



JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Data Mining Techniques : A Review

¹Prof.K. V. Raghuwanshi , ²Dr.V. R. Dhawale , ³Y. P. Rawale , ⁴Y. G. Bule , ⁵R. R. Sakhare

¹Professor, Department of MCA, Vidyabharati Mahavidyalaya, Amravati, India

²Professor, K. K. Wagh Institute of Engineering Education and Research, Nashik, India

^{3,4,5}Student, Department of MCA, Vidyabharati Mahavidyalaya, Amravati, India

Abstract

Data mining is the procedure of extracting private and useful patterns and information from data. With collaborative filtering techniques becoming more and more mature, recommender systems are widely used nowadays, especially in electronic commerce and social networks. However, the utilization of a recommender system in academic research itself has not received enough attention. This paper aims to show the data mining process and how its technology helps decision-makers make better decisions. Practically, data mining is useful for any organization which has a huge amount of data.

Keywords

Data mining, Prediction, technology, Decision tree, Database, Classification, Association, Neural Networks

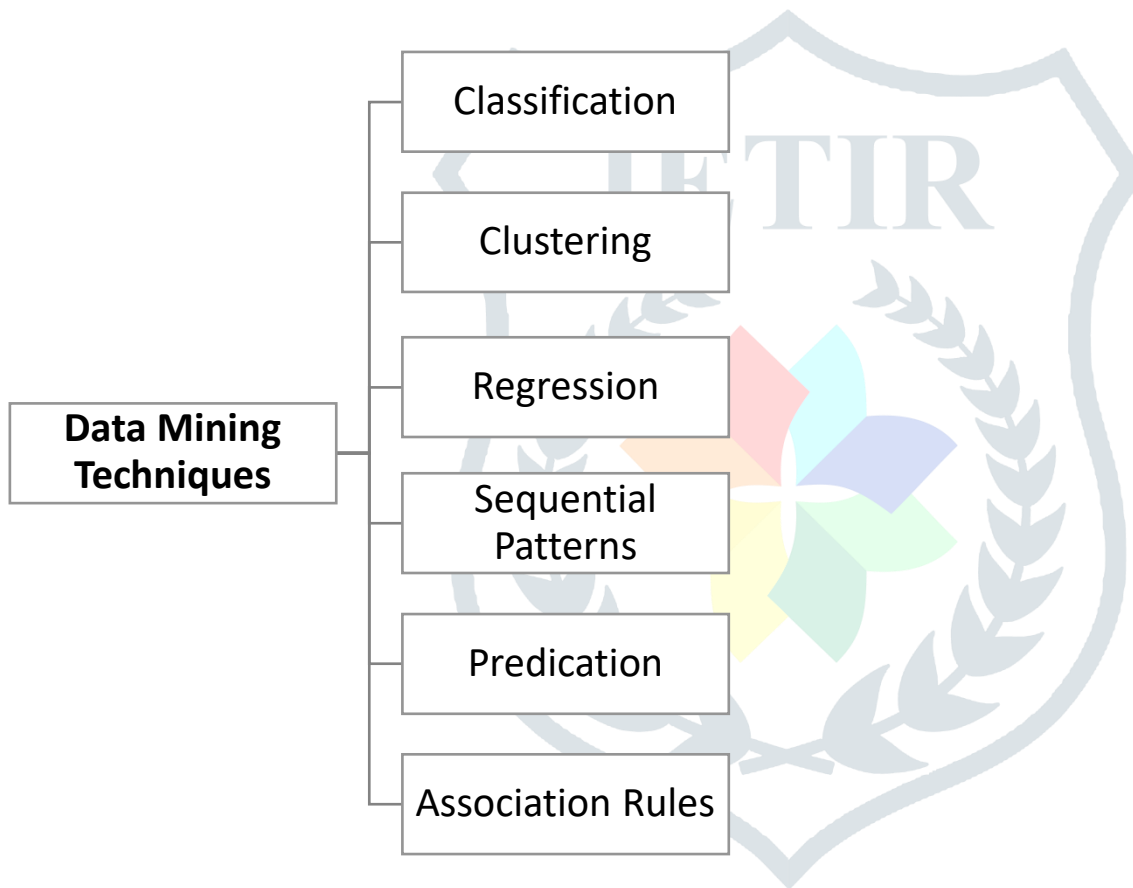
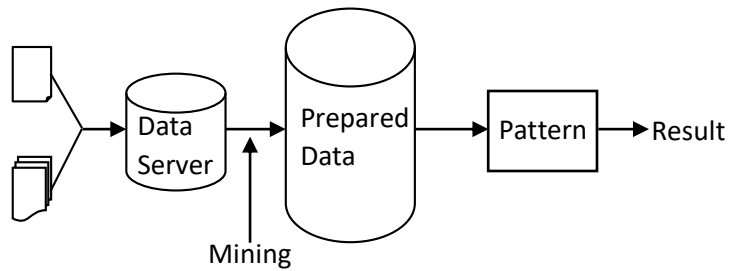
Introduction

Data mining is the procedure of using statistical methods to uncover patterns and insights within large datasets. Typically, the datasets used for data mining are so large that it would take days, weeks, or months for humans to read or analyze them. Consequently, data mining often requires using programs, machine learning, or artificial intelligence to do the work. However, human analysts or database administrators often need to be required, and you need to clean your data, so your datasets are adapted for analysis. With controlled data, your data stewards

need to be knowledgeable f these methods to train machines to uncover these insights and oversee their results to verify they are correct.

Data Mining Techniques:

Aim:



Data mining is a big area of data sciences, which aims to determine patterns and features in data, often large data sets. It includes regression, classification, clustering, outer, and others. It also includes sequential patterns, Prediction, Association Rules, and ultimately the making sense of the data sets.

How does it work?

Data mining works through the concept of predictive modeling. Suppose an organization wants to accomplish a particular result. By analyzing a dataset where that outcome is known, data mining techniques can, for example, build a software model that analyzes new data to predict the likelihood of similar results. Here is an overview:

Start with historical data

Let's say a company wants to know the best customer prospects in a new marketing database. It starts by examining its customers.

1. Analyze the historical data

The software scans the collected data using a combination of algorithms from statistics, artificial intelligence, and machine learning, looking for patterns and relationships in the data.

2. Write rules

Once the patterns and relationships are uncovered, the software expresses them as rules. A rule might be that most customers ages 51 to 65 shop twice a week and fill their baskets with fresh foods, while customers ages 21 to 50 tend to shop once a week and buy more packaged food.

3. Apply the rules

Here, the data mining model is applied to a new marketing database. If the company is a packaged food provider, it will be looking for 21-to-50-year-olds.

Types of Data Mining techniques

1) Classification

Classification is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. For example, a classification model could be used to identify loan applicants as having low, medium, or high credit risks.

2) Clustering

Cluster Analysis in Data Mining means that to find out the group of objects which are like each other in the group but are different from the object in other groups. In the process of clustering in data analytics, the sets of data are divided into groups or classes based on data similarity. Then each of these classes is labeled according to their data types. Going through clustering in data mining examples can help you understand the analysis more extensively. For example, consider a vehicle dataset that contains information about vehicles like cars, buses, bicycles, etc. As it is unsupervised learning there are no class labels like Cars, Bikes, etc. for all the vehicles, all the data is combined and is not in a structured manner

3) Regression

Regression is a data mining function that forecasts a number. Profit, sales, mortgage rates, house values, square footage, temperature, or distance could all be predicted using regression techniques. It is also used in various industries for business and marketing behavior, trend analysis, and financial forecast. The most popular types of regression are linear and logistic regressions. Other than that, many other types of regression can be performed depending on their performance on an individual data set.

Regression is divided into five different types

4) Sequential Patterns

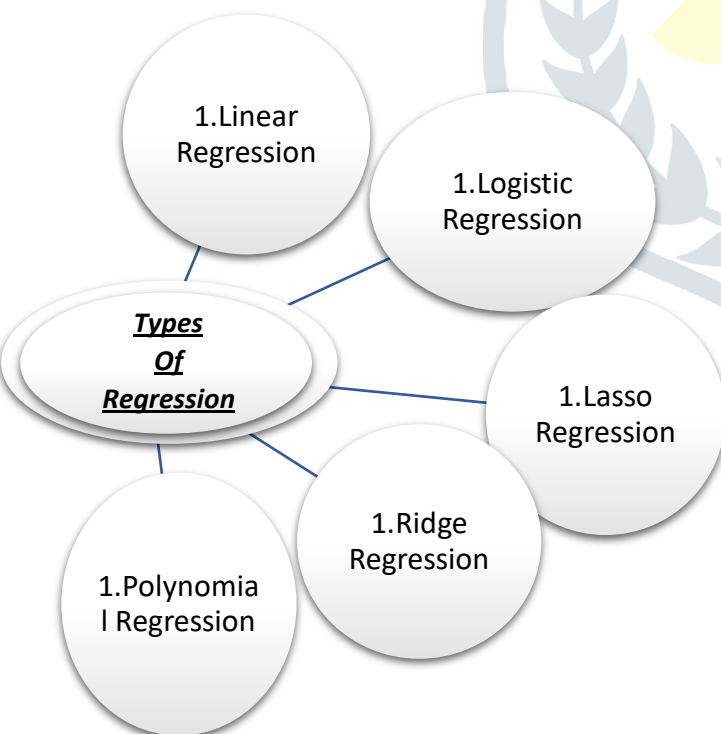
Sequential pattern mining is the mining of again and again appearing series of events or subsequences as patterns. A detail of a sequential pattern is users who purchase a Canon digital camera are to purchase an HP color printer within a month. Sequential pattern mining is an instance case of structured data mining.

For peddling information, sequential patterns are favorable for shelf placement and promotions. This industry, telecommunications, and separate businesses can also use sequential patterns for targeted marketing, user retention, and several tasks.

There are certain areas in which sequential patterns can be used such as Web access pattern analysis, weather prediction, production processes, and web intrusion detection.

5) Prediction

To find a numerical output, prediction is used. The training dataset contains the inputs and numerical output values. According to the training dataset, the algorithm generates a model or predictor. When fresh data is provided, the model should find a numerical output. This approach, unlike classification, does not have a class label. A continuous-valued function or ordered value is predicted by the model. A prediction is what someone thinks will happen. A prediction is a forecast, but not only about the weather. Pre means “before” and diction has to do with talking. So, a prediction is a statement about the future. It's a guess, sometimes based on facts or evidence, but not always. For example, suppose the marketing manager needs to predict how much a particular customer will spend at his company during a sale. We are bothered to forecast a numerical value in this case. Therefore, an example of numeric prediction is the data processing activity. In this case, a model or a predictor will be developed that forecasts a continuous or ordered value function.



6) Association Rule

Association rule mining finds interesting associations and relationships among large sets of data items. This rule shows how frequently an item set occurs in a transaction. A typical example is a Market Based Analysis.

Market-Based Analysis is one of the key techniques used by large relations to show associations between items. It allows retailers to identify relationships between the items that people buy together frequently.

Association rule mining, at a basic level, involves the use of machine learning models to analyze data for patterns, or co-occurrences, in a database. It identifies frequent if-then associations, which themselves are the association rules. An association rule has two parts: an antecedent (if) and a consequent (then). Example of Association rule in data mining A supermarket has 200,000 customer transactions. About 4,000 transactions, or about 2% of the total number of transactions.

Future Scope

The use of data mining in enrollment management is a fairly new development. Current data mining is done primarily on simple numeric and categorical data. In the future, data mining will include more complex data types. In addition, for any model that has been designed, further refinement is possible by examining other variables and their relationships. Research in data mining will result in new methods to determine the most interesting characteristics of the data. As models are developed and implemented, they can be used as a tool in enrollment management.

Conclusion

Strategic Enrollment Management to achieve desired enrollment targets in colleges and universities. In situations where it has been applied, it has been proven to successfully predict enrollment, at least to a degree. More research is needed to fully take advantage of the data mining processes and technologies.

References

Javatpoint -- <https://www.javatpoint.com/data-mining-techniques>

Google Books -- <https://www.shorturl.at/ciyOQ>

GeeksForGeeks-- <https://www.geeksforgeeks.org/data-mining-techniques/>

Educba -- <https://www.educba.com/data-mining-methods/>

Wikipedia--https://en.wikipedia.org/wiki/Data_mining

ABDULLAHI SIDOW OSMAN--Al-

dinah International University

<http://ojs.mediu.edu.my/index.php/IJDSR/article/view/1841>

Application of Data Mining Techniques to Healthcare Data

Published online by Cambridge University Press: 02 January 2015

<https://www.cambridge.org/core/journals/infection-control-and-hospital-epidemiology/article/abs/application-of-data-mining-techniques-to-healthcare-data/7EE5E7B1FA8B1C535FBC7A3881EC42E0>

Anurag Kumar and Kumar Ravi Singh, "A Study on Web Structure Mining," International Research Journal of Engineering and Technology (IRJET), vol. 04, no. 1, pp. 715-720, January 2017.