

A Study on Machine Learning Technique for Student Scholarship Forecasting

¹TATIKONDA SHIVA KUMAR, ²M. NARESH

¹PG Scholar, Dept. of MCA, Newton's Institute of Engineering, Guntur, (A.P)

²Associate professor, Dept. of CSE, Newton's Institute of Engineering, Guntur, (A.P)

Abstract: This review study employs Machine Learning and Data Mining to make predictions about scholarly output based on a systematic examination of the relevant literature. In addition, there is a brief explanation of the ML/DM techniques used in the study. It also emphasises the significance of data sets in ML/DM approaches. Machine learning is now the most talked about topic in the IT sector. Today, Machine Learning and Data Mining have emerged as a formidable tool with broad applications in areas as diverse as information technology, education, and business. All of this method is directed at the many ML/DM algorithm varieties. Algorithms like Naive Bayes, Decision Tree, and k-NN provide more precise results in identifying the continuation of each student's scholarship. In conclusion, the suggested model will produce a list of applicants deserving of a scholarship, and the correctness of each approach utilised to get this result has been discussed.

Keywords: Data mining, Machine Learning, Scholarship, Grade point average, recipients, prediction.

I INTRODUCTION

This study discusses several Machine Learning and Data Mining Techniques for predicting academic performance and other outcomes. Methods for Machine Learning and Data Mining are outlined, along with some of the challenges these fields provide. Both ML and DM approaches, as well as their benefits and drawbacks, are examined. This also proved

to be the greatest technique for predicting academic success.

Scholarships are cash awards given to deserving students to help them finish their degrees. Scholarships are often funded by either the government or a non-profit organisation. When kids are given accolades for their efforts, it is an inspiring moment. They get the confidence to set new objectives.

Scholarships may be divided into many categories, the most frequent of which are merit-based, student-specific, and career-specific.

Scholarships are awarded to students via a variety of programmes with varying requirements and benefits.

Most often, a student's Grade Point Average (GPA) is used in the selection process for academic scholarships. Many factors, including academic achievement, extracurricular activities, and community service, go into determining which athletes get athletic scholarships. This study examined the use of ML and DM to automatically predict which students would be eligible for scholarships based on criteria like GPA, SAT/ACT scores, family income, language proficiency, and other factors.

II LITERATURE SURVEY

Scholarships, as Thalia an Agnos and Eva shirring (2018) [1] explain, are crucial in helping students afford their education and boosting their motivation to succeed academically. However, students may lose their scholarship opportunities due to factors like their own health or the inability to keep up with their course load. The study's author recommended a structure that would encourage and assist

upper-class engineering students to graduate with the knowledge, disposition, and set of abilities necessary to assume leadership roles in the field. While the paper's descriptions of the leadership qualities and attitudes of scholarship recipients and their future pathways and college experiences are helpful, the suggested approach is time-consuming and inefficient.

In her paper, Angela Bielefeld (2015) [2] explains that differences between the cultures of engineering faculties and scholarship of teaching and learning (SOTL) in the engineering sector persist across all defining features of the faculty members who are actively engaged in the latter. Overall, SOTL evaluated US engineering faculty in terms of the proportion of assistant professors, full professors, women, and employees with a baccalaureate or master's degree [2]. However, the suggested system needs produced data and uses fewer criteria for predictions than are discussed in this study, which details the number of faculty members involved in scholarship-based teaching and learning.

In their 2017 paper "Artificial intelligence-based scholarship DGG credit

pre-assessment system," Sercan, saatci, hande cansiz, gulsha aslan, erkan ozhan describe how scholarships are doled out to deserving students attending institutions and organisations affiliated with various universities on the basis of various criteria.

[4] Data analysis may identify deserving students, but the procedure is time-consuming and can make it hard to choose just a few recipients. In this study, AI and ML were used to make predictions about which students would benefit most from a scholarship. The drawbacks of this approach of prediction based on questions and answers are that it is slower and produces less precise findings.

Okfalisa, Ratik a Fitreani, Yelfi Vitreana (2018) [4] explain that datamining is the technology most often used by researchers for analysis of data, including the prediction of scholarship recipients. [6] The author explains how K Nearest a neighbour (KNN) and linear regression techniques were used to forecast the outcome. Both approaches to determining who should get a scholarship are analysed and compared here. Important indicators including semester attendance, GPA, student activity letter, family card, id card, study results card, etc. are used by the

author here [4]. Author performs simulation testing on data from 8212 scholarship recipients, obtaining ratios of 90:10, 70:30, 50:50, 30:70, and 10:90. The aforementioned methods were analysed using the Rapid miner tool. According to the findings, the KNN approach outperforms the linear Regression algorithm in terms of both effectiveness and efficiency.

Furthermore, Jonalyn Joy B. Labayne, Lustre L. Mercado, and Jheanel Espiritu Estrada (2018) [8] emphasise that FAITH is an institution that provides scholarships to students on the basis of two criteria: a high GPA across the board and a minimum 85th percentile entry in any given topic. Here, we evaluate three different approaches to determining whether or not a scholarship applicant will keep eligibility. The decision tree, Naive Bayes, and K-NN algorithms are used in the suggested system.

The author used the Rapid miner tool for initial data cleansing. Methods used to improve the precision of processed data are the accuracy of the decision tree is 82.18%, that of the KNN is 79.27%, and that of the Naive Bayes is 79.34% [7]. As a result, the author reasoned, pupils who do

poorly in mathematics and English will be ineligible for a scholarship.

Today's e-commerce sector makes great use of data mining techniques for evaluating refurbished electronics[9]. Decision-making aids like decision trees and neural networks are being used by an increasing number of modern businesses[10]. Machine learning and regression models are used in the tourism business nowadays to analyse customer behaviour [11]. Sentiment analysis is another crucial idea utilised nowadays to evaluate goods and track their progress[12]. New and innovative forecast algorithms are employed in most box office analysis applications[13]. Many existing infrastructures use sentiment analysis algorithms in order to foretell process outcomes[14]. Probabilistic generative mining methods are often used to study online criminals[15].

III. METHODOLOGIES

Machine learning (ML) is a subfield of AI concerned with accumulating knowledge from examples. Multiple industries make use of ML, including those dealing with images, medicine, optics, and networking. In machine learning, the goal is to predict an unknown variable based on the inputs

and outputs of previously collected data. K-nearest neighbour, decision tree, support vector machine (SVM), artificial neural network (ANN), ID3, naive bayes, etc. are all examples of machine learning methods.

A. Choice Diagram

The tree-like graph structure used by the decision tree approach may be used for both classification and regression. There is a root node in the centre of the structure that is linked to child nodes and child nodes in turn are linked to leaf nodes. Each node in the graph stands for a set of characteristics, and the terminal leaf node indicates an expected result.

Using methods like ID3, classifier and regression tree(CART), and others, it is possible to generate decision tree rules from a training data set [8, 9].

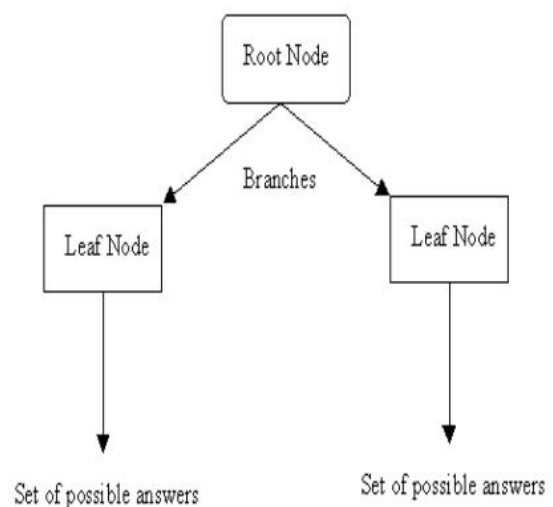


Figure 1. Decision tree

B. Algorithm K-Nearest Neighbour

It's a non-parametric model subset that uses a straightforward ML approach. The algorithm's primary goal is to use "Close" input-output pairings from the training set in order to accurately predict an outcome value from input data. The KNN technique may be used to big datasets without much difficulty.

It's a flexible algorithm. It's versatile enough to be put to use in search, regression, and classification.

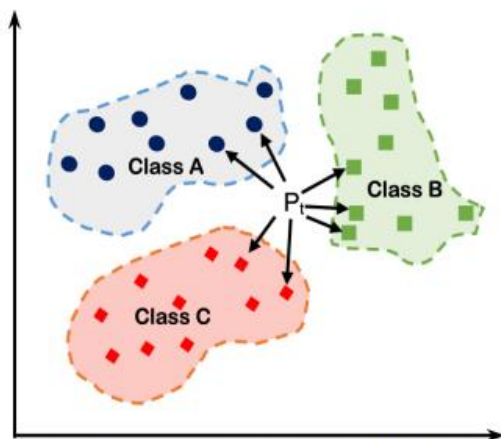


Figure 2. K-Nearest Neighbor Algorithm

C. SVM (Support Vector)

The name "Kernel methods framework" describes this particular algorithm. This method attempts to solve a problem by first translating the input attribute space to a lower dimensional space.

The hyperplane's parameters may then be predicted using the vectors generated by the Supervised Vector Machine. The SVM method is used to resolve regression-type issues by taking the hyperplane into account. When compared to K-NN and artificial neural networks, SVM provides superior performance in terms of accuracy and training time.

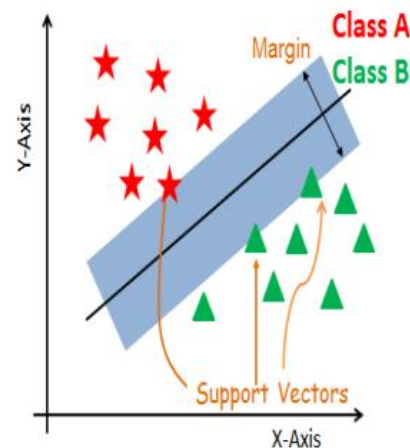


Figure 3. Analysis of Support Vector Machine (SVM)

D. An Attempt to Simulate the Brain

The field of artificial neural networks is based on the study of neural networks in living organisms. Artificial neural networks (ANNs) are composed of networks of interconnected nodes. A neuron is the central component of an ANN. The primary goal of this technique is to get a target value as an output by manipulating various neuron attribute/parameters inside a network.

Input, hidden, and output layers make up the fundamental neuronal structure of an ANN. This framework is sometimes referred to as "multi-layer perception." The input layer's job is to take a vector as input. After that, the input layer sends the input vector to every neuron in the hidden layer. The last layer, the output layer, produces a vector of results. The speed of the testing phase, which follows the training phase, is the algorithm's greatest strength.

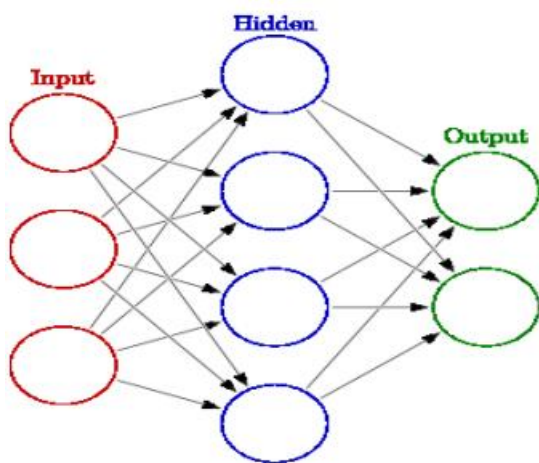


Figure 4. Artificial Neural Network

IV RESULT ANALYSIS

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

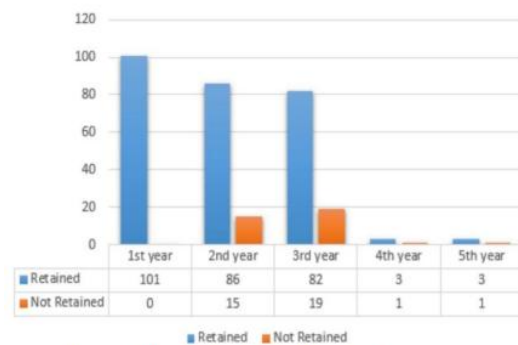


Figure 5. Retention of Scholarship of 5 years

Figure 5 above is a statistical analysis of students who kept their scholarships for five years [8]. All first-year students are guaranteed to keep their scholarship for the second year. Many pupils fell from the wagon in the second year.

Only three students were still receiving scholarships at the conclusion of the fifth year.

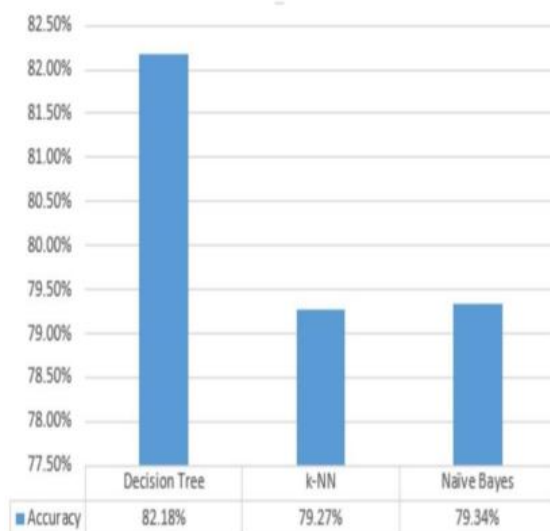


Figure 6. The accuracy of algorithms in processing data

Figure 6 represents result of the accuracy of three algorithms. By using of the

Decision tree [8] algorithm gives an accuracy of 82.18% , k-NN gives the 79.27% and lastly the Naïve Bayes which gives accuracy as 79.34%. It means Decision tree has huge accuracy rather than k-NN and Naive Bayes.

K Value	TP	FP	FN	TN	Accuracy	Precision	Recall	Classification error	Absolute error	Root mean squared error
K.1	401	0	0	422	100%	100%	100%	0%	0.000 +/- 0.000	0.000
K.3	401	0	0	422	100%	100%	100%	0%	0.000 +/- 0.000	0.000
K.5	401	0	0	422	100%	100%	100%	0%	0.000 +/- 0.000	0.000
K.7	401	0	0	422	100%	100%	100%	0%	0.000 +/- 0.000	0.000
K.9	401	0	0	422	100%	100%	100%	0%	0.000 +/- 0.000	0.000

Table 1. Confusion Matrix KNN 70:30

Accuracy, precision, recall, and classification error are all explained by 0.00% by the KNN matrix shown above the confusion one[4].

Analytical data for all pupils is handled in Figure 7. The acronym ALGEB3U explains how algebra consistently yields students that are unable to keep up with their academic requirements. Second, BASGRAM had the second highest rate of students who did not keep up with their academic obligations.

The courses that prove challenging to students' ability to keep their scholarship are shown in Figure 8. Based on our findings, the BSA course had the highest

"not retained" rate. When compared to other courses, its difficulty level is higher.

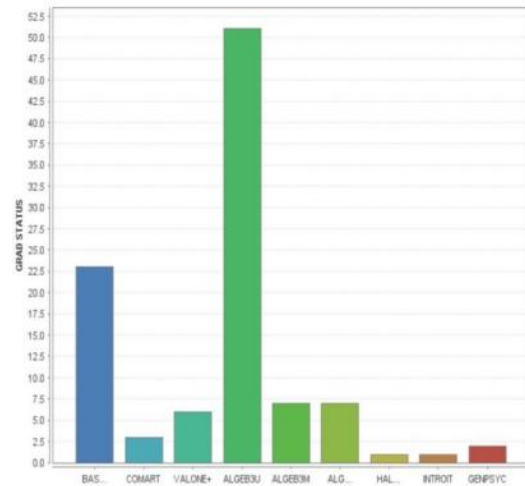


Figure 7. Analytical study data of all subjects

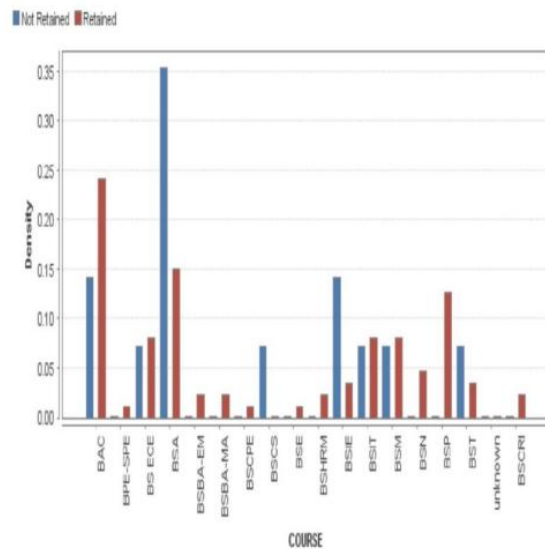


Figure 8. Analytical study data by Course of retained and non-retained students

V CONCLUSION

This study provides a comprehensive overview of the Machine Learning and Data Mining methods currently being employed in the process of scholarship prediction. Scholarship prediction is simplified and improved by applying ML and DM methods. There are a lot of factors to consider when trying to forecast data. Both ML and DM methods are useless without some kind of training data. Data collection becomes more challenging as a result. There is no way for us to compile the optimal prediction method.

Because predictions are dependent on the values in the training data, The ML and DM techniques allow the instructors to keep close tabs on the scholarship winners and their academic struggles.

REFERENCES

- [1] Thalia Anagnos, Eva Schiorring, "Helping Students Get More than Tier Money from an Engineering Scholarship Program" IEEE, 2018.
- [2] Angela R. Bielefeldt, "Characteristics of Engineering Faculty Engaged in the Scholarship of Teaching and Learning" IEEE, 2015.
- [3] Naga Lakshmi Somu, Prasadu Peddi (2021), An Analysis Of Edge-Cloud Computing Networks For Computation Offloading, Webology (ISSN: 1735-188X), Volume 18, Number 6, pp 7983-7994.
- [4] Sercan, Saatci, Hande cansiz, gulsha aslan, erkan ozhan "Artificial intelligence-based Scholarship DQG Credit pre-assessment system" IEEE, 2017.
- [5] G. Upton, "The Effects of Merit-Based Scholarships on Educational Outcomes," Journal of Labor Research, vol. 37, no. 2, pp. 235-261.
- [6] Okfalisa, Ratik a fitriani, Yelli Vitriana, " The Comparison of Linear Regression Method and k- Nearest Neighbors in Scholarship Recipient" IEEE, 2018.
- [7] N. T. Nghe, "A Comparative Analysis of Techniques for Predicting Academic Performance," IEEE, 2007.
- [8] Jonalyn Joy B.Labayne, Lester L. Mercado, Jheanel Espiritu Estrada, "Model Development of Students Scholarship Status at First at Asia Institute of technology And Humanities(FAITH)" IEEE,2018.
- [9] Suma, V., and Shavige Malleswaram Hills. "Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics." Journal of Soft Computing

Paradigm (JSCP) 2, no. 02 (2020): 101-110

[10] Kumar, T. Senthil. "Data Mining Based Marketing Decision Support System Using Hybrid Machine Learning Algorithm." Journal of Artificial Intelligence 2, no. 03 (2020): 185-193.

[11] Chakrabarty, Navoneel. "A Regression Approach to Distribution and Trend Analysis of Quarterly Foreign Tourist Arrivals in India." Journal of Soft Computing Paradigm (JSCP) 2, no. 01 (2020): 57-82.

[12] Prasadu Peddi (2016), Comparative study on cloud optimized resource and prediction using machine learning algorithm, ISSN: 2455-6300, volume 1, issue 3, pp: 88-94.

[13] Hur, Minhoe, Pilsung Kang, and Sungzoon Cho. "Box-office forecasting based on sentiments of movie reviews and Independent subspace method." Information Sciences 372 (2016): 608-624

[14] Prasadu Peddi (2023), Using a Wide Range of Residuals Densely, a Deep Learning Approach to the Detection of Abnormal Driving Behaviour in Videos, ADVANCED INFORMATION

TECHNOLOGY JOURNAL, ISSN 1879-8136, volume XV, issue II, pp 11-18.

[15] Parvathy G, Bindhu JS (2016) A probabilistic generative model for mining cybercriminal network from online social media: a review. Int J Computer 134(14):1-4.doi:10.5120/ijca2016908121
Google Schola