

University Administration Prediction Using Machine Learning Based Linear Regression Method

¹MEDA BALAMANIKANTA, 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: *In the present situation, students often have problems finding a suitable institution to conduct a higher search based on their profile. Some online application and counselling sites suggest universities, but they charge high consulting fees, and online applications are inaccurate. Therefore, this study aims to develop a version that appropriately takes into account the proportion of opportunities in the university. This model additionally offers ratings against predictions based on historical facts to allow scholars to assess whether their profile is a good fit. The proposed model uses linear regression and random forest algorithms; however, the cat lift rule set gives maximum accuracy.*

Keywords: *Machine learning, college admission, Linear Regression, Random Forest.*

I. INTRODUCTION

Specific training plays an essential role in your being. As a result, student teachers often question universities where they can get admission, scholarships, and housing. One main problem is getting admission to the university of their dreams [1]. Students are still pursuing their education in globally recognized universities. On the subject of global graduates, the USA is the first preferred by the general public. With recognized schools worldwide, various courses in every field exceptionally recognized teaching and learning programs, and undergraduate scholarships,

college students, are available worldwide. It is estimated that more than 10 million international students are enrolled in more than 4,200 universities and colleges across America, including private and public universities and colleges.

Most students studying in America are from Asian countries like India, Pakistan, Sri Lanka, Japan, and China. They are no longer deciding on the best of America but also deciding on Great Britain, Germany, Italy, Australia, and Canada. The number of people doing better research on these international sites is unexpectedly increasing. The historical motivation for

students who go to foreign universities to obtain a master's degree is no. Job opportunities are few, and the diversity of people for these jobs is very high in their countries [2]. This pushes many students into their careers to pursue their higher studies. It can be seen that there are a huge number of students in US universities who are pursuing a master's degree in computer science. The focus of this research could be on these students. Many colleges in the United States have similar requirements for student admission. Universities consider various factors in addition to the order of competency assessment and assessment of educational records [3]. The idea will count English language proficiency in their overall performance on the English Proficiency Test, which includes TOEFL and IELTS. The University Admissions Committee decides to accept or reject a particular candidate based on the overall profile of the applicant's application.

Additionally, when filling out the admissions application bureaucracy, educational institutions can use educational mining to focus on the most relevant details in the facts they collect. It reveals hidden records in statistics that queries and reports cannot detect. This approach evaluated a group of college applicants after collecting statistics from

admissions records using applicants for admission over several years.

This paper creates a machine proficiency model that considers barriers such as the GRE score, TOEFL score, university ranking, proposal statement, letters of recommendation, undergraduate GPA, and college experience. After receiving all inputs, anticipate entry risk. For test events that are difficult to understand, the structured version has extensive factual knowledge to assess the chance of confirmation and thus provides an unbiased impression of distance.

II. RELATED WORK

Acharya MS wrote an editorial, "A Comparison of Regression Models for Prediction of Graduate Admissions." In this, he used numerous fashions like Linear Regression, SVM, and Random woodland and compared their overall performance by computing error capabilities [1]. Narendra Gupta wrote the article "Will I Get in? Modeling the Graduate Admission Process for American Universities," they considered factors such as GPA and other scores and considered the problem a category problem, and used a huge dataset for modeling [2]. The allocated seats can be delivered to a record by the administrator, and the records are stored in the gadget. The universal time for

entrance allotment decreases, and the allotment process becomes faster. It assists college students in making knowledgeable selections about which university to wait for. Students can check in with personal and educational statistics to predict college admission, and administrators can assign seats to college students [3]. Sushrutha Mitra and Soumya Sahoo have published a paper on " A Quality Based Automated Admission System for Educational Domain," where it is treating the hassle as a class problem and predicts whether the scholar will get admitted into that specific college or not [4]

Dr. N. M. Saravana Kumar has applied Artificial Intelligence In Imparting Education, And Evaluating Student Performance, and AI technology have also been used [5]. It also indicates the evaluation of scores versus the hazard of admission so that scholars can apprehend how the threat of admission depends on scores. It also recommends universities where students with similar profiles have a higher percentage of chances.

Nagini Dharani and Sathya Raghava developed a college admissions index that takes details from college students, and tests cut-off scores, and predicts the probability of admission [5]. GRADE uses information beyond acceptance to determine the likelihood of each new

applicant being widely publicized through the panel. It provides for each prediction a score as provided by the human reviewers, along with a description of the candidate characteristics that influenced the prediction [6].

Bootstrapping is used to determine the value of explanatory variables. Notably, the records are free from admission-induced selection bias, which allows us to derive an independent estimate of predictive fees for bachelor's degree measures of graduate-level outcomes. Their results show that university performance can account for 54% of the difference in the overall performance of graduates [7]. Various classifiers are used, and their outputs are then evaluated against measures of accuracy, precision, forgetfulness, F measure, and proximity under the receiver operator curve by addressing whether or not the student assumes the acceptability will be assumed. Consider it accepting a binary classification problem.

III. PROPOSED WORK

In version optimization, the data set is continually divided into 80% and 20% learning and test set. The train set has 400 profiles, and the test set has 100 profiles. The dataset used in the modelling looks like this. Pre-processing is a critical step in

the approach. The purpose is to normalize and group the facts for use in the forecasting rule set. Statistics from Occidental College require some upgrades to be suitable for the proposed machine study algorithms. Determining how to handle lost data is a standard cleaning problem. Since the queried function is a good indicator of the result of the rule set, it is essential to find and discover the missing entries and find a solution based mainly on the variable form, which allows us to view the logs in the use of the version. The recordings are pre-processed and randomly divided into lessons: a training set and a log out set. We selected eighty percent of the 7,976 entries in our data set as our training set.

Table.1 Training Dataset

	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
0	337	118	4	4.5	4.5	0.65	1	0.92
1	324	107	4	4.0	4.5	0.87	1	0.75
2	316	104	3	3.0	3.5	0.90	1	0.72
3	322	110	3	3.5	2.5	0.67	1	0.80
4	314	100	2	2.0	3.0	0.21	0	0.65

The variable to expect is the chance of acceptance. The steps involved in upgrading the version are indicated below. Due to the limitations of university changes from year to year, we have included a fact in the code that the GRE score must be over 250, the TOEFL score must be over 50, and the CGPA must be

over five in all other positions. The education dataset is used to train the model using the Cat. reinforcement algorithm.

PROPOSED ARCHITECTURE

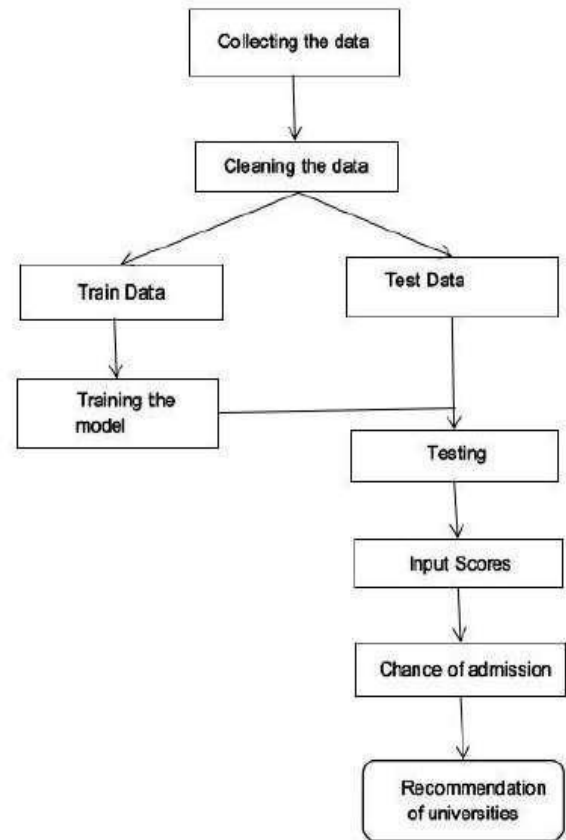


Fig.1 Proposed architecture

Now we can predict the percentage of chances. Recommended universities are also shown where students with similar type of profile having higher chance of prediction.

IV. PERFORMANCE ANALYSIS

Similarly, when the process repeated for the same dataset using catboost without tuning, we got highest accuracy.

Table.1 Base model evaluation

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

Table.2 catboost evaluation (before tuning)

MODEL	Catboost
MAE	0.03
MSE	0.001
R2 Score	0.89
ACCURACY	0.95

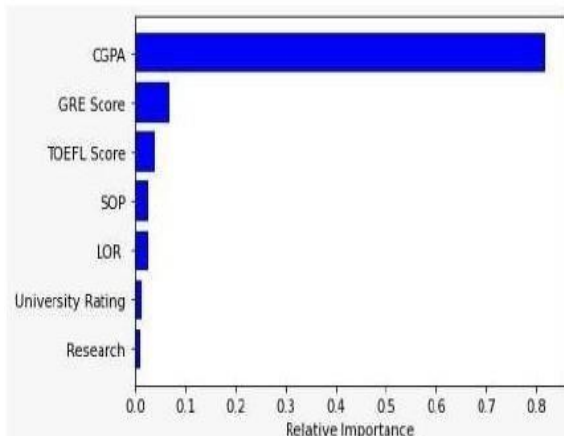


Fig.2 Feature importance

Since CGPA has more than eighty percent importance, it is considered an essential feature. The above determination gives the actual and projected percentage risk of college admission. After running around, the icon looks like this, and the input window looks like this. After introducing the input ratings, we can anticipate the acceptance risk.



Fig.3 predicting the chances of admission

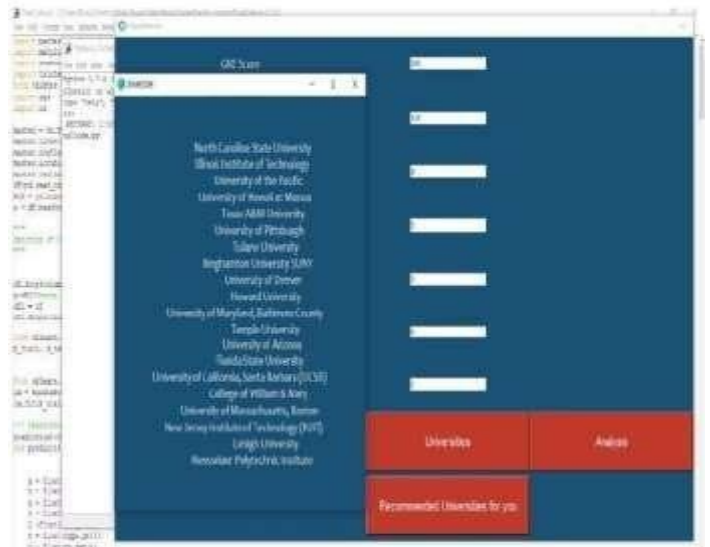


Fig.4 Getting the Universities

We can get university guidance. This can be achieved by searching the profile which is very similar to the specific entry profile and where the risk of entry is greater than the risk of recent entry

And finally, in the analysis. This indicates a relationship between all the skills and assessments students achieve and the number of college students who have a

better chance of being accepted. The relationship between the GPA and the chances of admission.

V. CONCLUSION

This paper discusses the proposed methodology, some algorithms, and their implementation, and additional steps in version training are discussed. Finally, the Catboost rule set has an accuracy of 95 and is the highest quantitative result of a positive incentive expectation release.

REFERENCES

- [1] Acharya MS, Armaan A, Antony AS (2019) A comparison of regression models for predict ionof graduate admissions.
- [2] Prasadu Peddi (2017) “Design of Simulators for Job Group Resource Allocation Scheduling In Grid and Cloud Computing Environments”, ISSN: 2319- 8753 volume 6 issue 8 pp: 17805-17811.
- [3] Roa, Annam Mallikharjuna, et al. “College Admission Predictor.” Journal of Network Communications and Emerging Technologies (JNCET), Volume 8, Issue 4, April (2018).
- [4] Mishra, S. and Sahoo, S. (2016). A Quality Based Automated Admission System for Educational Domain, pp. 221–223, International conference on Signal Processing, Communication, Power and Embedded System (SCOPE)- 2016.
- [5] Kumar, NM Saravana. "Implementation of artificial intelligence in imparting education and evaluating student performance." Journal of Artificial Intelligence 1, no. 01 (2019): 1-9.
- [6] L. Breiman, Accuracy Predictors, Machine Learning, 24(2):123-140,1996.
- [7] Data Cleaning and Analytics, Machine Learning <https://archive.ics.uci.edu/ml/index.php>
- [8] Data Visualizaton, Machine Learning <https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/>
- [9] Jupyter Notebook, Implementing the Algorithms, Machine Learning, <https://jupyter-notebook.readthedocs.io/en/stable/>
- [10] Prasadu Peddi (2022), A Hybrid-Method Neighbor-Node DetectionArchitecture for Wireless Sensor Networks, ADVANCED INFORMATION TECHNOLOGY JOURNAL ISSN 1879-8136, volume XV, issue II.