IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114173

DATA MINING TECHNIQUES AND APPLICATIONS

Rahul Keshav Bhalerao[1], Sheetal A Wadhai[2]

[1]Bachelor Of Computer Engineering, Universal College of Engineering & Research, Pune

[2]Guide, Computer Engineering, Universal College of Engineering & Research, Pune

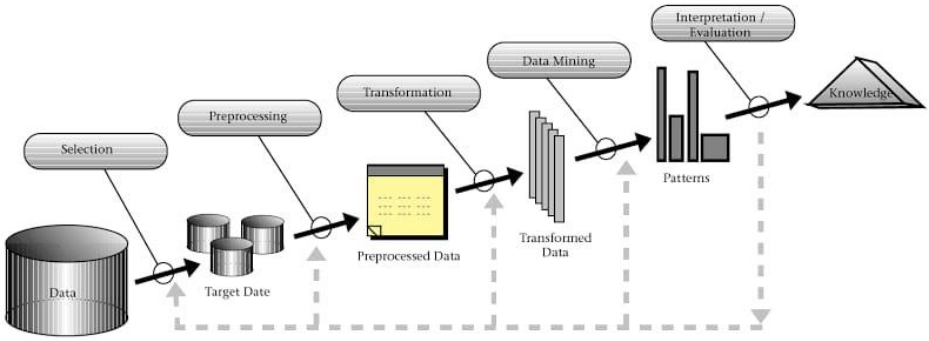
Abstract: Data mining is a technique for extracting meaningful patterns from massive amounts of information. The paper examines a few data mining approaches, algorithms, and some of the corporations that have successfully implemented data mining technologies to better their businesses.

Keywords: Data mining Techniques; Data mining algorithms; Data mining applications.

INTRODUCTION

1. Overview of Data Mining

The development of Information Technology has generated large number of databases and huge data in various areas. The research in databases and information technology has given rise to an approach to store and manipulate this precious data for further decision making. Data mining is a process of extraction of useful information and patterns from huge data. It is also called as knowledge discovery process, knowledge mining from data, knowledge extraction or data /pattern analysis.



Data mining is a logical process for searching through enormous amounts of data to locate meaningful information. The purpose of this technique is to discover previously unknown patterns. Once these patterns have been discovered, they may be used to make specific decisions about how to grow their enterprises. Data exploration begins with the cleaning and transformation of data into a new format, as well as the identification of essential factors and the nature of data relevant to the problem.

2. DATA MINING ALGORITHMS AND TECHNIQUES

Classification, clustering, regression, artificial intelligence, neural networks, association rules, decision trees, genetic algorithms, the nearest Neighbour method, and other algorithms and techniques are used to extract knowledge from databases.

2.1 Classification

The most prevalent data mining technique is classification, which uses a group of pre-classified samples to create a model that can categories the entire population of information. This form of research is particularly well suited to fraud detection and credit risk applications. This method usually employs classification algorithms based on decision trees or neural networks. Learning and classification are both involved in the data classification process. The training data is examined by a classification algorithm in Learning. Classification test results are used to estimate the classification rules' accuracy. The rules can be applied to the new data tuples if the accuracy is satisfactory. This would include entire records of both

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114173

fraudulent and lawful activity evaluated on a record-by-record basis for a fraud detection programmed. These pre-classified samples are used by the classifier-training method to determine the set of parameters required for proper discrimination. The algorithm then converts these values into a classification model.
Types of classification models:

- Classification by decision tree induction
- Bayesian Classification
- Neural Networks
- Support Vector Machines (SVM)
- Classification Based on Associations

2.2 Clustering

Clustering is the classification of items into comparable groups. We can detect dense and sparse regions in object space and discover overall distribution patterns and relationships among data variables using clustering techniques. Classification can be used to distinguish groups or classes of objects, but it is time consuming, therefore clustering can be used as a preprocessing method for attribute subset selection and classification. For example, to categories genes with similar functioning and generate groups of customers based on purchase patterns.

2.3 Neural networks

A neural network is a collection of connected input/output units, each with its own weight. During the learning phase, the network adjusts weights to predict the correct class labels of the input tuples. Neural networks have the remarkable ability to derive meaning from complicated or imprecise data, and they can be used to extract patterns and detect trends that are too complex for humans or other computer techniques to notice. These are ideal for inputs and outputs with continuous values. For example, handwritten character reorganization, training a computer to pronounce English text, and many real-world business problems have been successfully applied in a variety of industries. Neural networks are excellent at detecting patterns or trends in data and are well suited to prediction or forecasting requirements. Back Propagation is a type of neural network.

2.4 Predication

The regression technique can be used to predict outcomes. The relationship between one or more independent variables and dependent variables can be modelled using regression analysis. In data mining, independent variables are already known attributes, and response variables are what we want to predict. Unfortunately, many real-world problems do not lend themselves to simple prediction. Sales volumes, stock prices, and product failure rates, for example, are all difficult to predict because they can be influenced by complex interactions of multiple predictor variables. To forecast future values, more complex techniques (e.g., logistic regression, decision trees, or neural nets) may be required. The same model types are frequently used for regression and classification. For instance, the CART (Classification and Regression Technique)

Different types of regression methods
Regression Linear
Nonlinear Regression Multivariate Linear Regression
Nonlinear Multivariate Regression

2.5 Association Rule

The goal of association and correlation is to find frequent item set findings in large data sets. This type of discovery assists businesses in making decisions such as catalogue design, cross marketing, and customer purchasing behavior analysis. Association Rule algorithms must be capable of producing rules with confidence levels less than one. The number of possible Association Rules for a given dataset, on the other hand, is generally very large, and a large proportion of the rules are usually of little (if any) value.
Association rule types
Rule of multilevel association
Rule of multidimensional association
Rule of quantitative association

3. DATA MINING APPLICATIONS

Data mining is a relatively new technology that has not yet reached its full potential. Despite this, a variety of industries are already utilizing it on a regular basis. Retail stores, hospitals, banks, and insurance companies are examples of these organizations. Many of these organizations combine data mining with statistics, pattern recognition, and other critical

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114173

tools. Data mining can be used to discover patterns and connections that would be difficult to discover otherwise. Many businesses like this technology because it allows them to learn more about their customers and make better marketing decisions. Here is an overview of business problems and solutions discovered through the use of data mining technology.

3.1 FBTO Dutch Insurance Company

- ♦ To reduce direct mail costs.
- ♦ Increase efficiency of marketing campaigns.
- ♦ Increase cross-selling to existing customers, using inbound channels such as the company's sell center and the internet a one-year test of the solution's effectiveness.

3.2 ECtel Ltd., Israel

Fraudulent activity in the telecommunications industry.

3.3Provident Finacial's Home credit Division, United Kingdom

- There is no system in place to detect and prevent fraud Results
- Agent and customer fraud has been reduced in both frequency and magnitude.
- Early detection of fraud saved money.
- Investigators' time was saved, and the prosecution rate was increased.

CONCLUSION

In various corporate fields, data mining is important for detecting trends, forecasting, and knowledge discovery, among other things. Data mining techniques and algorithms, such as classification and clustering, aid in the discovery of patterns that can be used to forecast future business trends. Data mining has a wide range of applications in practically any industry that generates data, which is why it is regarded as one of the most important frontiers in database and information systems, as well as one of the most promising multidisciplinary advances in Information Technology.

REFFERENCE

1. Jiawei Han and Micheline Kamber (2006), Morgan Kauffman, 2nd ed., Data Mining Concepts and Techniques. Dr. Gary Parker, Data Mining: Modules in Emerging Fields, CD-ROM, vol. 7, 2004.
3. Crisp-DM 1.0 Data Mining Step-by-Step Guide (http://www.crisp-dm.org/CRISPWP-0800.pdf).
4. Customer Successes in Your Industry, courtesy of http://www.spss.com/success/?source=homepage&hpzone=nav bar.
5. Last obtained on 15th August 2010 from https://www.allbusiness.com/Technology/computer-software-data-management/ 633425-1.html.
6. http://www.kdnuggets.com