Available online at: http://www.advancedscientificjournal.com
http://www.krishmapublication.com
IJMASRI, Vol. 1, issue 1, pp. 135-138, Apr. -2025
https://doi.org/10.53633/ijmasri

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY ADVANCED SCIENTIFIC RESEARCH AND INNOVATION (IJMASRI)

ISSN: 2582-9130 **IBI IMPACTFACTOR 1.5** **DOI: 10.53633/IJMASRI**

RESEARCH ARTICLE

EDUCATIONAL DATAMINING AND STUDENTS' PERFORMANCE PREDICTION

Ms Sakthipriya B[1] Vennila S[2] and Sathya R[3]

[1]PG & Research Department, Department of Computer Science, St.Ann's College of Arts and Science Tindivanam. Email: priyamuruganscs@gmail.com
[2]Principal & Head of the Department, PG & Research Department of Computer Science, St.Ann's College of Arts and Science Tindivanam
Email: sathiyaat@gmail.com
[3]Assistant Professor & Head, Department of Computer Science, St.Ann's College of Arts and Science Tindivanam. Email: moulikrishna19@gmail.com

Abstract

It is important to examine and analyze academic data, especially for high-achieving students. Educational data mining (EDM) is a research field that uses data mining to find interesting patterns and information in educational organizations. This study also focuses on this topic, especially high-achieving students. This study investigates various theoretical hypotheses that affect student performance in higher education and finds a model that best describes and predicts student performance based on personal and social influence.

Keywords: EDM, data mining, patterns and prediction

Introduction

Educational Data Mining (EDM) is a new form of data mining and knowledge discovery (KDD) in data It focuses on discovering valid patterns that are found to be good and useful Education Information Access, Registration Processes, Education Systems (Moodle, Blackboard etc.) and other methods related to boys as girls are studying in schools, colleges and universities at different levels of education. Researchers in this field aim to find good information that will help schools to manage students better or manage students' learning and presentation and improve their performance. The main aim is to analyse and understand students' learning data, report their learning and develop specific policies, classifications and predictions that will help students in their future education. Classification is the most well-known and

widely used data mining technique used to classify and predict values.

Educational Data Mining (EDM) is not unique, so this research paper uses it to analyse student data collected through research and provide a classification of the collected data to predict and classify student performance in the next semester. The purpose of this study is to determine the relationship between students' personal and social relationships and their learning. This new information can help students and teachers improve their academic performance by identifying students who are likely to under perform early in the semester/term and taking more care to help them succeed. They achieved better results in their studies.

In fact, this research can benefit not only low performers, but also high performers who will benefit from this research by putting in more effort, doing better work and research, and receiving more help and attention from their instructors. Each method or process has its own advantages and disadvantages. Therefore, this article uses various methods to explain and analyse the results of different types of classification. Ultimately, the best result can be selected based on facts and facts.

Data Mining Methodology

The aim of this study is to investigate the relationship between the personal and social well-being of students and their academic performance in the previous semester using survey data. Their performance in the next semester can be predicted. Therefore, we developed a questionnaire that includes various personal, social and educational questions that are pre-processed and converted into nominal data for data mining. The process is to find a relationship between the above factors and student performance. Student performance is measured and reported by grade point average (GPA), which is a real number with a maximum score of 4.0. This study was conducted by a group of students in various schools of Ajman University of Science and Technology (AUST), United Arab Emirates.

A. *Dataset:*

The data used in this study was collected from questionnaires distributed to different students in daily classes and in the form of an online survey using Google Forms. The initial size of the datais 270 records.

B. *Data Exploration:*

In order to understand the dataset at hand, it needs to be analysed statistically and visualized using graphs and charts. This step in data mining is important because it allows researchers and readers to understand the data before they start using more complex data mining processes and algorithms

C. *DataMining Implementation and Results:*

There are many well-known methods for data mining and knowledge discovery (KDD) in data, such as classification, clustering, organization learning, intelligence, etc. The distribution is a frequently used distribution and learn about the mining process. Scientists use and study classification because it is simple and easy to use. In particular, classification in data mining is a process of predicting the class or category of data objects, i.e. which classes of objects are known, based on previous learning from training data. There are many classification techniques in data mining, such as decision trees and naive Bayes Prediction university students at the end of the semester. This approach is recommended because it can provide an overview and understanding of the final results and conclusions, and allows comparison of the results with the study science. Additionally, 10-fold cross validation was used to check and verify the results of the algorithms used and to provide accuracy and precision measurements

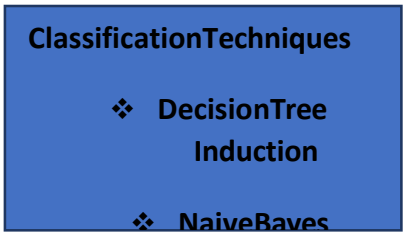

Fig:1

1. Decision Tree Induction:

Decision trees are classification algorithms that generate top-down tree models based on the characteristics of the given data. Decision trees are predictive techniques used to predict, classify or qualify a particular object based on a model created using training data with similar characteristics. The resulting tree structure consists of a root node, internal nodes and leaf (terminal) nodes. The root node is the first node in a decision tree that has no incoming edges but one or more outgoing edges; it has no outgoing edges; The leaf node is the end of a decision tree model and represents the final predicted category (label) of a data item..5 Decision Tree, ID3 Decision Tree, CART Decision Tree and CHAID.

Analysis and Summary:

Their performance and accuracy have been validated. As a final observation, it is worth noting that some algorithms perform better than others on certain datasets, the best accuracy of CART is 40%, which is a higher accuracy (Table1) as desired (pre-set pattern), followed by CHAID andC4.5, which are 34.07% and 35.19% respectively.

DecisionTree	Accuracy
CART	40%
CHAID	34.07%
C4.5	35.19%

Table:1

2. NaiveBayesClassification:

Naive Bayes classification is a simple probabilistic classification technique that assumes that all attributes given in the data are independent of each other, which is why it is called "naive". In this study, the use of classification techniques has been tested and validated in terms of their effectiveness and accuracy. Also, this section presents some ideas to find interesting patterns in Naive Bayes models.

Analysis and Summary:

As a final observation, this section presents the results of the research data obtained using the Naïve Bayes model and more detailed information that can be obtained from the Naive Bayes model in the future.

Conclusion

In this research paper, various data mining tasks were used to develop a robust prediction model that can effectively predict student performance. Based on the information provided, it was found to be good. First, a survey was created for university students and various personal, social and academic data were collected. Secondly, the collected data was pre- processed and searched to make sure that it only fits the data. Third, data usage is the presentation of the available data to build and test a classification model. Finally, the results obtained from the distribution model were expanded to find patterns that expanded in the Naive Bayes model. Four decision tree algorithms were applied, including the Naive Bayes algorithm. In recent studies, we see little evidence that students' performance does not depend solely on their academic performance, but there are many other factors that have equal or greater influence. In conclusion, this study can support and help universities to identify positive results and patterns by conducting regular data mining on student data, which can help universities and students in many ways.

Future Work

Using the same dataset, more data can be processed easily and more algorithms can be used. For now, it would be interesting to use the combined rule to see the detailed rules in the student profile. Also, since the collection of student data in this study is done with the classical sampling method, it is a time- consuming process, it would be better to collect the data as part of the school registration process, i.e. filling out the application. It will be better, the data sets will be larger and universities will be able to see interesting patterns and improve by continuously running this data mining on their students. Their performance is good.

References

1. Baradwaj, B.K. and Pal, S., 2011. Mining Educational Data to Analyse Students' Performance. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011.

2. Ahmed, A.B.E.D. and Elaraby, I.S., 2014. Data Mining: A prediction for Student's Performance
3. Computer Application and Technology, 2(2), pp.43-47.
4. Pandey,U.K and Pal,S., 2011. Data Mining: A prediction of performer or underperformer using classification. (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (2), 2011, 686-690.
5. Bhardwaj, B.K. and Pal, S., 2012. Data Mining: A prediction for performance improvement using classification. (IJCSIS) International Journal of Computer Science and Information Security, Vol. 9, No. 4, April2011.

Using Classification Method. World Journal of

6. Yadav, S.K., Bharadwaj, B. and Pal, S., 2012. DataMining Applications: AComparative Study for Predicting Student's Performance. International JournalofInnovative Technology & Creative Engineering (ISSN: 2045-711), Vol. 1, No.12, December.
7. Yadav, S.K. and Pal, S., 2012. Data mining: Apredictionfor performance improvement of engineering students using classification. World of Computer Science and Information Technology Journal (WCSIT). (ISSN: 2221- 0741), Vol. 2, No. 2, 51-56, 2012

******