WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Student Scholarship Prediction using Machine Learning Algorithms

Darshana Chaudhari[1], Rajashri Khadke[2], Smita Sarode[3], Sakshi Mahajan[4], Leena R. Waghulde[5]*

[1]*Student,Department of Computer Engineering*
cdarshana442000@gmail.com
[2]*Student,Department of Computer Engineering*
rajashrikhadke00@gmail.com
[3]*Student,Department of Computer Engineering*
smitasarode1317@gmail.com
[4]*Student,Department of Computer Engineering*
sakshimaha2126@gmail.com
[5]*Professor,Department of Computer Engineering*
waghuldeleena.2008@gmail.com

ABSTRACT

Today, predictive analytics applications became an urgent desire in higher educational institutions. Predictive analytics used advanced analytics that encompasses machine learning implementation to derive high-quality performance and meaningful information for all education levels. Mostly know that student grade is one of the key performance indicators that can help educators monitor their academic performance. During the past decade, researchers have proposed many variants of machine learning techniques in education domains. However, there are severe challenges in handling imbalanced datasets for enhancing the performance of predicting student grades. Therefore, presents a comprehensive analysis of machine learning techniques to predict the final student grades in the first semester courses by improving the performance of predictive accuracy. Two modules will be highlighted. First, we compare the accuracy performance of six well-known machine learning techniques namely Naïve Bayes (NB), C5.0 real student's course grade dataset. Second, we proposed a multiclass prediction model to reduce the overfitting and misclassification results caused by imbalanced multi - classification based on oversampling Synthetic Minority Oversampling Technique with features selection methods. This proposed model indicates the comparable and promising results that can enhance the prediction performance model for imbalanced multi-classification for student grade prediction multi- classification for student grade prediction

Keywords: Machine Learning, Scholarship, Grade Point Average, Prediction.

Introduction

Graded point average (GPA) is a commonly used indicator of academic performance. Many universities set a minimum GPA that should be maintained. Therefore, GPA still remains the most common factor used by the academic planners to evaluate progression in an academic environment. Many factors could act as barriers to student attaining and maintaining a high GPA that reflects their overall academic performance, during their tenure in university. These factors could be targeted by the faculty members in developing strategies to improve student learning and improve their academic performance by way of monitoring the progression of their performance. With the help of clustering algorithm and decision tree of data mining technique it is possible to discover the key characteristics for future prediction. Data clustering is a process of extracting previously unknown, valid, positional useful and hidden patterns from large data sets. The amount of data stored in educational databases is increasing rapidly. Clustering technique is most widely used technique for future prediction. The main goal of clustering is to partition students into homogeneous groups according to their characteristics and abilities. These applications can help both instructor and student to enhance the education quality. This study makes use of cluster analysis to segment students into groups according to their characteristics. Decision tree analysis is a popular data mining technique that can be used to explain different variables like attendance ratio and grade ratio. Clustering is one of the basic techniques often used in analysing data sets. This study makes use of cluster analysis to segment students in to groups according to their characteristics and use decision tree for making meaningful decision for the students.

Literature Review

This consists of literature survey to prediction of scholarship by using Machine Learning and Data Mining technique.

Suma, V., and ShavigeMalleshwara Hills et al.[1] had used decision trees to evaluate student performance. Here an educational data set is considered and entropy and information gain of all the attributes present in the dataset is calculated. The attribute which consists of the highest information gain is considered as the root node of the tree. This classification algorithm is used to identify students with poor performance.

Mitra, Ayushi et al[2]. had proposed a scrum methodology to track the performance of the student in web-based education .In this methodology, the teacher describes the learning objectives and that teacher is responsible for monitoring the progress. Later evaluation of all the members is done.

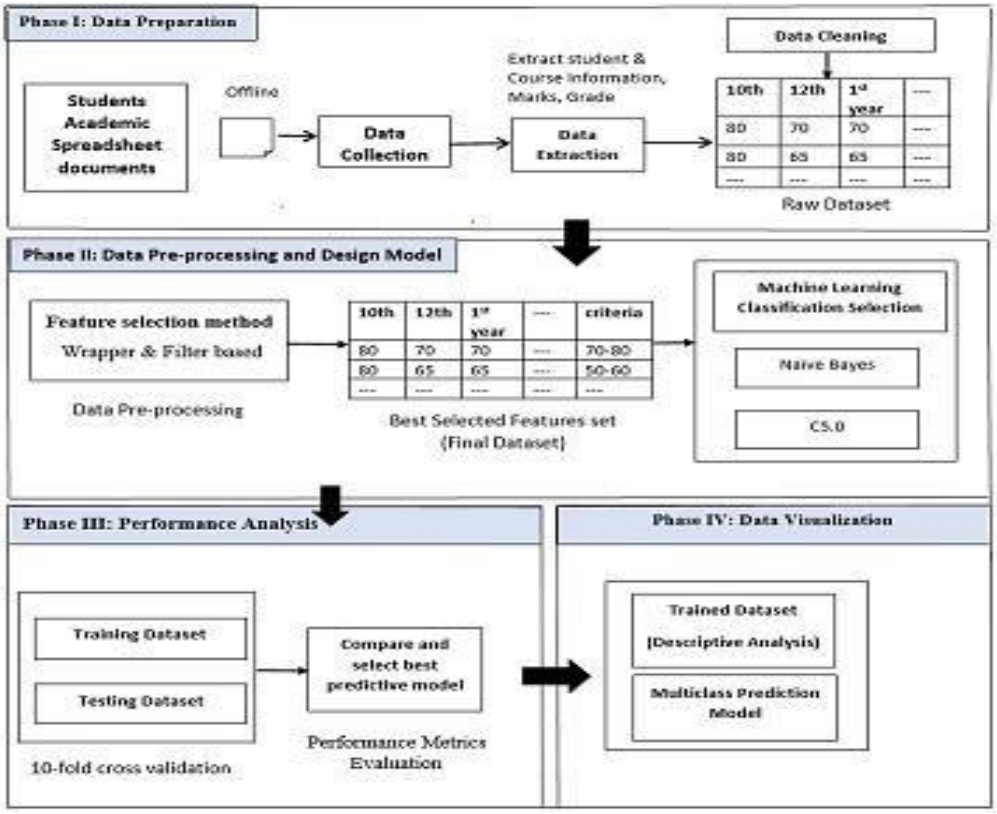
Madhav S. Vyas, et al. [3] made use of a decision tree model for academic performance prediction. The continuous values were converted to discrete values and the null values eliminated in the collection and pre-processing phase.

Okfalisa, Ratik a fitriani, YelliVitriana, et al,[5] describe among all technologies, the researcher used to analysis the data such as scholarship recipient prediction is data mining, Author describes the two methods which was used to predict the output such as k nearest neighbour‟s (KNN) and linear regression algorithms. This study compares both methods in solving the scholarship recipient problem. Here the Author uses key parameters such as semester attendance, Grade point average, statement letter of active student, Family Card, Identity Card, Study Result Card.

Angela R. Bielefeldt [6] in his paper explains that differences between the civilization of engineering faculties when compared to scholarship of teaching and learning (SOTL) in engineering sector maintain themselves with all characteristics of faculty who involved in their activities. SOTL compared overall US engineering faculty based on a assistant professors percentage, a full professors percentage, womens percentage, employees percentage who worked at Baccalaureate and Masters institutions.

Architecture of Student Scholarship prediction

Architecture defines the oral presentation of naïve bayes for student Scholarship prediction using machine learning algorithm.



Working of the Student Scholarship Prediction

Here figure shows the System Architecture of the system. This architecture shows that what input will provide by user and then what kind of output they will get in return as a result from system. Student: Student module consists of attributes of students which would help to generate performance of students and provide scholarships to students. Main attributes that would be considered are numerical parameters like 10^{th} marks, 12^{th} marks, 1^{st} year,2^{nd} year ,3^{rd} year marks. Data Collection: The raw data used in this study will be collected from a reputed engineering college. The data included attributes like 10, 12, 1^{st} year, 2^{nd} year, 3^{rd} year, diploma, gap, atkt of the student"s data variables. The data would range from 150-200 records, out of which 70% would be used for training and 30% would be used for testing. Preprocessing data and Feature Extraction: The data pre-processing will be carefully done to avoid incomplete records. The fields selected for the model include: 10^{th} marks, 12^{th} marks, 1^{st} year marks, 2^{nd} year marks, 3^{rd} year marks, diploma marks, gap, atkt and Result. The instrument and variable selected for this work were based on previous year student result data. The reliability and validity of the instruments used construe with the theories posited. Classification: The tasks involve machine learning and classification algorithms; hence in the research work a two-layered classifier system was designed to achieve the objective of the work. These classifiers have been selected because of their performances in various domains. They have both been successfully applied to a variety of real world classification tasks in industry, business, science and education with good performances. The Naïve bayes is known for its predictive accuracy, ability and aptitude to learn and remember.

Algorithms

Naïve Bayes:

Step 1: Convert the data set into a frequency table

Step 2: Create Likelihood table by finding the probabilities like Overcast

for example - probability = 0.29 and probability of playing is 0.64.

Step 3: Now, use Naive Bayesian equation to calculate the posterior

probability for each class. The class with the highest posterior probability is the outcome of prediction.

C5.0 Classifier

Algorithm for Experimental Model

Input: dataset.

Output: classified output.

1. Take a data set as input.
2. If that set has more features then apply the feature selection technique (PCA) as pre-processing technique
3. Apply parallelism from step 4 to step 6.
4. Evaluate the entropy value and information gain ratio of all three entropies (Shannon, havrda and Charvat's entropy and quadratic entropy).
5. Construct the models separately using C5.0 algorithm based on various entropies.
6. Find the accuracy and execution time of each model and store the value in array.
7. Find a model that has maximum Accuracy.
8. If two have maximum accuracy then
9. Find a minimum execution time of the model that has maximum accuracy.
10. Classify by that model which has minimum execution time.
11. Else classification done by the model which has maximum accuracy.
12. End

Advantages

1. Easily predicting large volume of data within less amount of time.
2. Classification of scholarship on different levels.
3. High accuracy with less time consumption.

Disadvantages

1. Speed may decrease if number of records are increased tremendously.
2. Accuracy may fluctuate as per large number of records used for testing dataset.

Conclusion

By using ML and DM techniques the prediction of scholarship much easier and effective also. While predicting the data it need to look upon many parameters. ML and DM techniques unable to work without training data. For collecting of datasets makes more difficult. It cannot be able to collect the best algorithm suitable for prediction. because prediction is based on training data values: By using the ML and DM methods, the instructions can easily track on the scholarships and the problems faced by the recipients in their studies**.**

References

1 Suma, V., and ShavigeMalleshwara Hills. &quot, "Student Scholarship Prediction using Machine Learning Algorithms". *Journal of Soft Computing Paradigm (JSCP) 2*, 2020 101-110.

2 Mitra, Ayushi. "Sentiment Analysis Using Machine Learning Approaches (Lexicon based on movie review dataset)." *Journal of Ubiquitous Computing and Communication Technologies (UCCT) 2*, 2020 145-152.

3 Madhav S. Vyas and ReshmaGulwani. "Predicting Student"s Performance using CART approach in Data Science*", 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, India, 2017, pp.

4 Huda Al-Shehri, Amani Al-Qarni, Leena Al-Saati, ArwaBatoaq, Haifa Badukhen, SalehAlrashed, Jamal Alhiyafi and Sunday O. Olatunji. "Student Performance Prediction Using Support Vector Machine and K-Nearest Neighbors",*2017 IEEE 30th Canadian on electrical and computer engineering(CCECE),Canada 2017*

5 Okfalisa, Ratik a fitriani, YelliVitriana, " The Comparison of Linear Regression Method and k- Nearest Neighbors in Scholarship Recipient" IEEE, 2018.

6 Angela R. Bielefeldt, "Characteristics of Engineering Faculty Engaged in the Scholarship of Teaching and Learning" IEEE, 2015.