Use Machine Learning to Predict the Probability of Being Admitted to a University

[1]S. Venkateswara Rao, [2]A.Navya

[1]Assistant Professor, Megha Institute of Engineering & Technology for Women, Ghatkesar.
[2]MCA Student, Megha Institute of Engineering & Technology for Women, Ghatkesar.

Abstract

Under the current circumstances, students frequently struggle to identify a university that fits their profile and allows them to pursue higher education. There are internet apps and advising services that suggest universities, but they charge exorbitant consulting fees and the recommendations are inaccurate. Therefore, the goal of this study is to create a model that can precisely forecast the likelihood of admission to a university. Students can determine whether or not their profile is appropriate by using this model's comparison of scores against prediction chance based on previous data. Although the suggested model makes use of random forest and linear regression methods, the cat boost algorithm yields the best accuracy.

Keywords

Catboost, Linear Regression, and Machine Learning.

I. INTRODUCTION

Since the quality of an individual's education determines their destiny, it plays a crucial role in his life. Following graduation, they frequently have They have a number of questions about going to college and selecting the finest university. The majority of students favor universities with international renown. Therefore, a greater proportion of Indian students choose to continue their education in the United States of America. Despite the existence of reputable universities in India, graduate students struggle to gain admission to these institutions and find employment because there aren't many options. Students spend time and money seeking help because they are unsure which university is preferable. In addition to advisers and consulting offices, there are websites and blogs that offer advice and encouragement to students regarding their chances of admission; nevertheless, these resources are not entirely reliable and cannot be relied upon. Educational institutions may utilize knowledge mining to focus on the most pertinent information in the data they have gathered when students fill out admission inquiry forms. It uncovers information that is concealed in the data and cannot be revealed by queries or reports. After gathering information from admission forms completed by applicants over a number of years, this method must be applied to assess a collection of trends of college applicants. This study develops a machine learning model that considers parameters including undergraduate GPA, study experience, proposal statement and recommendation letter power, university ranking, GRE and TOEFL scores, and more. It forecasts the likelihood of admission once all the inputs have been received. The produced model provides an unbiased picture of measurement on obscure test occasions since it contains significant factual discoveries for the (like) assessment of the chance of confirmation.

II. RELATED WORK

An essay titled "A Comparison of Regression Models for Prediction of Graduate Admissions" was written by Acharya MS. He employed a number of models in this, including SVM, Random, and Linear Regression. forest and computed error functions to compare their performance [1]. An essay by Narendhra Gupta titled "Will I Get in? In their study "Modeling the Graduate Admission Process for American Universities," they took into account variables like GPA and other scores, viewed the issue as a classification problem, and employed a sizable dataset [2]. The administrator can add the assigned seats to a file, and the data is saved in the system. The process of allocation gets faster and the total time for admission allotment reduces. It helps students make well-informed choices regarding which college to enroll in. In order to anticipate admittance to institutions, students can register with their academic and personal information, and administrators can then allot seats to individuals [3]. In their study

"A Quality Based Automated Admission System for Educational Domain," Sushrutha Mitra and Soumya Sahoo tackle the issue as a classification problem and forecast whether or not the student would be admitted to that specific university [4]. Artificial Intelligence has been used by Dr. N. M. Saravana Kumar to teach and assess student performance; additional technologies were also employed [5]. In order to help students understand how scores affect their chances of admission, it also displays the analysis of scores against admission possibility. Additionally, it suggests universities with a higher percentage of chances for students with similar profiles. A college admission predictor developed by Nagineni Dharani and Sathya Ragava uses student information to assess cutoff scores and forecast admission likelihood [5]. GRADE forecasts each new applicant's likelihood of being accepted by the committee based on historical admissions data. It assigns a score to each prediction that is comparable to the score given by human reviewers and provides a summary of the applicant attributes that had the biggest impact on the prediction [6]. Bootstrapping is performed to ascertain the value of the explanatory variables. Crucially, the data is devoid of selection bias brought on by admission, allowing us to estimate objectively how well undergraduate-level metrics predict graduate-level outcomes. According to their findings, 54% of the variation in graduate performance can be explained by undergraduate performance [7]. The question of whether a student offered admission would accept it is treated as a binary classification problem, and the output of a number of different classifiers is then assessed using the metrics of accuracy, precision, recall, F-measure, and area under the receiver operator curve [8].

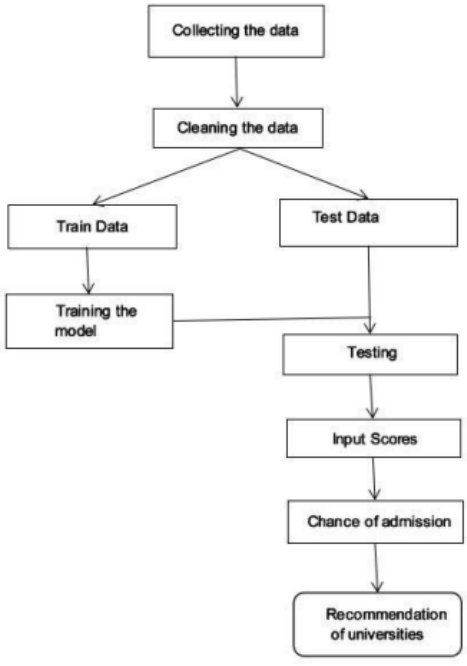
III. PROPOSED WORK

The dataset is regularly divided into train and test sets of 80% and 20% during the model development process. There are 400 profiles in the train set and 100 in the test set. profiles. This is how the modeling dataset appears. One important step in the process is preprocessing. Cleaning the data and getting it ready for a prediction algorithm is the goal. The data acquired from Occidental College just needs minor adjustments to be appropriate for the suggested machine learning techniques. One major issue in data cleansing is figuring out how to handle missing data. Finding missing entries, locating them, and applying a treatment based on the variable form that permits us to use the data in the model are crucial since the function in question may be a good predictor of the algorithm's result. A training set and a testing set were created by randomly dividing the pre-processed data into two classes. Eighty percent of our dataset's 7976 elements were chosen to serve as our training set.

	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
0	337	118	4	4.5	4.5	9.65	1	0.92
1	324	107	4	4.0	4.5	8.87	1	0.76
2	316	104	3	3.0	3.5	8.00	1	0.72
3	322	110	3	3.5	2.5	8.67	1	0.80
4	314	103	2	2.0	3.0	8.21	0	0.65

Fig 1: Training Dataset

Chance of Admission is the variable that needs to be forecasted. The following lists the steps that go into developing a model. as university cutoffs varies annually, thus we've added a requirement in the code that the GRE score must be higher than 250, the TOEFL score must be higher than 50, the CGPA must be higher than 5, and all other requirements must be met. Using the cat boost approach, the model is trained on the training dataset.



The percentage of chances can now be predicted. Additionally, suggested universities are displayed

where students with comparable profiles have greater probability of forecasting

IV PERFORMANCE ANALYSIS

MODEL	Linear Regression
MAE	0.04
MSE	0.003
R2 Score	0.84
ACCURACY	0.93

Table1:Base model evaluation

Similarly, we achieved the maximum accuracy when we used catboost to rerun the process for the same dataset without adjusting.

MODEL	Catboost
MAE	0.03
MSE	0.001
R2 Score	0.89
ACCURACY	0.95

Table2:catboost evaluation (before tuning)

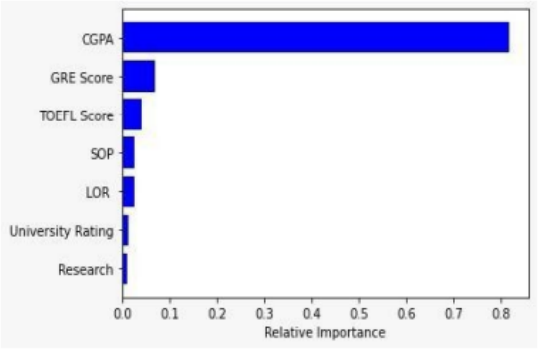


Fig3: Feature importance

CGPA is regarded as the most important attribute because it has more than 80% of the total. The figure above displays both real and anticipated proportion of applicants who will be accepted into a university. This is how the input window appears once the code has been executed. We can forecast the likelihood of admittance after providing the input scores.

Fig6: predicting the chances of admission

Fig7:Getting the Universities

We can obtain recommendations from universities. This can be accomplished by looking for a profile that closely resembles the input profile that was provided and where the likelihood of admittance is higher than the current admissions chance.

Fig8:Recommendation of Universities

Lastly, the analysis. This demonstrates the relationship between all the attributes, student scores, and the proportion of students have a better probability of being admitted. CGPA and Admission Chance Relationship.

V CONCLUSION

The stages involved in training the model are also covered in this work, along with the suggested approach, a few algorithms, and their implementation. talked about. Lastly, the catboost algorithm yields the highest quantitative result of a confirmatory incentive expectation model to date, with an accuracy of 95.

REERENCES

[1] Acharya MS, Armaan A, Antony AS (2019) A comparison of regression models for predictionof graduate admissions.

[2] Gupta N, Sawhney A, Roth D (2016) Will I get in? modeling the graduate admission process for American universities. In: 2016 IEEE 16th international conference on data mining workshops(ICDMW). IEEE

[3] Roa, Annam Mallikharjuna, et al. “College Admission Predictor.” Journal of Network Commu- nications and Emerging Technologies (JNCET), Volume 8, Issue 4, April (2018).

[4] Mishra, S. and Sahoo, S. (2016). A Quality Based AutomatedAdmission System for Educational Domain, pp. 221–223, International conference on Signal Processing, Communication, Power and Embedded System (SCOPES)- 2016.

[5] Kumar, NM Saravana. "Implementation of artificial intelligence in imparting education and evaluatingstudent performance." Journal of Artificial Intelligence 1, no. 01 (2019): 1-9. [6] Austin Waters and Risto Miikkulainen,” Austin Waters and Risto Miikkulainen”, Proceedings of the Twenty-Fifth Innovative Applications of Artificial Intelligence Conference,2013.

[7] Judith Zimmermann, Kay H. Brodersen, Kay H. Brodersen ,”A Model-Based Approach to Predicting Graduate-Level Performance Using Indicators of Undergraduate-Level Performance”, Journal of Educational Data Mining, Volume 7, No 3, 2015.

[8] Kanadpriya Basu , Treena Basu ,Ron Buckmire and Nishu Lal ,” Predictive Models of Student College Commitment Decisions Using Machine Learning”, Data,2019.