



International Journal of Marketing Management

ISSN 2454 - 5007



www.ijmm.net

Email ID: editor@ijmm.net , ijmm.editor9@gmail.com

Methods for Student Scholarship Prediction Using Data Mining and Machine Learning: A Review

¹ G.Sudheer Kumar, ² G.Vaishnavi,

¹Assistant Professor, Megha Institute of Engineering & Technology for Women, Ghatkesar.

² MCA Student, Megha Institute of Engineering & Technology for Women, Ghatkesar.

Abstract

From a literature review to a scholarship forecast utilizing Machine Learning and Data Mining techniques, this review article has it all. Also included is a brief explanation of the ML/DM tools utilized by the academics. Data sets are also characterized as crucial in ML/DM approaches. In the IT sector, machine learning is booming in popularity. In recent years, data mining and machine learning have emerged as potent tools with wide-ranging applications in industries as diverse as business, education, and information technology. Using all of these methods, we can tackle the various ML/DM algorithms. For the purpose of detecting the continuation of each student's scholarship, algorithms like Naïve Bayes, Decision Tree, and k-NN provide more accurate results. In the end, the suggested model would compile a list of deserving scholarship recipients, and we've already discussed how accurate each method was.

Keywords-

Machine learning, data mining, scholarship, GPA, forecast, and the like.

1. INTRODUCTION

Various Machine Learning and Data Mining Techniques for Scholarship and Outcome Prediction are Covered in This Paper. Methods for Machine Learning and Data Mining are detailed, along with some of the challenges encountered. We talk about the pros and cons of ML and DM techniques. Additionally, this proved to be the most effective strategy for scholarship forecasting. Students are able to continue their studies with the

financial support of a scholarship. Governments or non-governmental organizations (NGOs) often provide financial aid in the form of scholarships. A time when pupils are acknowledged for what they have achieved. It empowers individuals to pursue other aspirations. Award programs may be broadly categorized into three main types: merit-based, student-specific, and career-specific. Students are offered several scholarship programs according to their qualifying requirements. Most academic scholarships likely employ GPAs as a criterion for selection. There are a lot of factors that go into deciding who gets an athletic scholarship, including grades, extracurricular activities, community service, and overall performance. In this work, we looked at a scholarship prediction system that uses ML/DM methods to automatically determine which students are qualified for a scholarship based on criteria including percentage, GPA, marks, yearly income, communication abilities, and so on.

2. LITERATURE SURVEY

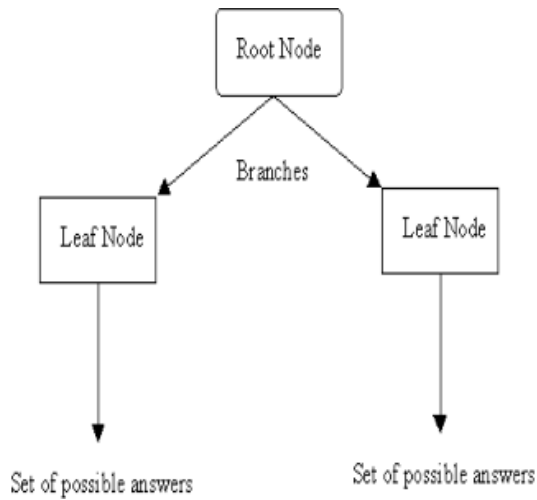
Scholarships are a great way for students to pay for college, but thalia anagnos and eva schiorring (2018) [1] say that students should do everything they can to keep themselves financially stable so that they don't have to worry about things like health problems or running out of money. In order to encourage and assist upper-class students to complete their engineering degrees with specific knowledge, attitudes, and abilities that emerge as leaders, the researcher created a system that offers generous assistance and programs. Although the suggested approach is inefficient and takes too long to analyze data, this study explains how scholarships may have the greatest possible effect on scholarship recipients' leadership abilities, attitudes, future careers, and college experiences. In his article, Angela R. Bielefeldt (2015) [2] indicates that there are distinctions between engineering faculty civilization

and engineering faculty scholarship of teaching and learning (SOTL) in the engineering sector. These disparities persist across all faculty features. SOTL analyzed the engineering faculty in the United States as a whole by looking at the proportion of assistant professors, full professors, women, and employees at baccalaureate and master's degree programs [2]. Although the suggested approach necessitates generation of a smaller number of factors for prediction, this study details the amount of faculty members involved in scholarly teaching and learning. Scholarships were awarded to eligible students according to various criteria in the 2017 article "Artificial intelligence-based scholarship DGG credit pre-assessment system" [4] by Sercan, saatci, hande cansiz, gulsha aslan, and erkan ozhan. These students had attended various institutions and organizations affiliated with various universities. [4] Scholarship eligibility is determined using data analysis, which is a time-consuming and challenging method for shortlisting deserving applicants. In this case, researchers employed AI and ML to forecast which students would be most deserving of a grant. The system's drawbacks include longer prediction times and less reliable outcomes due to its question-and-answer oriented prediction method. Okfalisa, Ratik, Fitreani, and Yelvi Vitreana (2018) state that datamining is one of the methods used to analyze data for purposes such as scholarship recipient prediction. The author explains how linear regression techniques and K-Nearest Neighbors (KNN) were used to forecast the results. The purpose of this research is to evaluate the two approaches to the scholarship recipient dilemma. Key characteristics used here include semester attendance, grade point average, active student statement letters, family cards, identification cards, study result cards, etc. [4]. The author simulates testing of training and testing data, which yields ratios of 90:10, 70:30, 50:50, 30:70, and 10:90. The data comes from 8212 scholarship recipients. The results of the aforementioned methods' analyses were examined using the rapid miner tool. According to the findings, when compared to the linear regression technique, the KNN approach produces more effective and efficient outcomes. Additionally, as stated by jonalyn joy B. Labayne, Luster L. Mercado, and Jheanel Espiritu Estrada (2018) [8], FAITH is a school that offers financial aid to students who meet two criteria: a minimum cumulative GPA of 90% and a minimum of 85% in any one subject. [8] To determine whether scholarship winners are qualified to keep their

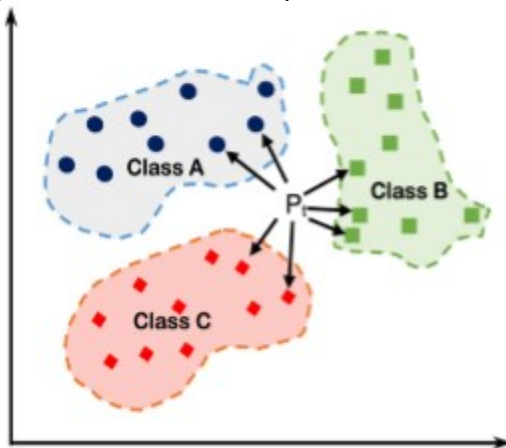
awards, this article compares three different ways. Decision trees, Naive Bayes, and K-NN are all part of the suggested system. In order to preprocess the acquired data and eliminate extraneous information, the author used the Rapid Miner program. Decision trees have an accuracy of 82.18%, KNN of 79.27%, and Naive Bayes of 79.34% when processing data using these approaches [7]. Students who do poorly in algebra and English will not be awarded scholarships, since the author concluded that decision trees were the most appropriate algorithms. At the moment, data mining techniques are being used by the e-commerce sector to examine used electronics products [9]. Nowadays, decision-tree and neural network-based decision-support systems are used by the majority of businesses [10]. Modern tourism makes use of regression models and machine learning to study visitors' habits and preferences [11]. Today, sentiment analysis is a crucial tool for evaluating goods and their effectiveness [12]. A lot of apps employ new forecast algorithms to look at box office predictions [13]. In order to foretell the process's conclusion, several systems use sentiment analysis algorithms [14]. Probabilistic generative mining methods find extensive usage in analyzing cybercriminals on social media [15].

3. METHODOLOGIES

Data used for training is the focus of machine learning (ML), a subfield of AI. Machine learning finds value in several domains, including optical networks, medical applications, image processing, and more. ML algorithms are used to make predictions about the unknown element by analyzing the training data and its inputs and outputs. The following are examples of machine learning algorithms: k-nearest neighbor, decision tree, Support Vector Machine (SVM), Artificial Neural Network, ID3, Naïve Bayes, and many more. A.Nestle Review With its classifier and regression built into its tree-like network structure, the decision tree method is a powerful tool. A structure's root node establishes a connection to its subnodes, which in turn are linked to its leaf nodes by links. The qualities are represented by the nodes, and the outcomes or desired results are at the end of each leaf node. In order to build decision tree rules, many methods may be used, including ID3, classifier, and regression tree (CART) [8].

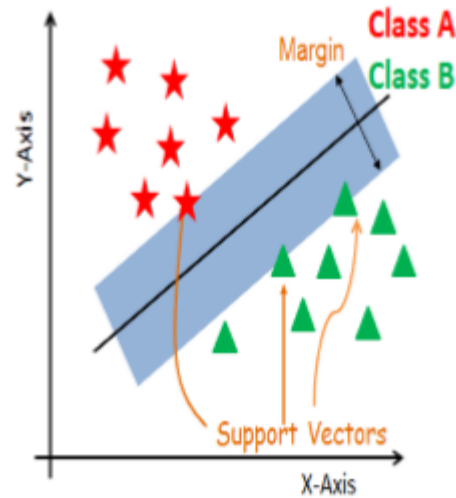
**Figure 1. Decision tree**

Section B. The K-Nearest Neighbor Method It is a non-parametric model subset that employs a straightforward ML approach. Predicting an output value from input data using "Close" input-output pairs available during training is the algorithm's primary objective. In addition to being suitable for big datasets, the KNN method is user-friendly. There is a lot of flexibility in the algorithm. Classification, regression, and search are all possible uses for it.

**Figure 2. K-Nearest Neighbor Algorithm**

Category C: Support Vector Machines This framework of algorithms is sometimes referred to as Kernel techniques. In order for this technique to function, it first attempts to map a low-dimensional input attribute space. In order to forecast the hyperplane's parameters, these Supervised Vector Machine vectors are used. When solving regression-

type issues, the SVM method takes the hyperplane into account. Instead of K-NN and ANNs, SVMs outperform them in terms of training time and accuracy.

**Figure 3. Analysis of Support Vector Machine (SVM)**

Section D: Neural Networks Artificial Biological neural networks serve as inspiration for artificial neural networks. A network of interconnected computational units known as artificial neurons makes up an ANN. A neuron is the fundamental building block of an ANN. The primary goal of this technique is to get a desired output by adjusting various neuron attributes and parameters in a network. The three neuronal layers that make up an ANN's fundamental structure are the input, hidden, and output layers. A multi-layer perceptual structure is another name for it. The input layer's job is to take a vector as input. The input layer then provided a vector of instructions to each neuron in the hidden layer. The output layer is used to produce an output vector. One major benefit of this algorithm is how quickly it can transition from training to testing.

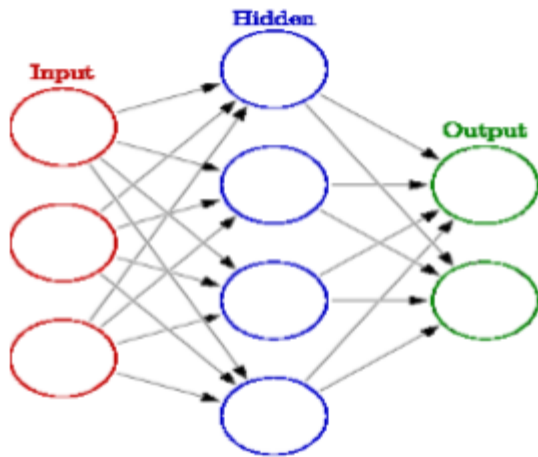


Figure 4. Artificial Neural Network

4. RESULT ANALYSIS

According to the aforementioned research, a plethora of data mining and ML techniques were used to forecast students' scholarship prospects.

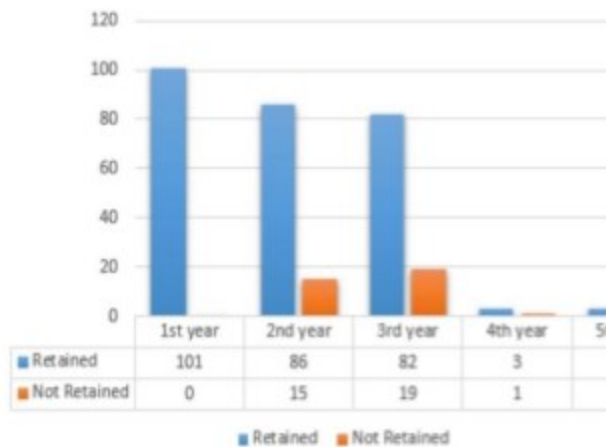


Figure 5. Retention of Scholarship of 5 years

The statistics on students' scholarship retention after five years is shown in figure 5 above [8]. Scholarships are renewable for all first-year students. Many pupils struggled to keep up in second year. Three students out of a total of 925 had their scholarships renewed at the conclusion of the fifth year.

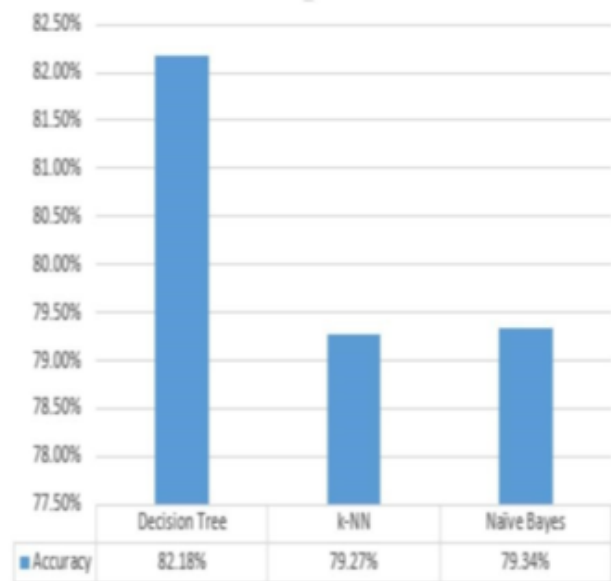


Figure 6. The accuracy of algorithms in processing data

The accuracy results of three methods are shown in Figure 6. The accuracy rates for the decision tree method (82.18 percent), k-NN (79.27 percent), and Naïve Bayes (79.34 percent) are as follows. As a result, decision trees outperform k-NN and Naive Bayes in terms of accuracy.

Table 1. Confusion Matrix KNN 70:30

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

All four of the above KNN confusion matrices—accuracy, precision, recall, and classification error—are 100% [4]. All pupils' acceptable analytical data is handled by Figure 7. Algebra has a significant proportion of results that do not preserve scholarship, as clearly described by ALGEB3U. The second most common reason for not keeping a scholarship was

BASGRAM. Figure 8 shows which classes are having trouble keeping their students' scholarships. The BSA course had the largest rate of "not" based on our observations. held on to. The difficulty level is higher than that of the other courses.

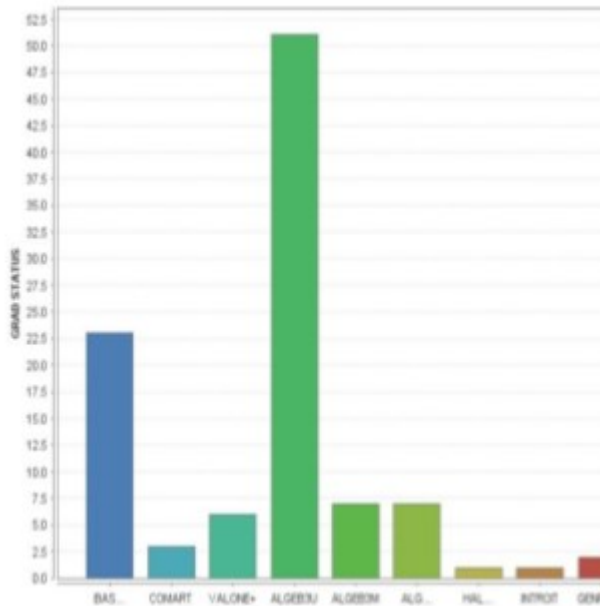


Figure 7. Analytical study data of all subjects

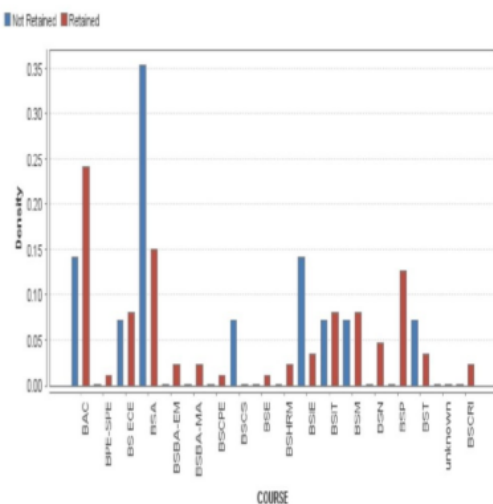


Figure 8. Analytical study data by Course of retained and non-retained students

5. CONCLUSION

Scholarship predictions are detailed in this article, which also includes a summary of data mining and machine learning methods. Scholarship predictions are made considerably simpler and more successful with the use of ML and DM approaches. We need to consider a lot of factors while making data predictions. Without training data, ML and DM algorithms will not be able to function. It becomes more challenging for dataset collection. Unfortunately, we will not be able to gather the most effective system for making predictions. Reason being, values from training data form the basis of prediction: The instructions may simply keep tabs on the scholarships and the challenges that the beneficiaries encounter in their studies by using the ML and DM approaches.

REFERENCES

- [1] Thalia Anagnos, Eva Schiorring, "Helping Students Get More than Teir 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] K. Bunkar, "Data Mining: Prediction for Performance Improvement of Graduate Students using," IEEE, 2012.
- [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 Malleshwara 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] Mitra, Ayushi. "Sentiment Analysis Using Machine Learning Approaches (Lexicon based on movie review dataset)." Journal of Ubiquitous Computing and Communication Technologies (UCCT) 2, no. 03 (2020): 145-152.

[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] Vohra, S. M., and J. B. Teraiya. "A comparative study of sentiment analysis techniques." Journal JIKRCE 2, no. 2 (2013): 313-317.

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