Phoenix: International Multidisciplinary Research Journal
Vol 1, No. 2, April-June, 2023
ISSN:2583-6897

FEATURE SELECTION ALGORITHM FOR PREDICTING STUDENTS' ACADEMIC

[1]***Anisha,*** *Master of Computer Application BKIT-Bhalki*
[2]***Prof. Poojarani,*** *Master of Computer Application BKIT-Bhalki*

Abstract -Predicting the performance of pupils is necessary in order to evaluate whether or not there is room for improvement. Evaluations should be done on a regular basis since they not only assist students enhance their performance but also shed light on areas in which they need improvement. Due to the fact that a single institution might have thousands of students, the assessment procedure requires a significant amount of human labour to be completed. In this article, a comparison was made between two different automated approaches to predicting the students' performance using machine learning. Because educational databases include such a vast amount of information, it is becoming more difficult to accurately forecast the performance of pupils. There are primarily two explanations for why this is taking place. To begin, the research that has been done on the many current prediction approaches is not nearly enough to determine which techniques are most suited for forecasting the performance of students. The second reason is that there haven't been enough studies done to determine the elements that influence students' performance in various classes. As a result, in order to raise student accomplishment levels, a comparative research on forecasting student performance via the use of machine learning technologies has been conducted. The primary purpose of this study is to determine which machine learning methods have shown to be the most accurate when used to forecast the performance of pupils. This study also focuses on how the prediction algorithm may be used to determine which characteristics of a student's data are the most relevant to concentrate on.

Keywords— Student performance, Educational Data Mining; Learning Analytics model; FPSO; SVM; KNN; Navie Bayes

1.INTRODUCTION

In places of higher education, one of the most important aspects is the performance of the students. This is due to the fact that an outstanding track record of academic accomplishments is one of the requirements for a university to be considered of a high level [1]. According on the research that came before, there are a variety of different ways to define how well children do. According to Usamah et al. (2013), one may collect information on the performance of students by assessing both the learning assessment and the cocurriculum [2]. On the other hand, the vast majority of research referred to graduation as the indicator of students' level of accomplishment. In most cases, the majority of At the moment, a wide variety of approaches of assessing the performance of pupils are being contrasted. Mining data is quickly becoming one of the most common methods used to evaluate the performance of pupils. In recent years, there has been a widespread use of data mining in the field of education [10]. The approach is known as "educational data mining." The technique of extracting usable information and trends

from a large educational database is known as educational data mining [11]. The beneficial information and trends may be employed in the process of performance forecasting for the pupils. As a consequence of this, it would be of assistance to the educators in developing an efficient method of instruction. In addition to that, teachers are able to keep track of their pupils' accomplishments. It would be possible for students to enhance their learning activities, which would then enable the administration to improve the functioning of the system. As a result, the implementation of data mining methods may be tailored to meet the particular requirements of a variety of organisations. A comprehensive evaluation and comparison is carried out in order to identify and address the issues. The following are the primary goals of this work:

1. To investigate and pinpoint the areas of weakness in the currently available forecasting methodologies.

2. To investigate and determine which factors are taken into account when evaluating the performance of pupils.

3. To do research on the many approaches of performance forecasting that are currently available.

2. Literature survey:

1)A DETERMINING DOMINANT FACTOR FOR STUDENTS PERFORMANCE PREDICTION BY USING DATA MINING CLASSIFICATION ALGORITHMS

AUTHORS: E. Osmanbegović, M. Suljić, and H. Agić

The identification of a representative data set, on the basis of which a classification model will be developed, is the first challenge that must be overcome in order to extract information from data in the context of the educational data mining process. This is the core difficulty. This study shows the research achievements in reducing the dimensionality of data, in the classification issues of predicting student performances using high schools in the Canton of Tuzla as an example. In this article, many algorithms that are used to construct a data mining model for predicting the performances of students based on their personal demographic and social characteristics are presented. These algorithms are used to lower the dimensionality of data, which is a measure of the amount of space occupied by the data. It was discovered that the algorithms Random Forest and J48 are able to produce classification models with an accuracy that is more than 71%.

2) Predicting academic performance

AUTHORS: P. Golding and O. Donaldson

The value that is put on matriculation rises with each successive degree of academic achievement, with the tertiary level assigning the highest amount of significance. In order to

accomplish the job of standardising entry-level criteria at the tertiary level, standardised tests like the SAT, GMAT, and GRE were developed and implemented. Previous study that was carried out at the University of Technology in Jamaica (UTECH) suggested that the process of identifying effective academic performance indicators is still not complete. Within the context of the Bachelor of Science and Information Technology (BSCIT) programme offered by UTECH, this research investigates the connection between students' grade point averages (GPAs) and their ability to meet the program's matriculation criteria in the first year of their coursework. Undergraduate students who graduated from the BSCIT programme in the year 2005 are the focus of this study's evaluation. Specific data was acquired after a survey was administered to the files of each and every BSCIT undergraduate student from the year 2005. According to the data, performance in gateway courses taken in the first year has some amount of importance in terms of predicting success in subsequent years. The results of this research will be very helpful in reorganising the criteria for participation in the programme, so stay tuned for that!

3)Selecting optimal subset of features for student performance model

AUTHORS: H. M. Harb and M. A. Moustafa,

The topic of educational data mining (also known as EDM) is a new one that is quickly expanding as a study area. The fundamental ideas behind data mining are being used in the realm of education with the objective of gleaning relevant information on how students behave while they are engaged in the learning process. The educational data may be classified using techniques such as decision trees, rule mining, and Bayesian networks in order to predict student behaviour, such as how well they would do on an examination. This forecast could be helpful in evaluating the students.

Because the selection of features has such an impact on the degree to which a performance model can accurately predict future outcomes, it is vital to conduct in-depth research on the efficiency of student performance models in relation to feature selection strategies. The primary purpose of this study is to accomplish excellent predictive performance by using a variety of feature selection strategies in order to improve forecast accuracy while using the fewest possible features. The results demonstrate a decrease in the amount of computing time as well as the cost of constructional work during both the training and classification phases of the student performance model.

4)Correlation Based Feature Selection (Cfs) Technique To Predict Student Perfromance

AUTHORS: M. Doshi,

Education data mining is a burgeoning field that helps students mine academic data in order to address a wide range of challenges. Choosing the appropriate academic path is one of the challenges that must be overcome. There are a lot of different considerations that go into deciding whether or not to let a student into an engineering programme. In this study, we

attempted to develop a categorization method that would be of use to students in determining the likelihood of their being accepted into one of the engineering specialisations.We have performed an analysis on the data set that consists of information regarding students' academic and sociodemographic factors. These variables include characteristics such as family pressure, interest, gender, XII marks and CET rank in entrance tests, as well as historical data from the students who came before them. A strategy that helps enhance the prediction accuracy of classifiers is called feature selection. This procedure involves deleting unnecessary and redundant characteristics from the dataset. In this research, we have initially attempted to forecast the important features by using the feature selection attribute methods Chi-square.InfoGain, and GainRatio. Then, we proceeded to apply the rapid correlation base filter on the characteristics that were provided. In the latter stages of categorization, NBTree, MultilayerPerceptron, NaiveBayes, and Instance-based -K closest neighbour are used. As a result of the analysis, the student model exhibited both decreased computing costs and times as well as increased predicted accuracy.

3. OBJECTIVE:

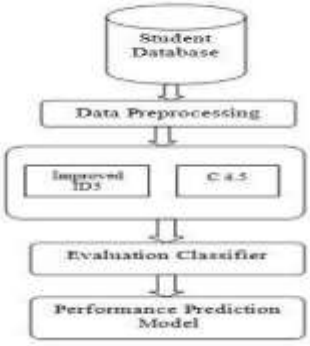
4. SYSTEM ANALYSIS:

Existing System:

As a result of the vast number of data included in educational databases, accurate forecasting of student performance is becoming more difficult. This is because of the two primary causes. To begin, the research that has been done on the many current prediction techniques has not been adequate to determine which approaches are the most appropriate for forecasting the performance of students who are enrolled in the university. The second reason is that there haven't been enough studies done to determine the elements that influence students' performance in different classes. As a result, it is essential to conduct literature evaluations on the topic of employing data techniques to forecast student performance.

Proposed System

4. ARCHITECTURE



The database of students at the university, which was established so that their grades could be analysed and it could be determined which academic departments they would be accepted into. Experimentation will be carried out using the information gained during training. The information about the training includes the names and grades of the students.
Data pre-processing: data pre-processing is a data processing method that is used to rework the data into a format that is both beneficial and inexpensive. information pre-processing is also known as information pre-processing.

a. Data purification:
the information purification is the process of analysing the raw data so that it may be reborn as knowledge. In this process, good data is preserved while bad data is removed. The user is given the ability to search for information that may be erroneous or incomplete as a result of this. Incorrect records from a database or dataset are identified as incomplete, less trustworthy, incorrect, or non-relevant components of the information; as a result, the information is restored, remodelled, and any unclean or crude information is removed.

b. Data Integration: information integration may be a strategy for information pre-processing that combines information from a variety of disparate information sources into a coherent information store and offers a unified reading of the information.
Integration of information may refer to either the process of combining information sets, files, or information cubes, or it can represent the end result of such a process.

c. Data Transformation: This step is taken to rework the info into acceptable forms appropriate for the mining method. This involves the subsequent ways:
1. Normalization: it's done to scale the info values during such vary (-1.0 to 1.0 or 0.0 to 1.0)
2. Attribute Selection: From the given set of attributes, new attributes square measure generated or created
3. Concept Hierarchy Generation: Here attributes square measure born-again from level to higher level within the hierarchy.

d. Data Reduction: Data processing is a method that is used to handle an excessive amount of information. One of the steps in this method is data reduction. When dealing with a large number of information, it is difficult to do analysis. This study often makes use of information reduction in an effort to induce or remove something. Its goals are to increase storage capacity while simultaneously lowering the costs of information storage and processing.

5. MODULES:

6. Results and Analysis:

A. Indices of Operational Effectiveness

There are a variety of efficiency indicators that may be used in order to determine how effective the methods to machine learning are. In order to evaluate how effective something is, this study makes use of the Detection Accuracy.

Accuracy of the Detection

Detection Accuracy is a measuring method that determines the degree of closeness of measurement between the original findings and the properly predicted outcomes. It does this by comparing the original results to the correctly predicted results.

Accuracy= (TP+TN)/(TP+FP+TN+FN)

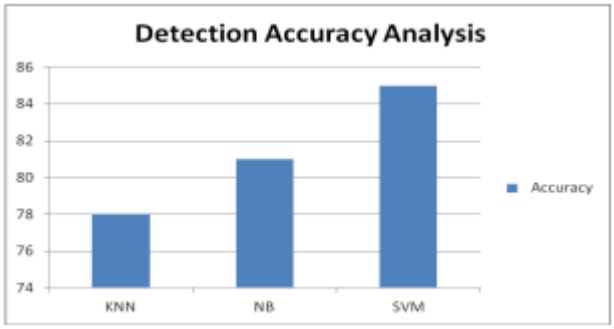
B. The Results of Experiments Experiment: An Accuracy Analysis of Different Approaches to Machine Learning

In this experiment, this work will evaluate the contribution of each classifier strategy that is used in the work by determining how well each approach performs. In an ideal world, a high Accuracy value is expected to be present in a machine learning scheme that is considered to be good. The results of the accuracy analysis performed by FPSO using SVM are shown in Table 1.

Table 1: Detection Accuracy Analysis of Machine Learning Approaches

Metrics	Accuracy
KNN	78
NB	81
SVM	85

As observed from Table 1, the Accuracy of the FPSO with SVM in range 85, which is excellent when compared to other approaches. Therefore, the FPSO with SVM classifier is thought to be the most effective method for doing sentiment analysis. The Detection Accuracy of several classifier techniques was shown in Fig.1.



Conclusion:

Predicting the performance of students is primarily important for the purpose of assisting teachers and students in enhancing their respective learning and teaching processes. In this study, a number of different techniques to machine learning and feature selection were evaluated in terms of their ability to predict students' performance using a number of different analytical methodologies. The purpose of student classification is to make predictions about their placement in various class groups, such as high, medium, and low. The accuracy and

precision of the outcomes of using different feature selection methodologies in conjunction with machine learning were analysed and contrasted. When compared to other classifiers, it was discovered and discovered that classification carried out by SVM with FPSO is more efficient when compared to other classifiers as evidenced in the accuracy and precision. According to the findings, the SVM approach combined with the FPSO technique is superior to other techniques in terms of its ability to accurately anticipate the performance of students.

ACKNOWLEDGEMENT

The heading should be treated as a 3rd level heading and should not be assigned a number.

REFERENCES

[1] M. of Education Malaysia, National higher education strategic plan (2015).
[2] U. bin Mat, N. Buniyamin, P. M. Arsad, R. Kassim, An overview of using academic analytics to predict and improve students' achievement: A proposed proactive intelligent intervention, in: Engineering Education (ICEED), 2013 IEEE 5th Conference on, IEEE, 2013, pp. 126–130.
[3] I. No, S. Anam, and S. Gupta, "A research Review on Comparative Analysis of Data Mining Tools, Techniques and Parameters," Int. J. Adv. Res. i, vol. 8, no. 7, pp. 523– 529, 2017.
[4] J. Xu, K. H. Moon, and M. van der Schaar, "A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs," IEEE J. Sel. Top. Signal Process., vol. 11, no. 5, pp. 742–753, 2017.
[5] I. ĐurđevićBabić, "Machine learning methods in predicting the student academic motivation," Croat. Oper. Res. Rev., vol. 8, no. 2, pp. 443–461, 2017.
[6] A. Kumar, J. Naughton, and J. M. Patel, "Learning Generalized Linear Models Over Normalized Data," Proc. 2015 ACM SIGMOD Int. Conf. Manag. Data - SIGMOD '15, pp. 1969–1984, 2015.
[7] M. Pandey and V. K. Sharma, "A Decision Tree Algorithm Pertaining to the Student Performance Analysis and Prediction," Int. J. Comput. Appl., vol. 61, no. 13, pp. 2– 6, 2013.
[8] M. A. Al-Barrak and M. Al-Razgan, "Predicting Students Final GPA Using Decision Trees: A Case Study," Int. J. Inf. Educ. Technol., vol. 6, no. 7, pp. 528–533, 2016. [9] C. Romero and S. Ventura, "Data mining in education," Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 3, no. 1, pp. 12–27, 2013.
[10] C. Romero, S. Ventura, Educational data mining: A review of the state of the art, Trans. Sys. Man Cyber Part C 40 (6) (2010) 601–618.
[11] D. M. D. Angeline, Association rule generation for student performance analysis using apriori algorithm, The SIJ Transactions on Computer Science Engineering & its Applications (CSEA) 1 (1) (2013) p12–16.