ISSN: 2096-3246
Volume 52, Issue 02, April, 2020

https://doi.org/10.5281/zenodo.12707063

A Study on Machine Learning Based Student Scholarship Prediction

DUDEKULA RESHMA, SREEKANTAM VASUDHA, K BALAJI SUNIL CHANDRA

Assistant Professor [1,2,3]

reshma.dudekula1993@gmail.com, vasudhasvit87@gmail.com, hod.cse@svitatp.ac.in
Department of Computer Science and Engineering, Sri Venkateswara Institute of Technology, N.H 44, Hampapuram, Rapthadu, Anantapuramu, Andhra Pradesh 515722

KEYWORDS:

Machine learning, scholarship, GPA, prediction, Naive Bayes (NB), and the C5.0 dataset are all relevant terms.

ABSTRACT

In today's world, top-tier educational institutions are increasingly opting for predictive analytics tools. In order to get implementation insights, generate high-quality performance, and create relevant records for all areas of education, predictive analytics applied system-covering super-analytics. One of the most important metrics for gauging a teacher's effectiveness in the classroom is the grade they get. Researchers have offered a plethora of different forms of mechanical knowledge acquisition about domain name techniques for instructional objectives throughout the last decade. Improving performance via grade prediction presents unique challenges when dealing with imbalanced data sets. Consequently, it offers a comprehensive evaluation of machine learning methods for enhancing the prediction accuracy of first-semester course grading guidelines. It is possible to emphasise two modules. We start by checking how well six popular tool-learning methods—including Naive Bayes (NB) and the C5.0 dataset—performed on the real student path class. A multi-class prediction model based on over-sampling from the artificial minority over-sampling approach with power-selection techniques was developed as a second measure to mitigate the effects of overfitting and misclassification caused by several imbalanced classes. This revised version shows how the overall performance version of the student score prediction may improve upon the imbalanced multi-category or multi-class student score prediction with similar and promising consequences.

ISSN: 2096-3246
Volume 52, Issue 02, April, 2020
https://doi.org/10.5281/zenodo.12707063

I. INTRODUCTION

A student's life is greatly enhanced by obtaining a higher degree. Analysis of every thing is the main emphasis of higher education institutions. as a result of individual engagement. Marketing, healthcare, online development, fraud detection, engineering, and many more domains make use of data mining methods. The primary Find previously unknown information is the goal of data mining.dm offers a wide range of analytic techniques, such as k-means, decision trees, regression, time series, neural networks, classification, and many more. Analysis of educational system participants is directly aided by data mining's application to the education system. Pupils also suggest a plethora of things to do and accomplish [1]. Data mining may also reveal how students apply course content. Teachers are able to evaluate their pupils' progress in a classroom setting. One popular way to measure academic success is by looking at the grade point average (GPA). Maintaining a certain minimum grade point average is required by several schools. Consequently, academic planners still rely on GPA as the primary metric to measure success in the classroom. Achieving and maintaining a high GPA, which is indicative of a student's overall academic success, throughout their time at university may be challenging for many reasons. Teachers might use these as targets when devising plans to raise students' learning and academic performance, and they could track their progress to see how well those plans worked. Implementing a clustering method and making a employing a data mining approach, it is feasible to identify the crucial traits for future forecasting. Data clustering is a way to find hidden patterns in big datasets that were previously undiscovered. These patterns might be valid, positionally helpful, or both. There has been a meteoric rise in the quantity of data kept in academic databases. When looking to the future, the clustering method is by far the most popular choice. Clustering primarily aims to divide students into similar groups based on their traits and skills [2]. The use of these apps in the classroom has the potential to benefit students and teachers alike. Using cluster analysis, this research divides pupils into subgroups defined by shared traits. A number of factors, such as the attendance ratio and the grade ratio, may be explained using decision tree analysis, a well-liked data mining approach. When evaluating data sets, clustering is a common and fundamental strategy. This research use cluster analysis to categorise students into groups based on their traits, and then uses decision trees to help students make informed choices [3].

I. LITERATURE SURVEY

This includes everything from a literature review to academic forecasting utilising ML and DM methods. One study that used decision trees to assess student performance was that of Suma, V., and ShavigeMalleshwara Hills et al. [4]. The entropy and information gain of each attribute in an educational dataset are computed here. The node at the very top of the tree represents the attribute with the most information gain. Students who aren't doing well are found using this classifying system. The scrum technique was suggested by Mitra, Ayushi et al. [5] as a means of monitoring students' progress in online courses. The approach relies on the instructor outlining the goals of the lesson and then keeping tabs on how well students are meeting those goals. All of the members are evaluated thereafter. In order to forecast students' academic success, Madhav S. Vyas et al. [6] used a decision tree model. While collecting and pre-processing, the continuous data were transformed into discrete ones and the null values were removed. Among the many technologies utilised for data analysis, such as scholarship recipient prediction, Okfalisa, Ratik, Fitriani, YelliVitriana, et al. [7] mention data mining. The author goes on to discuss two approaches that were used for output prediction, namely k-nearest neighbours (KNN) and linear regression.

regression algorithms. The scholarship recipient dilemma is addressed in this research by comparing the two approaches. Key characteristics used by the author here include semester attendance, grade point average, active student statement letters, family and identity cards, and study result cards. In his study, Angela R. Bielefeldt [8] illustrates how the distinctions between engineering faculty civilization and engineering faculty scholarship of teaching and learning (SOTL) persist throughout all faculty members' activities. SOTL found the proportion of assistant professors, full professors, women, and workers at baccalaureate and master's degree programmes in

ISSN: 2096-3246
Volume 52, Issue 02, April, 2020

https://doi.org/10.5281/zenodo.12707063

the United States' engineering faculty. Abdulrashid Qasem A. [9] -"Mining student data using decision tree" is the title of the study. To aid in the administration of higher education, they use a data mining approach to assess how well students are doing in their classes.A lot of things influence how well it works. To construct a trustworthy classification model, they apply a classification approach; specifically, they use the CRISP-DM (cross-industry standard procedure for data mining). The five-step process that these approaches follow is as follows: ascertaining the pertinent aspects of the issue, I will prepare the data, Developing the categorization model, testing it, and then making predictions about the future. After correctly formatted data was gathered in a table, a classification model was constructed using the decision tree approach. Lots of regulations were put into place. When using the WEKA toolbox, Various categorization approaches, such as ID3, C4.5, and naïve, were used. The results table included both Bayes and accuracy. Kalpana J.K. Jothi The use of data mining tools for the purpose of analysing pupils' intellectual performanceSchool level predictions (elementary, secondary, and tertiary) are the main topic of this article. We employ clustering methods like density-based clustering and centroid-based distribution-based clustering. Research at Villupuram College provided the data. Their strategy for improving pupils' efficiency

II. MACHINELEARNINGMETHODOLOGIES

It is the branch of science that works with the system in such that they automatically learn. It means that recognizing and understanding the input data and moving decision on the support data.

The name machine learning was come in 1959 by Arthur Samuel. They evolved rom the study of pattern, AI, computational theory. Machine learning constructs the algorithm that can learn and make predictions. Machine learning closely related to statistics which help in prediction. It is very difficult to take the division for their problem and algorithm is developed. Some of the Machine learning algorithms are k-nearest neighbour,Decision tree, Support Vector Machine (SVM), Artificial Neural Network, ID3, Naïve Bayes etc.

A. Decision Tree

Decision tree algorithm uses a tree-like graph structure having both classifier and regression. The structure contains root node which connects to subnodes, through links which connects to leaf nodes. Each node represents attributes and each leafnodes ends with an outcome or target result. Decision tree rules can be constructed from a training data set with help of many algorithms such as ID3, classifier and regression tree (CART).

ISSN: 2096-3246
Volume 52, Issue 02, April, 2020
https://doi.org/10.5281/zenodo.12707063

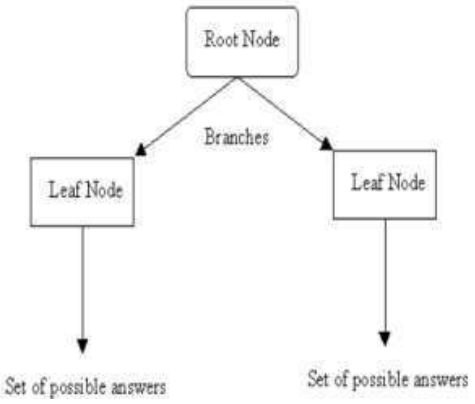


Fig.1 Decision tree

B. K-Nearest Neighbor Algorithm

It is a simple ML algorithm which comes under a division of a non-parametric model. The main aim of the algorithm is topredict an outcome value from input data by "Close" input output pairs that are

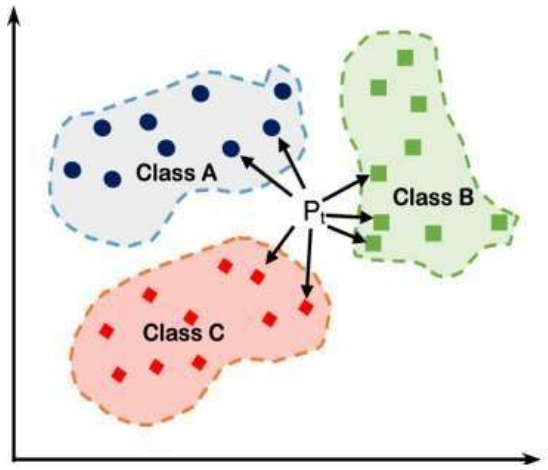


Fig.2 K-Nearest Neighbor Algorithm

C. Support Vector Machine (SVM)

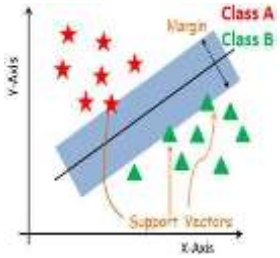
This algorithm widely known as Kernelmethods framework. This algorithm workson basis of mapping a low dimensionalinput attribute space and then it tries tosolve. These Supervised Vector Machine vectors are used to predict the parameters of the hyperplane. By considering the hyperplane, SVM algorithm used to solve regression kind of problems. SVM providebetter performance based on accuracy and training speed rather than K-NN andArtificial Neural Network.

ISSN: 2096-3246
Volume 52, Issue 02, April, 2020

https://doi.org/10.5281/zenodo.12707063

D. Artificial Neural Network

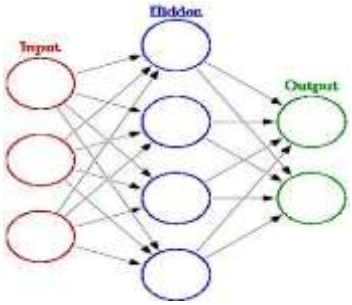
A neural network that is both artificial and based on the principles of biological brain networks. Artificial neurons are the building blocks of artificial neural networks (ANNs). Neurons are the fundamental building blocks of ANN. Applying various attributes and characteristics of neurons in a network to get a desired output is the primary goal of this approach. The input, hidden, and output layers of neurons make up the fundamental architecture of an ANN. Perception with several layers is another name for this arrangement. The input layer's job is to take a vector as input. After that, each neuron in the hidden layer received an input vector from the input layer. The output layer yields a vector of outputs. One major benefit of this algorithm is how quickly it can transition from training to testing.

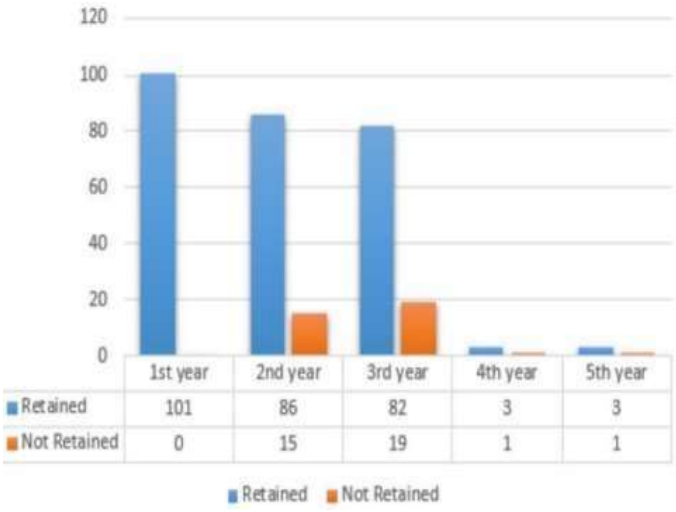


III. RESULT ANALYSIS

Above study describes so many numbers of Data Mining and Machine Learning Methods was used to predicting the scholarship of students.

below figure 5 represents the analytical data of Scholarship Retention in five years by students. In 1st year all students are eligible to maintain their scholarship. In2nd year, many numbers of students were not able to maintain it. By the end of 5th year only 3 students retained their scholarship





ISSN: 2096-3246
Volume 52, Issue 02, April, 2020
https://doi.org/10.5281/zenodo.12707063

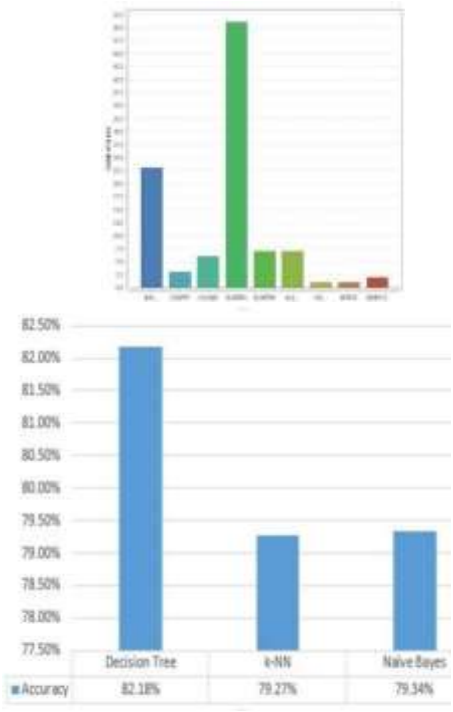


Fig.5 Retention of Scholarship of 5 years

Fig.6 The accuracy of algorithms in processing data

Figure 6 represents result of the accuracyof three algorithms. By using of theDecision tree algorithm gives an accuracy of 82.18%, k-NN gives the 79.27% andlastly the Naïve Bayes which gives accuracy as 79.34%. It means Decision tree has huge accuracy rather than k-NN and Naive Bayes.

Fig.7 Analytical study data of all subjects

IV. CONCLUSION

In this study, we take a look at how grant forecasting makes use of data mining and machine learning. The use of ML and DM methods has greatly improved the accuracy and efficiency of scientific prediction. We need to consider a number of factors while making factual predictions. Without our educational data, ML and DM tactics will not function. As time goes on, gathering data sets becomes increasingly difficult. Finding an appropriate, high-quality algorithm for prediction could be challenging. Instructions may be easily constructed to address the academic challenges faced by grantees and beneficiaries via the use of ML and DM approaches, since the prediction is grounded in the values of school information.

REFERENCES

1. "Student Scholarship Prediction using Machine Learning Algorithms" by V. Suma and T. Shavigemalleshwara Hills. Chapters 101–110 of the 2020 issue of the Journal of the Soft Computing Paradigm (JSCP) are accessible online.

ISSN: 2096-3246
Volume 52, Issue 02, April, 2020

https://doi.org/10.5281/zenodo.12707063

Second, Ayushi Mitra. Journal of Ubiquitous Computing and Communication Technologies (UCCT) 2, 2020 145-152. "Sentiment Analysis Using Machine Learning Approaches (Lexicon based on movie review dataset)."
The authors of the 2017 International Conference on Electronics, Communication, and Aerospace Technology (ICECA) in India, Madhav S. Vyas and Reshma Gulwani, "Predicting Student‟s Performance using CART method in Data Science," were present at the conference.
The following individuals are listed: Sunday O. Olatunji, Arwa Bathoaq, Haifa Badukhen, Saleh Alrashi, Jamal Alhiyafi, Amani Al-Qarni, Leena Al-Saati, and Huda Al-Shehri. The 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE) in Canada included the paper "Student Performance Prediction Using Support Vector Machine and K-Nearest Neighbours."
5. "The Comparison of Linear Regression Method and k-Nearest Neighbours in Scholarship Recipient" by Ratik Okfalisa and Yelli Vitriana Fitriani In 2018, the IEEE published.
6. "Identifying Engineering Faculty Members Who Are Active in the Scholarship of Teaching and Learning" by Angela R. Bielefeldt 2015, IEEE.
7. "The Comparison of Linear Regression Method and k-Nearest Neighbours in Scholarship Recipient" by Okfalisa, Ratik and Fitriani, Yelli Vitriana In 2018, the IEEE published.
8. "A Comparative Analysis of Techniques for Predicting Academic Performance," IEEE, 2007, by N. T. Nghe.
2.
Jarolyn (9th) "Model Development Of Students Scholarship Status at First at Asia Institute of Technology and Humanities (FAITH)" (Joy B. Labayne, Lester L. Mercado, and Jheanel Espiritu Estrada) In 2018, the IEEE published...
Ten. Shavige Malleshwara Hills and V. Suma. Journal of Soft Computing Paradigm 2, no. 02 (2020): 101-110. "Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics."
11. Senthil Kumar. "A Marketing Decision Support System Built on Data Mining and a Hybrid Machine Learning Algorithm." International Journal of Artificial Intelligence 2, no. 3 (2020): 185-193.