IJDPP International Journal of Development and Public Policy

| e-ISSN: 2792-3991 | www.openaccessjournals.eu | Volume: 1 Issue: 5

Increasing Trend in Accuracy Score for Machine Learning Algorithms

Dr. Rudra Govinda[1], Zenab Waheed[2]

[1]Associate Professor, Management and Production Technologies of Northern Aveiro

[2]Independent Researcher, Delhi-NCR, India

Abstract: Most business applications rely on historical data to predict their future. This process helps them gather important information about their customers and develop effective marketing strategies. Due to the nature of data, the development of various industries depends on it. For instance, education and health services need to analyze and predict the future growth of these sectors. Data mining is a type of statistical process that uses techniques to predict statistical data in various business applications. One of the most widely used techniques is the classification. The paper aims to introduce a comprehensive analysis of the literature on classification algorithms for different business applications. It shows that the most accurate algorithm is the FFBPN. The Random forest algorithm is the most accurate method to classify social networks activities. The Nave Bayes algorithm is also the most accurate method to classify agriculture datasets.

Introduction

Since data collection is the core of business operations, digitalizing it helps gather and analyze critical information about a company. Data analysts can use historical data to predict the future behavior of a business. However, dealing with such a huge amount of data requires a lot of time and effort. Data mining is a technique that uses statistical methods and information technology to find potential sources of useful information for decision-making. It is commonly used to collect large amounts of data. Data mining is a process that uses artificial intelligence to extract valuable information from large amounts of data. This method helps organizations identify potential threats and opportunities, and it can also help minimize the risk of unauthorized disclosure.

Data mining technique known as association is commonly used for extracting patterns from large sets of data. Clustering is a technique that mines large sets of data for useful features. One of the most popular techniques is the decision tree. One of the most challenging parts of implementing a data mining framework is deciding which method to use and when. In this article, we will focus on the classification technique.

The two types of data that are needed for the classification process are training data and testing data. The former are used for training and the latter are used for testing. Data mining techniques are commonly used in business to predict the export capabilities of a company.

In social media applications, sometimes, the link between a social network and another site becomes a missing link. Through data mining, it is possible to analyze soil nutrients to improve the profitability of farming. In the construction industry, it is also used to determine the energy efficiency of multi-family housing complexes. Understanding the causes of crime and how to prevent it can help us improve our lives. Data mining tools are used in the healthcare sector to analyze and predict the infection rate of various diseases. Data mining techniques are used in education to develop learning strategies and predict a student's performance. One more goal is to predict the student's salary after graduation. This can be done by analyzing the previous studies and behavior of the student.

Accuracy of Machine Learning Algorithms

Various classification algorithms are updated in terms of accuracy to address different business applications. Classification is a process utilized for extracting data from various types of data. There are two types of data that need to be classified: training data and testing data. Two sets of text articles are used to classify training and testing data. The algorithms used to classify these data are compared.

The nearest neighbor classifier, K-NN, is usually used for classifying test data. The Nave Bayes classifier and the Centroid classifier are also considered. The Centroid classifier combines the average vector for each class with the data structure to classify new data. It is very fast compared to the K-NN classifier. Data mining techniques were used to identify companies that can export. The K-means method is commonly used to group a sample into three different groups. The GRNN algorithm is used to minimize the error in the model's predictions. Feed forward back procuration neural network is a technique that uses machine learning to study the patterns of output and input behaviour of a set of

ISSN 2792-3991 (online), Published under Volume: 1 Issue: 5 in October -2021

data. Support Vector Machine is a classification technique that works by classifying a set of data according to its similarities. A decision tree is a method that works by asking questions about the classes. Naive Bayes is a method that combines the functions of various data sets into one classification scheme. It was mainly used to classify data sets with a probability concept known as Bayes theorem.

Social media applications are built on the OSN concept. These techniques are used to predict the missing link between two Facebook data sets. There are two datasets with high density, one with high links and one with low density. Random Forest is the best performer among the others in terms of precision, accuracy, and area under the receiver's operating characteristic curve.

On the other hand, the high-density data sets can be predicted with the help of Random Forest or Decision Tree. Analyzing soil nutrients is an important aspect of an agricultural survey. It can help the growers make a large profit. A choice tree is a conceptual framework where the inner hubs explain a test on a characteristic. A decision support model was used to determine the target multi-family housing complex for green remodeling. It was derived by analyzing data mining techniques. The concept of the decision support model was then developed to help people identify the goal of availing home financing for green remodeling.

One of the main goals of police is to prevent the offense against the female. On recent study analyzed the causes of various offenses and the relationships between them. Data mining techniques are used to analyze and forecast offensive statistics.

Most researchers use classification and clustering techniques in their work. In the discussion, the techniques that were used were discussed. A research also proposed a method for extracting data mining techniques for cyber-attacks. These applications are vulnerable to exploitation due to their nature. Data mining techniques can help identify hidden information in these applications. The use of viruses and malicious software is the main reason for the infringement of privacy in cyberspace. It can also lead to the unauthorized access of legitimate users.

One recent research aims to survey the various techniques used in crime investigations. It focuses on the various types of crimes that are carried out in Thailand. Most of the techniques studied failed to detect crime prediction due to the lack of research gaps. Further, these techniques require analysis and data preparation to find the right algorithms. Data mining in the healthcare industry is an important aspect of the research field. This paper reviews the various studies that were published in an attempt to improve the understanding of the removal of health care records.

Data mining tools have been utilized for the detection of various infectious diseases. Some of these include breast cancer, skin diseases, and blood diseases. Another paper presented various techniques that are used in data mining for healthcare research. They show how these techniques can help researchers extract valuable information from massive datasets. Data mining techniques help in the early diagnosis of various medical diseases, which can save a lot of time and money. Among the various medical diseases, cardiovascular is considered the most critical. Data mining techniques are being utilized to treat this condition. One of the recent study aims to find the best data mining techniques that are based on healthcare data. It highlights the various characteristics of the data mining techniques. Results show that the J RIPPER, OneRule, and PART techniques are the best at measuring precision and false-positive rate. The OneR technique is the best for accuracy while the Bayesian Network is the best for sensitivity.

Data mining techniques were used by education researchers to develop effective learning strategies.

A corporative learning model was proposed to group students into learning groups on the web. A comparison of the various algorithms used to predict the performance of students is performed. C4.5 was the most accurate algorithm for predicting the academic success of students. It was also the most reliable method for estimating the correct data correctness and processing time. The C4.5 algorithm has the highest correctness percentage. It is mainly due to its reliability and simple design. The ID3 algorithm has the lowest correctness percentage. Using the C4.5 algorithm, a better learning level can be achieved by selecting individuals to form learning groups.

To improve the rendering of education data and make a decision that helps the student proceed in education. Techniques that have been used include the J48 algorithm, the Support Vector Machine algorithm, the Random Forest algorithm, and the Multilayer Perceptron algorithm. The results of the tests revealed that the Multilayer Perceptron technique performed better than other techniques in terms of accuracy and performance metrics. Another study presents a literature survey on the various techniques used in educational data mining (EDM). The study focused on the prediction of student performance using a set of survey questions. The results were analyzed using a data mining model known as the CCRISP-DM.

Data mining techniques were used to create an incentive for students to study by predicting the future salary of graduates. It can also cause students to drop out of college. This is because of the lack of motivation to continue their studies. This can also affect their grades. This concept can be achieved by modeling the student's salary after graduation

ISSN 2792-3991 (online), Published under Volume: 1 Issue: 5 in October -2021

using the data collected during the study. This model is mainly used for predicting future salary. It utilizes various data mining techniques such as K-Nearest Neighbors, Naive Bayes, and Gaussian models. A WEKA tool was used to analyze the outputs of various data mining techniques. The tool performed better than a KNN technique in predicting the probability of recall and precision. Other techniques that were studied included Naive Bayes, J48, and the multilayer perceptron. A survey was then conducted to see if the model would improve the students' motivation and academic performance.

In another paper, researchers proposed an idea to use technology data mining 11th grade data in Oman. Through data mining techniques, we can help improve the dropout rate of students in 11th grade. This method helps identify the appropriate math for the students. Data mining refers to the process of extracting knowledge from vast amounts of data. This discipline can help educational institutions make informed decisions regarding the curriculum and teaching methods.

The various algorithms used in this study confirm the notion that data mining can predict the performance of students. Data mining techniques were used to extract useful information from academic databases. These algorithms were then used to predict the degree-accomplishments of bachelor's degree students. They present the C4.5 algorithm for data mining. It achieves a better performance in terms of rendering gauge with high precision and minimal class recall. Anoop Kumar and Rahman utilized data mining techniques in education to uncover potential educational data. Data mining techniques can also be used to enhance the learning process by discovering patterns and improving the efficiency of teachers. These are some of the many steps that help students reach their academic goals. They are also used to ensure that their progress is not impeded by certain factors.

This type of analysis can help identify the most effective way to teach and improve the academic performance of slow learners. It can also predict their future performance. Data mining techniques are used to modify raw data to make it easier to use in an educational environment. In another study, we compare the various techniques and algorithms used for data mining. The objective of this paper is to determine which of them can be utilized for educational data mining. Some of the most commonly used terms in educational data mining include association rule mining, clustering, and classification.

Conclusion

Data mining is a process utilized for predicting statistical data in various business applications. This paper presents various classifications algorithms that are commonly used in data mining. Experiments have revealed that the Centroid algorithm is the most accurate method for classifying text documents. The other algorithms that are commonly used to classify various types of datasets are the Random Forest, the Centroid, and the Fatabola Bayes. The OneR algorithm is the most accurate tool to classify cases in the health domain. This paper will help users to improve their knowledge about the domain names. It will also be used as a reference point for analyzing data.

References

1. Arcinas, Myla M., et al. "ROLE OF DATA MINING IN EDUCATION FOR IMPROVING STUDENTS PERFORMANCE FOR SOCIAL CHANGE." *Turkish Journal of Physiotherapy and Rehabilitation* 32.3: 204-226.
2. Wu, Xindong. "Data mining: an AI perspective." *Intelligent Informatics* (2003): 23.
3. Pallathadka, Harikumar, et al. "Impact of machine learning on management, healthcare and agriculture." *Materials Today: Proceedings* (2021).
4. Ganesh, R. Senthil, et al. "Artificial Intelligence Based Smart Facial Expression Recognition Remote Control System." *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 2021.
5. Sajja, Guna Sekhar, et al. "An Investigation on Crop Yield Prediction Using Machine Learning." *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, 2021.
6. Pallathadka, Harikumar, et al. "Applications of artificial intelligence in business management, e-commerce and finance." *Materials Today: Proceedings* (2021).
7. Kavitha, S., et al. "An Internet Of Things For Data Security In Cloud Using Artificial Intelligence." *International Journal of Grid and Distributed Computing* 14.1 (2021): 1257-1275.
8. Khurana, Lokesh, et al. "Speech Recognition with Deep Learning." *Journal of Physics: Conference Series*. Vol. 1854. No. 1. IOP Publishing, 2021.
9. Kshirsagar, Pravin R., et al. "Automation Monitoring With Sensors For Detecting Covid Using Backpropagation Algorithm." *KSII Transactions on Internet and Information Systems (TIIS)* 15.7 (2021): 2414-2433.

ISSN 2792-3991 (online), Published under Volume: 1 Issue: 5 in October -2021