends or patterns from data, we can use data mining



techniques intelligent conclusions about business outcomes and testing them. When historical data doesn't provide insights, businesses can recommend products to customers based on that information. These recommendations can lead to profitable deals and satisfied customers.

B. Decision Tree:

In this system, the decision tree is a division of the original dataset. Each data in the partition is compared with the estimated information. The root node of the tree is taken as the initial point and the appropriate class label is estimated for each data entry, which is used to classify the data until further classification is not possible.

C. Prediction:

In this data mining framework, historical data and the sequence of events are used to understand patterns and estimate future event outcomes. One widely used application of prediction is to forecast sales and determine the profitability for businesses. Prediction is a use of data mining techniques, such as sequential patterns, trends, clustering, and classification. They decode past events or examples in a suitable order to estimate future events.

D. Association:

Association or relationship mining is related to data analysis. Using this data mining technique, data scientists and analysts can discover associations or correlations between two or more data attributes. It informs about the frequency events of objects in behaviour dataset. One benefit of this system is that it passes through the database with fewer numbers. This makes it suitable for solving problems like analysing customer behaviour.

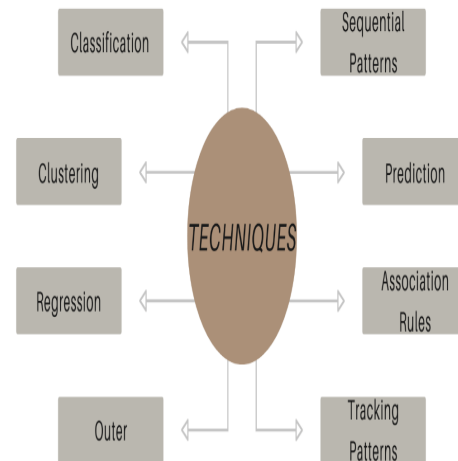
E. Neural networks:

Neural networks, or artificial neural networks, are computational models inspired by biological neural networks. They are extensively used for data mining tasks, such as artificial intelligence and machine learning, which involve significant utilization of computational power and extensive data analysis neural networks, there is a notable ability to extract meaning from complex or noisy data and to be used for cleaning impure data and for trend analysis, which can be quite intricate. They are similar to a human or another computer system.

F. Sequential Patterns:

Association mining helps discover relationships between events and actions. For example, let's say there is a customer who makes more purchases than usual in the first quarter of the year. There is a possibility that they will make at least one purchase in the second quarter as well.

Data Mining Techniques



IV. Advantages:

- Direct mail, online marketing, and other new marketing initiatives can elicit responses from targeted customers, resulting in profitable sales.
- Data mining can help businesses automate processes and identify inefficiencies in their operations, thereby aiding in recognizing patterns and improving productivity.
- Each customer in the market is unique in their behaviour and characteristics. Understanding their fundamental traits and preferences is essential.
- Data mining can help businesses identify market differences and assist in developing new products or services to fill those gaps.
- It is also useful for creating algorithms for machine learning and is suitable for designing specific AI applications and programs.

V. Disadvantages:

- To extract more accurate and efficient information, a large database, storage space, and processing power are necessary.
- The previous bullet point mentioned that data mining requires investment in database, storage, and processing power.



Future Scope:

By incorporating various specified algorithms in our future work, we can increase accuracy, resulting in more precise outcomes. Data sets that identify hidden patterns will be utilized, primarily focusing on research and analysis to benefit from those insights. The system is very relevant. It can help users in various ways, such as researching substances and providing valuable treatment options.

It can also assist in preserving the value of human life. Data mining is widely used and popular for extracting necessary information. Data mining also helps banks understand their customers' online behaviour and preferences in a better way, which is very helpful for creating new marketing campaigns.

VI. Result:

According to the results of the analysis, there is an increasing focus on performance in educational data mining. With the improvement in our studies, it has been concluded that data mining is generally cantered around the estimation of student performance. Selecting the right curriculum and providing timely warnings about students' education is an important task in SPP, but a significant component for achieving meaningful outcomes in educational practices is to identify the major factors influencing academic performance through research.

VII. Conclusion:

The Internet of Things (IoT) has emerged to manage, automate, and explore all devices, tools, and sensors in the world. It has been brought together with data mining for supporting decision-making for people and things in IoT, and for system optimization. In the application view, we review data mining applications in e-commerce, industry, healthcare, and public services. The discussion includes technical views, knowledge views, and application views.

Reference:

- [1]. https://www.researchgate.net/publication/49616224_Data_mining_techniques_and_applications
- [2]. https://www.academia.edu/45618032/Top_10_Datamining_Papers
- [3]. https://www.worldwidejournals.com/paripex/recent_issues_pdf/2013/June/review-of-data-mining-techniques-applications
- [4]. <https://arxiv.org/ftp/arxiv/papers/1211/1211.5723.pdf>
- [5]. <https://www.slideshare.net/ernancy/a-review-on-data-mining>
- [6]. <https://www.sciencedirect.com/science/article/pii/S1877050915023479>
- [7]. <https://www.techtarget.com/searchbusinessanalytics/definition/data-mining>
- [8]. <https://www.tableau.com/learn/articles/what-is-data-mining>