ISSN: 2249-7196
IJMRR/April. 2025/ Volume 15/Issue 2s/237-242
Naidu Devi Prasanna / International Journal of Management Research & Review

Machine Learning Framework for Prediction of Admission in Engineering College

Naidu Devi Prasanna
PG scholar, Department of MCA, DNR College, Bhimavaram, Andhra Pradesh.
A.Naga Raju
(Assistant Professor), Master of Computer Applications, DNR college, Bhimavaram, Andhra Pradesh.

Abstract: ***Utilizing machine learning (ML), enormous amounts of information can be re-evaluated and discover particular patterns that might not be immediately noticeable or recognizable to humans. ML strategies have increasingly been used to assess educational data such as student class performance. In the pursuit of the academic well-being of students, the utilization of neoteric technologies such as data mining, data management, and ML has increased. The idea of extracting undisclosed information from many raw databases is called data mining. Consequently, the exploration of knowledge acquisition relates to predictive ML models and subsequent decision-making. State-of-the-arts of data mining and ML have become more acceptable in predicting student examination evaluations such as grades, achievement, etc. Generally, conventional data mining for educational data analysis aimed at solving problems in an educational context can be described as educational data mining. Currently, intelligent computer-based methods such as artificial intelligence and data mining have been successfully applied to improve people's daily lives. A couple of million students participate in the bachelor's entrance examination at government-run universities each year in India. Nevertheless, only a few thousand are admitted after this competitive examination. In some cases, it was observed that many candidates struggled hard during this period. However, they could not get admission to a public university in India, resulting in an unforeseeable future. Numerous factors could be behind their unsuccessful admission to a public university, such as family circumstances, frustration, admission test anxiety, etc. However, Indian students need admission to a public university because private university education costs are too high for middle-income and low-income families. In contrast, the government primarily covers public university costs. Therefore, this project implements the prediction of college admission for engineering or college students using machine learning algorithm.***

Keywords: ***Machine learning, Engineering college, university.***

I. INTRODUCTION

Today, all higher education institutions, especially computer and engineering colleges, face challenges in the admissions process. Each university should strive for an admissions system based on valid and reliable admissions criteria that select candidates likely to succeed in its programs. In addition, each university should use the best possible techniques for predicting applicants' future academic performance before admitting them [1].

This would support university decision makers as they set efficient admissions criteria. However, most higher education institutions face challenges when they analyse their large educational databases to predict students' performance [2]. This is because they use only conventional statistical methods rather than new and efficient predictive techniques such as Educational Data Mining (EDM), which is the most popular technique to evaluate and predict students' performance.

EDM is the process of extracting useful information and patterns from a huge educational database [3], which can then be used to predict students' performance. As aresult of better information, student performance can be more effectively improved through more effective strategic programs. Today, all higher education institutions face difficulties in the admission process. Every college ought to make a choice in its admission form which is dependent on legitimate and credible admissions procedures that select the student candidates prone to prevail in its programs.

Furthermore, every college should use the most ideal [4] methods for foreseeing candidates' future academic performance before conceding them. This result would uphold college chiefs as they set effective admissions criteria. Recently, educational data mining (EDM), a subfield of

datum mining, has appeared that has practical experience in educational datum that is the most common method to value and foresee students' execution. EDM is the way toward extricating helpful information and examples from an enormous educational database [5], which would then be able to be used to predict students' performance.

Classification is one of the most frequently studied problems by DM and machine learning (ML) researchers [1]. It consists of predicting the value of a (categorical) attribute (the class) based on the values of other attributes (the predicting attributes) [4]. The aim of this paper is to determine the factors estimating & guiding the students to select engineering college for their first year admission. Most of students & parents are spending unnecessary efforts, time & money on selecting right engineering college for first year admission.

Sometimes the students who are seeking admission is not eligible to take admission into engineering program based on their past academic record. Also sometimes the students are seeking the admission to the college for she or he are not eligible as per the merit of that college. So here researcher has built predictive model to guide the students about their admissibility in the desired college & also suggest the college where they will get the admission. So to achieve this objective we may include machine learning capabilities that allow to improve their performance based on experience, just as humans do. [3] As right College plays very vital role form the students' placement and career point of view the researcher has implemented various algorithms to achieve this objective.

II. LITERATURE SURVEY

[1] L. H. Son and H. Fujita, "Neural-fuzzy with representative sets for prediction of student performance," Applied Intelligence, vol. 49, no. 1, pp. 172–187, 2019.

In this paper, a new method for handling the Multi-Input Multi-Output Student Academic Performance Prediction (MIMO SAPP) problem is proposed. The MIMO SAPP aims to predict the future performance of a student after being enrolled into a university. The existing methods have limitations of using a parameter set and an unsuitable training strategy. Thus, the new method called MANFIS-S (Multi Adaptive Neuro-Fuzzy Inference System with Representative Sets) uses multiple parameter sets and a special learning strategy to resolve those weaknesses. Specifically, the idea of multiple parameter sets is to approximate the MANFIS-S model with many meaningful parameters to ensure the performance of system.

[2] S. Bharara, S. Sabitha, and Bansal, "Application of learning analytics using clustering data mining for students' disposition analysis," Education and Information Technologies, vol. 23, no. 2, pp. 957–984, 2018.

Learning Analytics (LA) is an emerging field in which sophisticated analytic tools are used to improve learning and education. It draws from, and is closely tied to, a series of other fields of study like business intelligence, web analytics, academic analytics, educational data mining, and action analytics. The main objective of this research work is to find meaningful indicators or metrics in a learning context and to study the inter-relationships between these metrics using the concepts of Learning Analytics and Educational Data Mining, thereby, analyzing the effects of different features on student's performance using Disposition analysis. In this project, K-means clustering data mining technique is used to obtain clusters which are further mapped to find the important features of a learning context. Relationships between these features are identified to assess the student's performance

[3] C. M. D. Bondoc and T. G. Malawit, "Classifying relevant video tutorials for the school's learning management system using support vector machine algorithm," Global Journal of Engineering and Technology Advances, vol. 2, no. 3, pp. 1–9, 2020.

Today many schools, universities and institutions recognize the necessity and importance of using Learning Management Systems (LMS) as part of their educational services. This research work has applied LMS in the teaching and learning process of Bulacan State University (BulSU) Graduate School (GS) Program that enhances the face-to-face instruction with online components. The researchers uses an LMS that provides educators a platform that can

motivate and engage students to new educational environment through manage online classes.

[4] K. Coussement, M. Phan, A. De Caigny, D. F. Benoit, and A. Raes, "Predicting student dropout in subscription-based online learning environments: the beneficial impact of the logit leaf model," Decision Support Systems, vol. 135, 2020.

Online learning has been adopted rapidly by educational institutions and organizations. Despite its many advantages, including 24/7 access, high flexibility, rich content, and low cost, online learning suffers from high dropout rates that hamper pedagogical and economic goal outcomes. Enhanced student dropout prediction tools would help providers proactively detect students at risk of leaving and identify factors that they might address to help students continue their learning experience. Therefore, this study seeks to improve student dropout predictions, with three main contributions.

In reviewing LLM segments, these results show that different insights emerge for various student segments with different learning patterns. This notable result can be used to personalize student retention campaigns. Online learning refers to educational activities in cyberspace; it has transformed educational markets [1]. Educational content providers, both academic and commercial, use the Internet as a primary content delivery channel [2,3], to leverage its advantages in relation to content diversity, flexibility, scalability, accessibility, and cost effectiveness [4]. Among the various business models, online subscription-based learning platforms (e.g., Datacamp, Coursera, Lynda) are the most popular.

[5] Dhankhar, K. Solanki, and A. Rathee, "'Predicting student's performance by using classification methods," Journal of advanced trends in computer science and engineering, vol. 8, no. 4, pp. 1532–1536, 2019.

Most of the developing countries are facing the problem of ever-rising low-quality population. To convert this low-quality population into a high-quality one, efforts are required to be laid down. These efforts include investment in Research and development and in the education sector. If the people living in an area will be educated then they will be productive for the nation and eventually contribute towards its GDP. The advancement of technology helps educational institutions to turn raw data into actionable insights to achieve desirable results. This study has worked towards prediction models so that student's performance be evaluated timely so that necessary steps be taken in due time to improve their performance.

[6] H. A. Mengash, "Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems," in IEEE Access, vol. 8, pp. 55462- 55470, 2020, doi: 10.1109/ACCESS.2020.2981905.

An admissions system based on valid and reliable admissions criteria is very important to select candidates likely to perform well academically at institutions of higher education. This study focuses on ways to support universities in admissions decision making using data mining techniques to predict applicants' academic performance at university. A data set of 2,039 students enrolled in a Computer Science and Information College of a Saudi public university from 2016 to 2019 was used to validate the proposed methodology.

[7] M. S. Acharya, A. Armaan and A. S. Antony, "A Comparison of Regression Models for Prediction of Graduate Admissions," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), 2019, pp. 1-5, doi: 10.1109/ICCIDS.2019.8862140.

Prospective graduate students always face a dilemma deciding universities of their choice while applying to master's programs. While there are a good number of predictors and consultancies that guide a student, they aren't always reliable since decision is made on the basis of select past admissions. In this paper, we present a Machine Learning based method where we compare different regression algorithms, such as Linear Regression, Support Vector Regression, Decision Trees and Random Forest, given the profile of the student. We then compute error functions for the different models and compare their performance to select the best performing model. Results then indicate if the university of choice is an ambitious or a safe one. The Graduate Program is an

exhaustive task that requires thorough preparations, both in terms of building a noteworthy profile and choosing universities that offer relevant programs.

III. PROPOSED METHOD

3.1 Dataset description

- GRE Score - Out of 340 TOEFL Score - Out of 120
- University Rating - Between 1 to 5 (5 being the best)
- SOP - Between 1 to 5 (5 being the best)
- LOR - Between 1 to 5 (5 being the best)
- CGPA - Out of 10

Research - 1 if student has research experience, else 0 Chance of Admit - Probability of getting accepted into graduate program

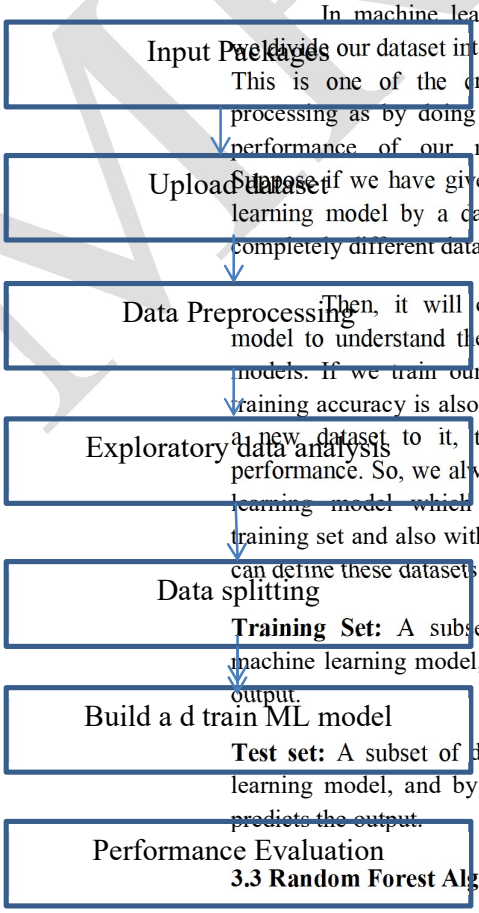


Fig.5 Flowchart

3.2 Data Pre-processing

In Machine learning Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So, for this, we use data pre-processing task.

Why do we need Data Pre-processing?

A real-world data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used for machine learning models. Data pre-processing is required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model.

3.2.1 Splitting the Dataset into the Training set and Test set

In machine learning data pre-processing, we divide our dataset into a training set and test set. This is one of the crucial steps of data pre-processing as by doing this, we can enhance the performance of our machine learning model. Suppose if we have given training to our machine learning model by a dataset and we test it by a completely different dataset.

Then, it will create difficulties for our model to understand the correlations between the models. If we train our model very well and its training accuracy is also very high, but we provide a new dataset to it, then it will decrease the performance. So, we always try to make a machine learning model which performs well with the training set and also with the test dataset. Here, we can define these datasets as:

Training Set: A subset of dataset to train the machine learning model, and we already know the output.

Test set: A subset of dataset to test the machine learning model, and by using the test set, model predicts the output.

3.3 Random Forest Algorithm

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both

Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

IV. RESULT

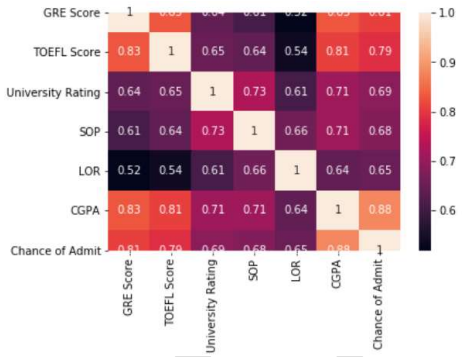
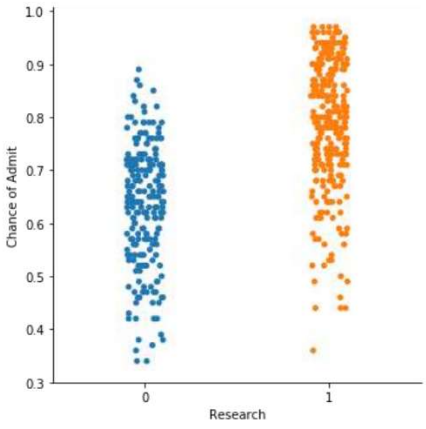


Fig. 6 Correlation matrix



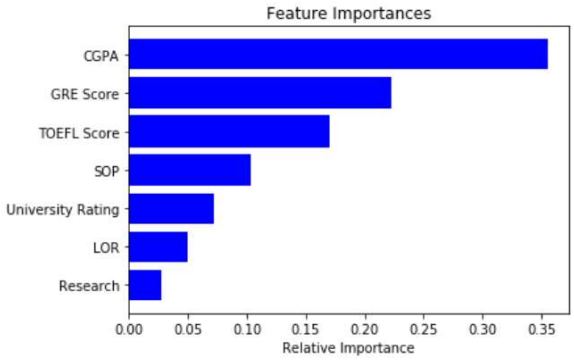


Fig.7. Graphical representation

V. CONCLUSION

Academic performance is the primary concern for most colleges in most countries. There are extensive quantities of data generated in learning systems. This data holds hidden knowledge that could be used to heighten the students' academic success. In this research, a suggested model of student achievement prediction was constructed totally on ensemble methods. The predictive model by classifiers random forest method (bagging and boosting) deal with raising these classifiers' benefits. The retrieved results expose that there is an enhancement in these models over the conventional classifiers. Then, the proposed method combines two different classifiers with one of the bagging or boosting process. This method gave better results than previous methods that contribute to the growth of the accomplishment of students and educational systems. We will assemble information from numerous understudies of different instructive organizations and use some great data mining techniques to deliver a substantial yield. This project empowers instructional frameworks, foundations, understudies, and instructors to fortify their performance.

REFERENCES

[1] L. H. Son and H. Fujita, "Neural-fuzzy with representative sets for prediction of student performance," Applied Intelligence, vol. 49, no. 1, pp. 172–187, 2019.

[2] S. Bharara, S. Sabitha, and Bansal, "Application of learning analytics using clustering data mining for students' disposition analysis,"

Education and Information Technologies, vol. 23, no. 2, pp. 957–984, 2018.

[3] C. M. D. Bondoc and T. G. Malawit, "Classifying relevant video tutorials for the school's learning management system using support vector machine algorithm," Global Journal of Engineering and Technology Advances, vol. 2, no. 3, pp. 1–9, 2020.

[4] K. Coussement, M. Phan, A. De Caigny, D. F. Benoit, and A. Raes, "Predicting student dropout in subscription-based online learning environments: the beneficial impact of the logit leaf model," Decision Support Systems, vol. 135, 2020.

[5] Dhankhar, K. Solanki, and A. Rathee, "'Predicting student's performance by using classification methods," Journal of advanced trends in computer science and engineering, vol. 8, no. 4, pp. 1532–1536, 2019.

[6] H. A. Mengash, "Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems," in IEEE Access, vol. 8, pp. 55462-55470, 2020, doi: 10.1109/ACCESS.2020.2981905.

[7] M. S. Acharya, A. Armaan and A. S. Antony, "A Comparison of Regression Models for Prediction of Graduate Admissions," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), 2019, pp. 1-5, doi: 10.1109/ICCIDS.2019.8862140.

[8] G. Tsang, S. -M. Zhou and X. Xie, "Modeling Large Sparse Data for Feature Selection: Hospital Admission Predictions of the Dementia Patients Using Primary Care Electronic Health Records," in IEEE Journal of Translational Engineering in Health and Medicine, vol. 9, pp. 1-13, 2021, Art no. 3000113, doi: 10.1109/JTEHM.2020.3040236.

[9] R. El-Bouri, D. W. Eyre, P. Watkinson, T. Zhu and D. A. Clifton, "Hospital Admission Location Prediction via Deep Interpretable Networks for the Year-Round Improvement of Emergency Patient Care," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 1, pp. 289-300, Jan. 2021, doi: 10.1109/JBHI.2020.2990309.