Prediction Probability of Getting an Admission into

a University using Machine Learning

A.Sivasangari
Dept of IT
Sathyabama Institute of science and Technology
Chennai,India.
sivasangarikavya@gmail.com

V.Shivani
Dept of CSE
Sathyabama Institute of Science and Technology
Chennai,India.
vutla.shivani789@gmail.com

Y.Bindhu
Dept of CSE
Sathyabama Institute of science and Technology
Chennai,India.
yarrambindu1026@gmail.com

D.Deepa
Dept of CSE
Sathyabama Institute of Science and Technology
Chennai,India
deepa21me@gmail.com

R. Vignesh
Dept of CSE
Sathyabama Institute of Science and Technology
Chennai,India
vignesh.cse@sathyabama.ac.in

Abstract—In the present conditions, students regularly have difficulty finding a fitting institution to pursue higher studies based on their profile. There are some advisory administrations and online apps that recommend universities but they ask huge consultancy fees and online apps are not accurate. So, the aim of this research is to develop a model that predict the percentage of chances into the university accurately. This model provides also the analysis of scores versus chance of prediction based on historical data so that students can understand whether their profile is suitable or not. The proposed model uses linear regression and random forest algorithms but cat boost algorithm is giving highest accuracy.

Keywords— Machine learning, Linear regression, Catboost .

I. INTRODUCTION

A person's education plays a very important role in his life because the quality of education decides their future. After they graduate, they often have several doubts that comes in their mind regarding pursuing higher studies and to choose the best university. Most of the students prefer the universities with global recognition. So, higher percentage of students from India prefer united states of America to pursue higher studies. Even though there are universities with good reputation in India, graduate students are facing difficulties to get an admission in good rated universities and also getting placed is quite difficult as the number of working opportunities is quite low. As students are not sure about which university is better, they invest time and money for the guidance. Apart from consultancy offices and advisors, there are some

blogs and websites that encourage and guide the students about the chances of admission but those resources are not very high accurate and cannot completely depend on that.

When filling out admission inquiry forms, educational institutes may use knowledge mining to concentrate on the most relevant details in the data they have collected. It finds information hidden in the data that queries and reports are unable to disclose. This technique must be used to evaluate a collection of trends of students seeking admission in college after collecting data from admission forms filled out by students seeking admission over many years.

This paper establishes a machine learning model, which takes into account boundaries such as GRE Score, TOEFL Score, the University Ranking, the Proposal Statement and the Recommendation Letter Power, the Undergraduate GPA and the Study Experience. After getting all the inputs, it predict the chance of admission. On obscure test occasions, the prepared model has substantial factual findings for the (like) estimate of the probability of confirmation and, accordingly, offers an unprejudiced impression of measurement.

II. RELATED WORK

Acharya MS, he wrote an article called "A Comparison of Regression Models for Prediction

of Graduate Admissions". In this he used various models like Linear Regression, SVM, Random forest and compared the performance of them by computing error functions [1]. Narendhra Gupta wrote 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 as classification problem and used very large dataset for modelling [2]. The allocated seats may be added to a file by the administrator, and the information are stored in the system. The overall time for entrance allotment decreases, and the process of allotment becomes quicker. It assists students in making educated decisions about which college to attend. Students can register with personal and academic information in order to predict admission to schools, and administrators can assign seats to students[3]. Sushrutha mitra and Soumya Sahoo has published a paper on " A Quality Based Automated Admission System for Educational Domain" where it is treating the problem as classification problem and predicts whether the student will get admitted into that particular university or not [4] .

Dr. N. M. Saravana Kumar has implemented Artificial Intelligence In Imparting Education And Evaluating Student Performance, along with AI other technologies were also used [5]. It also shows the analysis of scores versus chance of admission so that the students can understand how chance of admission is depending on scores. It also recommends universities where students with similar type of profiles having higher percentage of chances.

Nagineni Dharani and Sathya Ragava has created college admission predictor where it takes details of students and check the cutoff scores and predict the chance of admission [5]. GRADE uses past admissions data to predict how likely each new applicant is to be accepted by the committee. It gives each prediction a score similar to that provided by human reviewers, along with a description of the applicant characteristics affected the prediction the most [6].

To determine the value of explanatory variables, bootstrapping is used. Importantly, the data is free of admission-induced selection bias, enabling us to derive an unbiased estimation of the predictive

value of undergraduate-level measures for graduate-level results. Their findings show that undergraduate performance can account for 54% of the difference in graduate performance [7]. A variety of different classifiers are used and then evaluated their output by using the metrics of accuracy, precision, recall, F-measure, and area under the receiver operator curve by treating the question of whether a student offered admission would accept it as a binary classification problem [8].
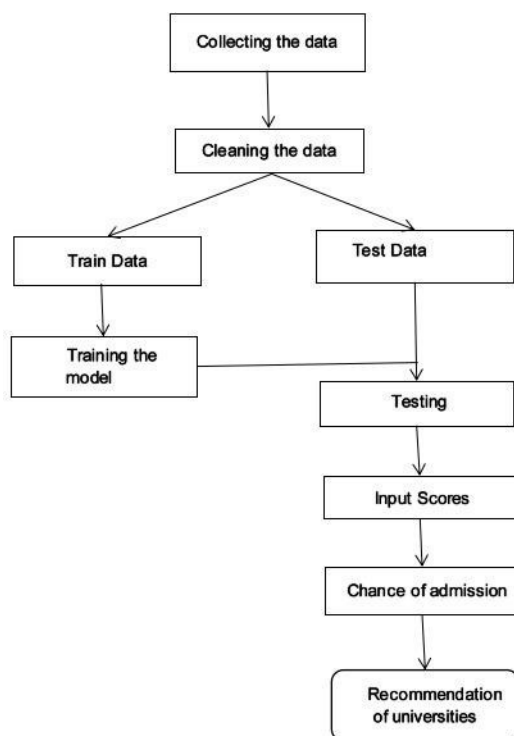
III. PROPOSED WORK

In the model development, the dataset is consistently split into train and test set of 80% and 20%. Train set has 400 profiles and test set has 100 profiles. The dataset used for modelling looks like this. Preprocessing is a crucial step in method. The aim is to clean the data and prepare it for use in a prediction algorithm. Few improvements are required for the data obtained from Occidental College in order to make it suitable for the proposed machine learning algorithms. Determining how to deal with missing data is a common problem in data cleaning. Since the function in question could be a good predictor of the algorithm's outcome, it's critical to find missing entries, locate them, and apply a treatment based on the variable form that enables us to use the data in the model. The data was pre-processed and split into two classes at random: a training set and a testing set. We selected 80 percent of the 7976 entries in our dataset as our training collection.

	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

The variable to be predicted is Chance of Admit. The steps involved in model development are mentioned below. As cutoffs of universities changes year to year, we have put a condition in the code that GRE score should be greater than 250 and TOEFL Score should be greater than 50 and CGPA should be greater than 5 and all other conditions. The training dataset is used training the model using cat boost algorithm

MODEL	Catboost
MAE	0.03
MSE	0.001
R2 Score	0.89
ACCURACY	0.95

Table2:catboost evaluation (before tuning)

Fig3: Feature importance

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

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

Table1:Base model evaluation

Similarly, when the process repeated for the same dataset using catboost without tuning, we got highest accuracy.
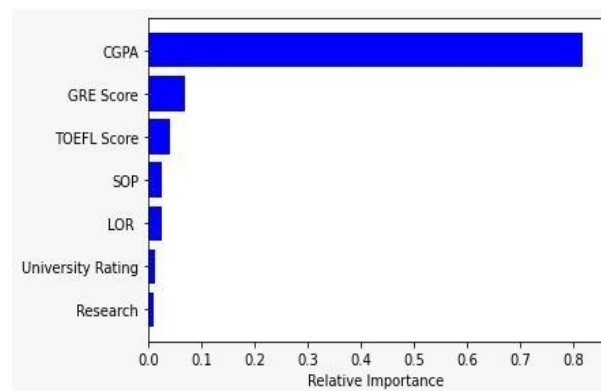
Since CGPA is having more than 80 percentage of importance, it is considered as the most importance feature. The above figure shows actual and predicted percentage of chance of admission into a university. After running the code, the input window looks like this. After giving the input scores, we can predict the chance of admission

Fig6: predicting the chances of admission

Fig7:Getting the Universities

We can get university recommendations. This can be done by searching the profile which is very similar to given input profile and where chance of admission is greater than current chance of admission.
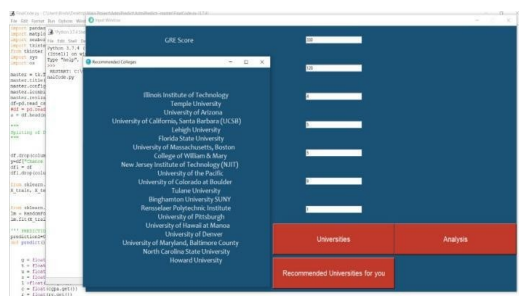

Fig8:Recommendation of Universities

And finally the analysis. this shows relation between all the features and what are the scores that students are getting and what percentage of students are getting higher chance of admission. Relation between CGPA and Chance of Admit.

V CONCLUSION

In this paper, proposed methodology and couple of algorithms and its implementation is discussed and also steps involved in training the model are discussed. Finally, there is an accuracy of 95 with catboost algorithm and is the highest quantitative result of a confirmatory incentive expectation model till now.

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.