

AN APPROACH TO GRADUATE ADMISSION CHANCE PREDICTION USING MACHINE LEARNING

Submitted in partial fulfillment of the requirements
for the award of
Bachelor of Engineering Degree in Computer Science and Engineering

by

YARRAM BINDU (Reg.No.37110867)

VUTLA SHIVANI (Reg.No.37110851)



**DEPARTMENT OF COMPUTER SCIENCE AND
ENGINEERING SCHOOL OF COMPUTING**

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited with Grade "A" by NAAC|12B Status by UGC|Approved by
AICTE
JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119, TAMILNADU.

MARCH - 2021



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited with Grade "A" by NAAC|12B Status by UGC|Approved by
AICTE
JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119, TAMILNADU.
www.sathyabama.ac.in



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDECERTIFICATE

This is to certify that this Project Report is the bonafide work of **YARRAM BINDU (Reg.No 37110867), VUTLA SHIVANI (Reg.No 37110851)** who carried out the project entitled "**AN APPROACH TO GRADUATE ADMISSION CHANCE PREDICTION USING MACHINE LEARNING**" under our supervision from **December 2020 to March 2021**.

Internal Guide

Dr.A.Sivasangari, M.E.,Ph.D.,

Head of the Department

Submitted for Viva voce Examination held on _____

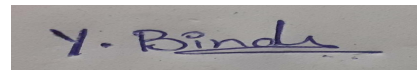
Internal Examiner

External Examiner

DECLARATION

I **YARRAM BINDU** (Reg.no 37110867) here by declare that the Project Report entitled "**AN APPROACH TO GRADUATE ADMISSION CHANCE PREDICTION USING MACHINE LEARNING**" done by me under the guidance of **Dr.A.Sivasangari, M.E.,Ph.D.**, Department of Computer Science and Engineering at Sathybama Institute of Science and Technology is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Computer Science and Engineering.

DATE:

A rectangular box containing a handwritten signature in blue ink that reads "Y. Bindu".

PLACE:CHENNAI

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management of Sathyabama Institute Of Science and Technology** for their kind encouragement in doing this project and for completing it successfully.

I convey my thanks to **Dr.T.Sasikala,ME.,Ph.D., Dean**, School of Computing and **Dr.S.Vigneswari,M.E.,Ph.D., and Dr.L.Lakshmanan, M.E., Ph.D., Heads Of the Department** of Computer Science and Engineering for providing me necessary Support and details at the time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to My Project Guide **Dr.A.Sivasangari, M.E.,Ph.D.**, for his valuable guidance, Suggestions and constant encouragement paved way to the successful completion Of my project work.

I wish to express my thanks to all Teaching and Non-Teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of project.

ABSTRACT

In today's era we see a lot of students pursuing their education away from their home countries. The main country targeted by these international students is The United States of America. Majority of the international students in the United States of America are from India and China. In the past decade the number of Indian students pursuing post graduate education from the USA has rapidly increased. With the increase in the number of international students studying in the USA, each applicant has to face a tough competition to get admission in their dream university. Generally as the students don't have much idea about the procedures, requirements and details of the universities in the USA they seek help from the education consultancy firms to help them successfully secure admission in the universities which are best suitable for their profile, for this they have to invest huge amount of money as consultancy fees. Apart from these the education consultancy firms there are few websites and blogs that guide the students on the admission procedures. The drawback of the currently available resources is that they are very limited and also they are not truly dependable taking into consideration of their accuracy and reliability. The aim of this research is to develop a system using machine learning algorithms, we will name it as Student Admission Predictor (SAP). It will help the students to identify the chances of their application to an university being accepted. Also it will help them in identifying the universities which are best suitable for their profile and. A simple user interface will be developed for the users to access the SAP system.

Keywords: Student Admission Predictor; Machine Learning; Data Mining.

TABLE OF CONTENTS

CHAPTER NO	CHAPTER NAME	PAGE NO
1	INTRODUCTION	1
2	LITERATURE SURVEY	4
3	AIM AND SCOPE OF THE PRESENT INVESTIGATION	8
	3.1 RESEARCH QUESTION	8
	3.2 METHODOLOGY	8
	3.3 RESEARCH SIGNIFICANCE	9
	3.4 RESEARCH LIMITATIONS	9
4	EXPERIMENTAL OR MATERIALS AND METHODS	10
	4.1 DATA-SET	10
	4.2 DATA-SET EXTRACTION AND TRANSFORMATION	10
	4.3 ALGORITHMS	11
	4.4 ARCHITECTURE	12
5	RESULTS AND DISCUSSION, PERFORMANCE ANALYSIS	14
	5.1 PERFORMANCE	14
	5.2 RESULTS	15
	CONCLUSION AND FUTURE WORK	19
	REFERENCE	20

LIST OF FIGURES

FIGURE NO	FIGURE NAME	PAGE NO
3.1	Flow Chart	9
4.1	Data-Set	10
4.2	Architecture	13
5.1	Graph	15
5.2	User Interface	15
5.3	Predicting the chance of Admission	16
5.4	List of Universities	17
5.5	Recommended Universities	17
5.6	Analysis	18

LIST OF TABLES

TABLE NO	TABLE NAME	PAGE NO
5.1	Linear Regression vs Catboost	14

LIST OF ABBREVIATIONS

ABBREVIATION	EXPANSION
GRE	General Record Examination
TOEFL	Test Of English as a Foreign Language
IELTS	International English Language Testing System
LOR	Letter Of Recommendation
SOP	Statement Of Purpose
SAP	Student Admission Predictor
SAT	Scholastic Assessment Test
CRISP	CRoss-Industry Standard Process
CGPA	Cumulative Grade Points Average
GPA	Grade Point Average
OLS	Ordinary Least Squares
GUI	Graphical User Interface

1. INTRODUCTION

A persons education plays a vital role in their life. While planning for education students often have several questions regarding the courses, universities, job opportunities, expenses involved, etc. Securing admission in their dream university is one of their main concerns. It is seen that often students prefer to pursue their education from universities which have global recognition. And when it comes to international students the first choice of the majority of them is the United States of America. With the majority of worlds highly reputed universities, wide range of courses offered in every sector, highly accredited education system and teaching, scholarships provided to students, best job market and many more advantages make it the dream destination for the international students. According to research, there are above 8 Million international students study- ing in more than 1700 public and 2500 private universities and colleges across the USA. (MasterPortal (2017))

The majority of international students studying in the USA are from India and China. In the past decade, India has seen a huge increase in the number of students opting to pursue their education from foreign universities in countries like The USA, Ireland, Australia, Germany, etc. Although there are significant universities and colleges in India, students are finding it difficult to get admission in the highly ranked colleges and also getting a job is a challenge as the ratio of number students to the number work opportunities available is quite high. India is one of the leading counties in the number of software engineers produced each year; it becomes tough for the students to find jobs in elite companies due to high competition. This motivates a good number of students to pursue post-graduation in their field. It is seen that the number of students pursuing Masters in Computer Science field from universities in the USA is quite high; the focus of this research will be on these students.

(Times (2015))

Majority of universities in the USA follow similar guidelines for providing admission to students. Universities take into consideration different factors

like score on aptitude based examination like the General Record

Examination (GRE), command over the English language is judged based on their score in English competency test like Test Of English as a Foreign Language (TOEFL) OR International English Language Testing System (IELTS), their work experience in same or other fields, the quality of the Letters Of Recommendation (LOR) and the Statement Of Purpose(SOP) documents provided by the student etc. Based on the overall profile of the student decision is taken by the universities admission team to admit or reject a particular candidate.

Every candidate has to take all the required examination and build a strong profile to secure admission in their dream universities in the USA. Once the candidates have made their profile ready, they apply to the universities where they aim to secure admission. The students have to shortlist the universities which are best known for the courses they are looking for and also they should have an idea about their chances of securing admission in those universities based on their profile. This task of shortlisting the universities where the student has high chances of admission is difficult for mainly for the international students, so they end up with applying to many universities in hopes of getting admission in few of them thus investing an extra amount of money in the applications. There are several portals and websites which provide information and help to students in shortlisting the universities, but they are not reliable. Most of the students don't take the risk of evaluating the colleges by themselves, and they seek the help of the education consultancy firms to do it for them. Again for this students have to pay a huge amount of fee to the education consultant.

The primary objective of this research is to develop a system to solve the problems the international students are facing while applying for universities in the USA. We will be developing a Student Admission Predictor (SAP) system which will help the students to predict the chances of their application being selected for a particular university for which they wish to apply based on their profile. Also, the system will provide a

recommendation of universities to the student to which the student has a high possibility of getting admission. Multiple machine learning classification algorithms were evaluated develop the system. Finally, Linear Regression and Decision Tree algorithms were used as they were found to be the best fit for the system developed. Also, we will be creating a simple user interface which will help the users to input the data related to student profile and get the admission chance. This research will thus eventually help students saving the extra amount of time and money they have to spend at the education consultancy firms. And also it will help them to limit their number of application to a small number by proving them the suggestion of the universities where they have the best chance of securing admission thus saving more money on the application fees. Today all the work at the time of admission of the students is done manually by ink and paper, which is very slow and consuming much efforts and time. In the modern world of technology, computer are affecting our lives in more ways than we probably are aware of computerized management maintaining information of an educational institute, colleges, other the list is endless. The main principle behind the need of college admission system is easy supervision of institutes. It can handle the details of students such as fee details or marks details. This Student Database has been designed taking into account the practical needs to manage a Students data. Moreover, it provides security at product level as well as user level. This Database follows a typical event flow seen in such a system. The design and implementation of a comprehensive student information system and user interface is to replace the current paper records. This system is intended for communication purpose between users of academic institutions. This system helps the administrator to easy access the information of students. This system is also helpful for the administrator because he/she can easily bring changes to the records of the students.

2. LITERATURE SURVEY

This section provides the literature review of the work that has previously done on predicting the chances of students enrolment in universities. There have been several project and studies performed on topics related to students admission into universities. (Bibodi et al. (n.d.)) used multiple machine learning models to create a system that would help the students to shortlist the universities suitable for them also a second model was created to help the colleges to decide on enrolment of the student. Nave Bayes algorithm was used to predict the likelihood of success of an application, and multiple classification algorithms like Decision Tree, Random Forest, Nave Bayes and SVM were compared and evaluated based on their accuracy to select the best candidates for the college. Limitation of this research as that it did only relied on the GRE, TOEFL and Undergraduate Score of the student and missed on taking into consideration other important factors like SOP and LOR documents quality, past work experience, technical papers of the studentsetc.

Bayesian Networks were used by (Thi et al. (2007)) to create a decision support sys- tem for evaluating the application submitted by international students in the university. This model was designed to predict the performance of the aspiring students by compar- ing them with the performance of students currently studying in the university and had similar profile during their application. In this way based on the current students profile the model predicted whether the aspiring student should be granted admission to the university. Since the comparisons were made only with the students who were already admitted in the university and the data of the students who were denied admission were not included in the research this model proved to be less efficient due to the problem of class imbalance.

(Abdul Fatah S; M (2012)) developed a model that can provide the list of universities/colleges where the which best suitable for a student based on their academic records and college admission criteria. The model was

developed by applying data mining techniques and knowledge discovery

rules to the already existing in-house admission prediction system of the university. (Mane (2016)) conducted a similar research that predicted the chance of a student getting admission in college based on their Senior Secondary School, Higher Secondary School and Common Entrance Examination scores using the pattern growth approach to association rule mining. The performance of both the models was good the only drawback was the problem statement was single university centric.

(Mishra and Sahoo (2016)) conducted a research from a university point of view to predict the likelihood of a student enrolling in the university after they have enquired about courses in the university. They used K-Means algorithm for clustering the students based on different factors like feedback, family income, family occupation, parents qualification, motivation etc. to predict if the student will enroll at the university or not. Depending upon the similarity of the attributes among the students they were grouped into clusters and decisions were made. The objective of the model was to increase the enrolment of the students in the university.

(Eberle et al. (n.d.)) used machine learning and predictive modelling to develop a model that to evaluate the admission policies and standards in the Tennessee Tech University. A well known version of the C4.5 algorithm, J48 was used to create the model. Like the models mentioned above they used the different factors of the student profile to evaluate the chances of their admission in the university. The model worked well in predicting the true positive scenarios where the student was had good profile to secure the admission, but it failed in efficiently identifying the true negatives because of which student who does not satisfy the defined criteria.

In research conducted by (Jamison (2017)) the yield of college admission was predicted using machine learning techniques. Yield rate can be defined as the rate at which the students who have been granted admission by the university actually enrol for the course. Multiple machine learning algorithms like Random Forest, Logistic Regression and SVM were used to create the

model; the models were compared based on their performance and accuracy,

Random Forest outperformed the other models with 86% accuracy and was thus used to create the system. The factors that proved to be significant in predicting successful application were also highlighted.

H.Gurler and A.Istanbullu has performed some research on student performance using knowledge discovery in higher educational systems but his model do not predict accurately. H.Aldowa has done educational data mining for 21st century but his algorithm may vary from year to year.

Al-Qahtani has done Comparisons on predictive values of admission criteria for academic achievement among undergraduate students of health and non-health science professions but he is focusing on only one particular field. Acharya suggested the following strategy in constructing four model AI recurrent recurrence: direct re occurrence, upholding vector machine, selection tree, and random backwoods for advanced graduate validation study.

In Guptaetal, an emotional supporting selection network was established by selecting machines to anticipate graduate confirmation in the United States by taking into account such boundaries such as state permits, GPA and the prestige of institutes.Ghai has created an American model of verifying alumni's preferences, allowing students to select a well-suited institution, forecasting whether or not it is given.In relation to graduate affirmation.GRADE system was developed by (Waters and Miikkulainen (2013)) to support the admission process for the graduate students in the University of Texas Austin Department of Computer Science.

The main objective of the project was to develop a system that can help the admission committee of the university to take better and faster decision.Linear Regression and Catboost were used to create the model, both models performed equally well and the final system was developed using Catboost due to its simplicity. The time required by the admission committee to review the applications was reduced by 74% but human intervention was required to

make the final decision on status if the application.(Nandeshwar et al. (2014)) created a similar model to predict the enrolment of the student in the university based on the factors like SAT (Scholastic Assessment Test)score, GPA score, residency race etc. The Model was created using the catboost algorithm, it was able to achieve accuracy rate of 95%

3 AIM AND SCOPE OF THE PRESENT INVESTIGATION

3.1 RESEARCH QUESTION

Can machine learning algorithms improve the accuracy in predicting the application status of students aspiring to enrol for Masters in Computer Science course at universities in the USA?

3.2 METHODOLOGY

CRoss-Industry Standard Process (CRISP) methodology (Azevedo, 2008) was followed in this research.

Business Understanding: Initially good amount of time was spent on understanding the problem statement by understanding the concerns of students regarding the current application process, the objectives of the research were defined in this process.

Data Understanding: Data required for the research was collected from multiple data sources. Different features of the data were analyzed based on their importance and relevance. Data-set would be explained in more detail further.

Data Preparation: In this phase, the data from multiple data sources were integrated into a final data-set. Further the data was cleaned by removing unwanted columns, performing transformation and cleaning activities on the data.

Modelling: Multiple machine learning models were developed to predict the likelihood of success of the student's application in a particular university. The user interface was developed to allow the users to access these models.

Evaluation: Models developed were evaluated based on their performance and accuracy. More information will be presented in the evaluation section of the paper.

Deployment: Once the models were evaluated they were integrated with code developed for user interface using the Shiny package in R.

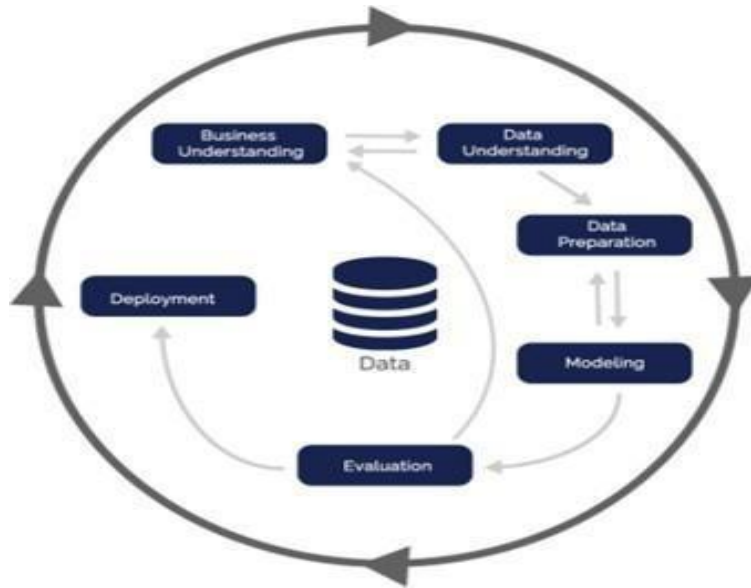


Figure 3.1: FLOW CHART

3.3 RESEARCH SIGNIFICANCE

The principal objective of the research is to help the students who are aspiring to pursue their education in the USA. The SAP system will help them to evaluate the chances of success in a particular university without being dependent on any education consultancy firm. It will help them in saving a huge amount of time and money spent in the application process.

Also, it will help them to limit the number of applications made by the students by suggesting them the best universities where they have high chances of securing admission thereby by saving the amount of money spent by the students by applying in universities where they have less chance to secure admit based on their profile

3.4 RESEARCH LIMITATIONS

Student Admission Predictor system will only take into consideration the data related to the Indian students pursuing Masters in Computer Science from universities in the USA.

4. EXPERIMENTAL OR MATERIALS AND METHODS

4.1 DATA-SET

This section describes, in brief, the data that has been used for the research. Data from multiple sources was used in this project, the major amount of data was extracted from public website Yocket(Yocket), data regarding the rankings, fees and enrolment in colleges was obtained from a leading educational consultancy firm The Mentors Circle in India. Data from both the sources was integrated together to form a staging data-set. For predicting the chance of a student getting shortlisted in universities the final data-set was divided into multiple data-sets each representing a particular university. For predicting the list of universities suitable for students based on their profile data of all the students the staging data-set was updated only to have records of students who had successfully secured admission in the universities. Below figure shows the different features of the data-sets.

	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 4.1: Data Set

4.2 DATA-SET EXTRACTION AND TRANSFORMATION

Data related to the college ranking was collected in .csv format, the data related to students profile was extracted from (Yocket (2017)) using data extraction tool provided by (Mozenda (n.d.))in .csv files. Data being from public portal had multiple records with missing and irrelevant values; data cleaning was performed in Microsoft Excel by deleting the records having unwanted and missing values. Unwanted columns were removed from the data-set. Once the data-set was cleaned data was transformed to be suitable for the model. The original data-set had TOEFL or IELTS score as a representation of language, to have a consistent metrics for the language score of all the records were converted to IELTS scale using the conversion

table.(Prescholar (2017)). Similarly, the GPA score of the students was represented in terms of percentage and CGPA; all the records of percentage were converted to CGPA by multiplying percentage score by 9.5.

4.3 ALOGORITHMS

Multiple machine learning algorithms were used for this research, Linear Regression, Random forest, Catboost algorithms were used to predict the likelihood of the students getting admission into university based on their profile. Catboost algorithm was used to predict the rank of the college that would be suitable for the students based on their profile and suggest the list of universities accordingly.

Linear Regression: It is an algorithm which is used widely for regression problems. Due to its simplicity and effectiveness, it is easy to implement and understand. It is a supervised machine learning algorithm that uses available data to create the model and further that model can be applied to predict the target variable. Linear regression quantifies the relationship between one or more predictor variable(s) and one outcome variable. Linear regression is commonly used for predictive analysis and modeling. Linear regression is also known as multiple regression, multivariate regression, ordinary least squares (OLS), and regression. Linear regression with a single predictor variable is known as simple regression. In real-world applications, there is typically more than one predictor variable. Such regressions are called multiple regression

Random Forest: Random Forest is a powerful and versatile supervised machine learning algorithm that grows and combines multiple decision trees to create a “forest.” It can be used for both classification and regression problems in R and Python. As we know, the Random Forest model grows and combines multiple decision trees to create a “forest.” A decision tree is another type of algorithm used to classify data. In very simple terms, you can think of it like a flowchart that draws a clear pathway

to a decision or outcome; it starts at a single point and then branches off

into two or more directions, with each branch of the decision tree offering different possible outcomes.

Catboost : CatBoost is an algorithm for **gradient boosting on decision trees**. It is developed by Yandex researchers and engineers, and is used for search, recommendation systems, personal assistant, self-driving cars, weather prediction and many other tasks at Yandex and in other companies, including CERN, Cloudflare, Careem taxi. Main advantages of CatBoost are Superior quality when compared with other GBDT libraries on many datasets and Best in class prediction speed. CatBoost originated in a Russian company named Yandex. It is one of the latest boosting algorithms out there as it was made available in 2017. There were many boosting algorithms like XGBoost, LightGBM etc. but none comes close to the CatBoost for several reasons. CatBoost means Categorical Boosting because it is designed to work on categorical data flawlessly, If you have Categorical data in your dataset

Tkinter Library: Tkinter is the most commonly used library for developing GUI (Graphical User Interface) in Python. It is a standard Python interface to the Tk GUI toolkit shipped with Python. As Tk and Tkinter are available on most of the Unix platforms as well as on the Windows system, developing GUI applications with Tkinter becomes the fastest and easiest. Tkinter is an inbuilt Python module used to create simple GUI apps. It is the most commonly used module for GUI apps in the Python. You don't need to worry about installation of the Tkinter module as it comes with Python default.

4.4 ARCHITECTURE

In this section, we will describe the architecture of the Student Admission Predictor system. The student will enter his/her profile details using the user interface developed in shiny. The user interface code will interact with the catboost models to provide the users with the required result.

This algorithm will be used to determine the chance of the student of

- securing admission in a particular university based on his/her profile and

will be used to determine the rank of college to which is most suitable for the student based on his/her profile and provide the student with the list of universities which fall in that rank. Once the model have been executed the result will be provided to the student as the output on the user interface. After giving the chance of admission, the universities function gives the list of universities under that rating and recommendation of universities gives the universities that are relevent to the profile and the analysis function gives the analysis of the model. Below is the flowchart that explains steps that are involved in model building. The input is the collected data and output is the chance of admission into particular universities of given input profile.

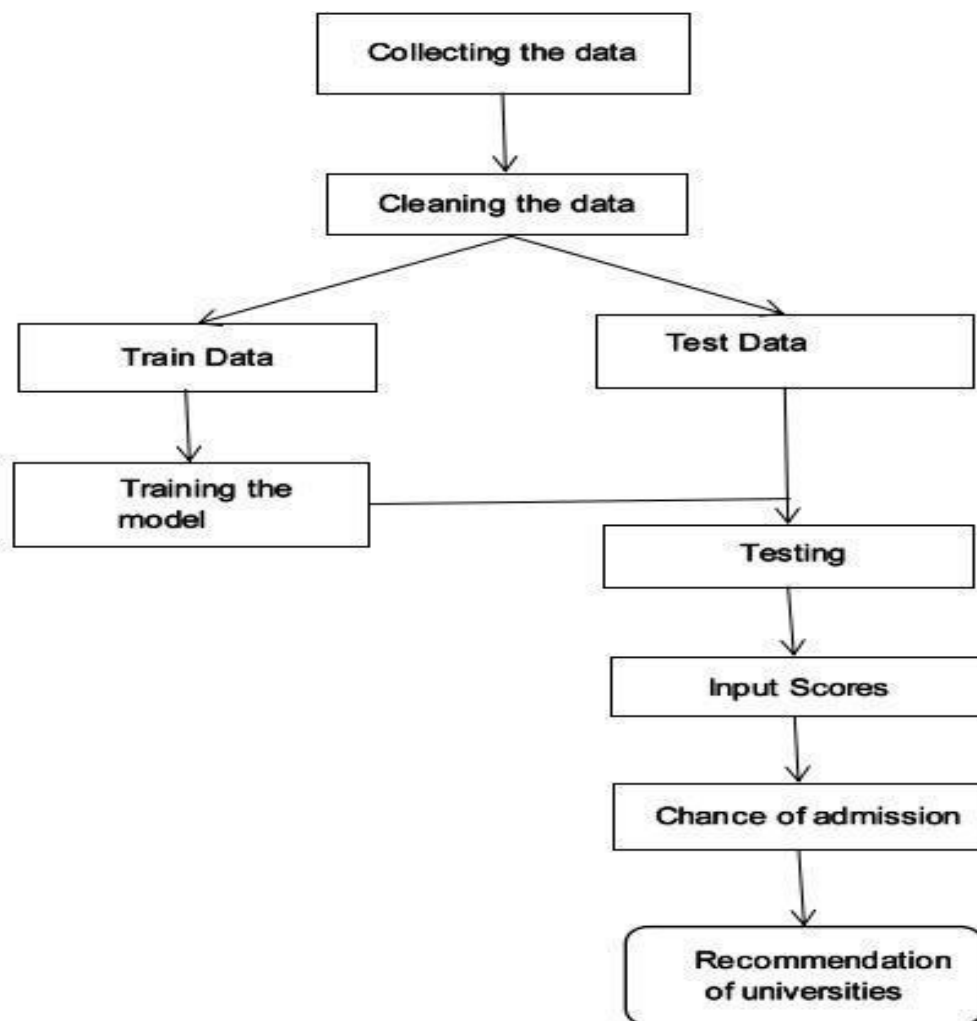


Figure 4.2: Architecture

5.RESULTS AND DISCUSSION, PERFORMANCE ANALYSIS

5.1 PERFORMANCE

Linear Regression and Catboost algorithms were used to create a model that can be used to predict the likelihood of success of a students application to the university based on his/her profile. Both algorithms were tested and their performance was evaluated based on different factors like Mean absolute Error, Mean Square Error, R2 Score, Accuracy value. As can be seen the table given below model created using Catboost outperformed the model created using Linear Regression on all the performance measures. Also by looking at the variance in the values of the data Catb seemed to be the best-fit algorithm to create the Student Admission Predictor System.

	Linear Regression	Catboost
Mean Absolute Error	0.04	0.03
Mean Square Error	0.003	0.001
R2 Score	0.84	0.89
Accuracy	0.93	0.95

Table 5.1: Linear Regression vs Catboost

Accuracy was considered to be main metric to evaluate the performance of the models, as the data used for creating the models was balanced. Also, prediction of the true positive and true negative scenarios was equally equivalent. The Linear Regression performed well with an overall average accuracy of 93%. The Catboost model which was created to predict the chance of admission of the universities suitable for the student provided the result with an accuracy of 95%. Below figure shows actual and predicted values. As we can see they are very closer.

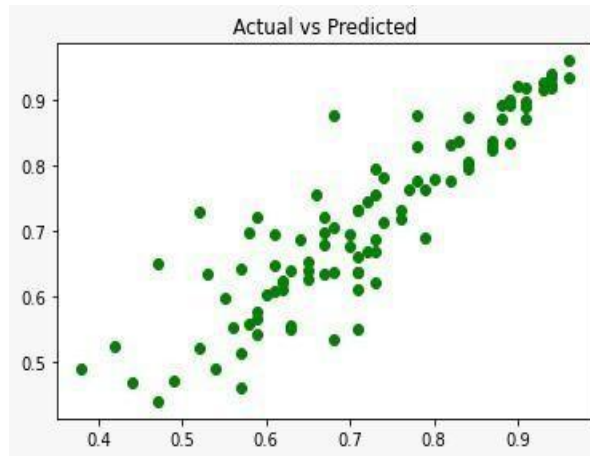


Fig 5.1: Graph.

5.2 RESULTS

Below is the screen-shot of the user interface before submitting the student profile.

The screenshot shows a Python application window titled "FinalCode.py - C:\Users\Bindu\Desktop\Main Project\AdmiPredict-master\FinalCode.py (3.7.4)". The interface is divided into two main sections: an input area and a result area. The input area has a dark blue background and contains seven white input fields for the following attributes: GRE Score, TOEFL Score, University Rating, SOP Rating, LOR Rating, CGPA, and Research. Below these fields are three red buttons labeled "Predict", "Universities", and "Analysis". The result area, located at the bottom, has a dark blue background and contains a red button labeled "Recommended Universities for you". The Python code in the background shows the use of pandas, matplotlib, seaborn, and sklearn libraries, along with a Random Forest model for prediction.

Figure 5.2: User interface

When a student with top grades evaluates his/her profile for admission, After clicking predict button the model predicts the chance of admission with that university rating

Below is the screen-shot of the user interface after submitting the student profile. Predict button predict the chance of admission in percentage.

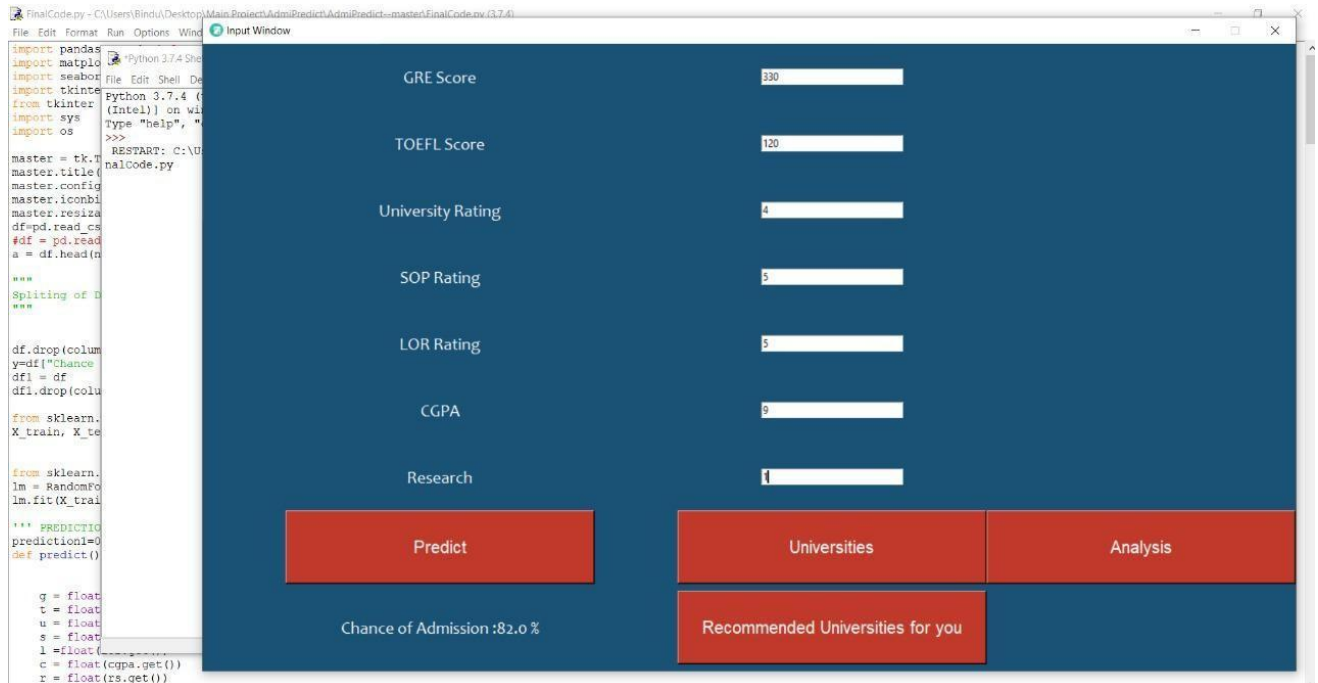


Figure 5.3: Predicting the chance of admission

When a student wants to check the list of universities under that rating, he can use universities button. It gives the list of universities under that rating in USA. A new window will be opened and the universities will be shown.

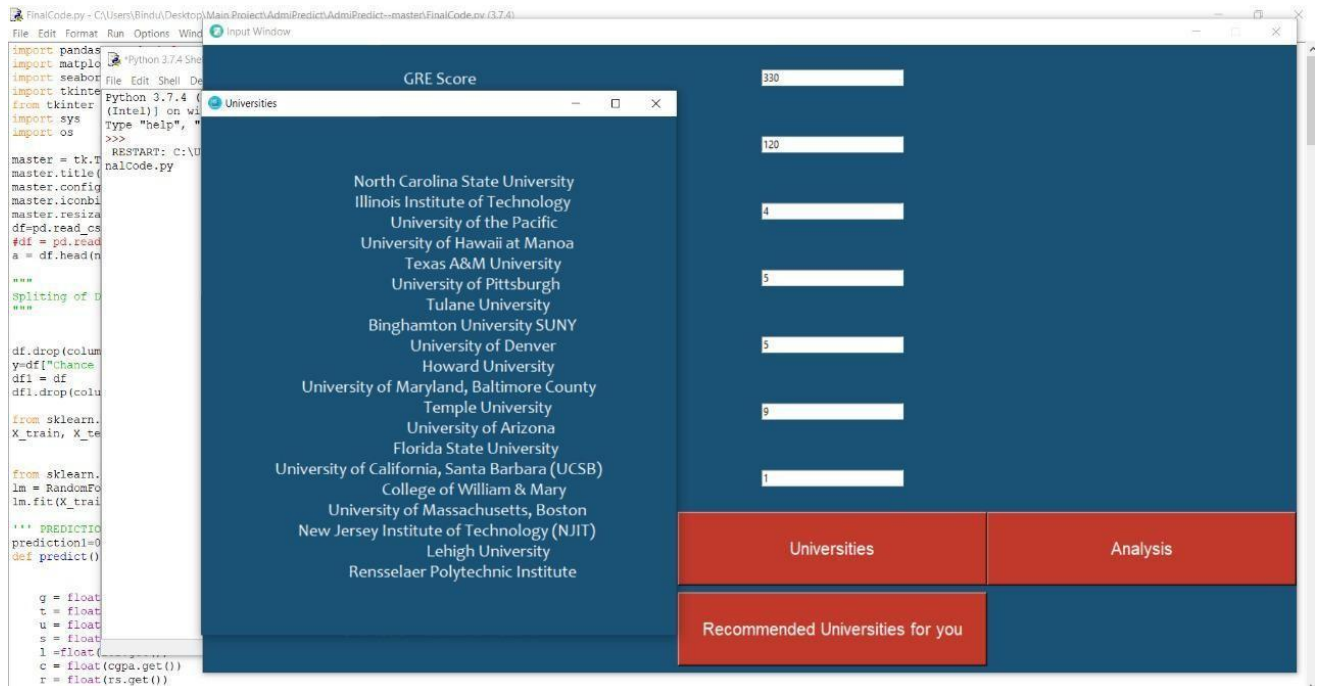


Figure 5.4: List of Universities

When a student wants to check recommended universities for him then he should use Recommended universities for you button. It searches for the similar type of profile in the dataset and give the universities that the student get higher chance of admission.

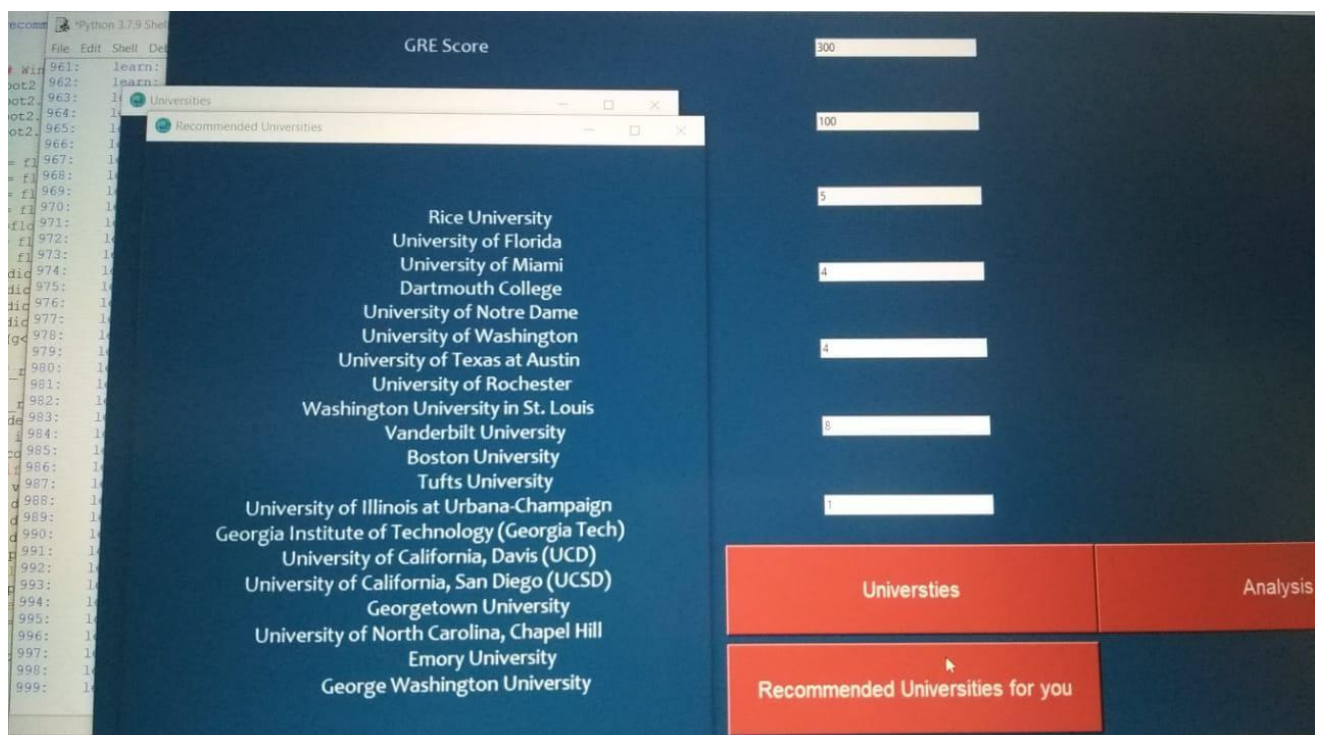


Figure 5.5: Recommended universities

Finally if a student wants to know which score is having highest importance for the highest chance of admission and the range of scores that students are having .

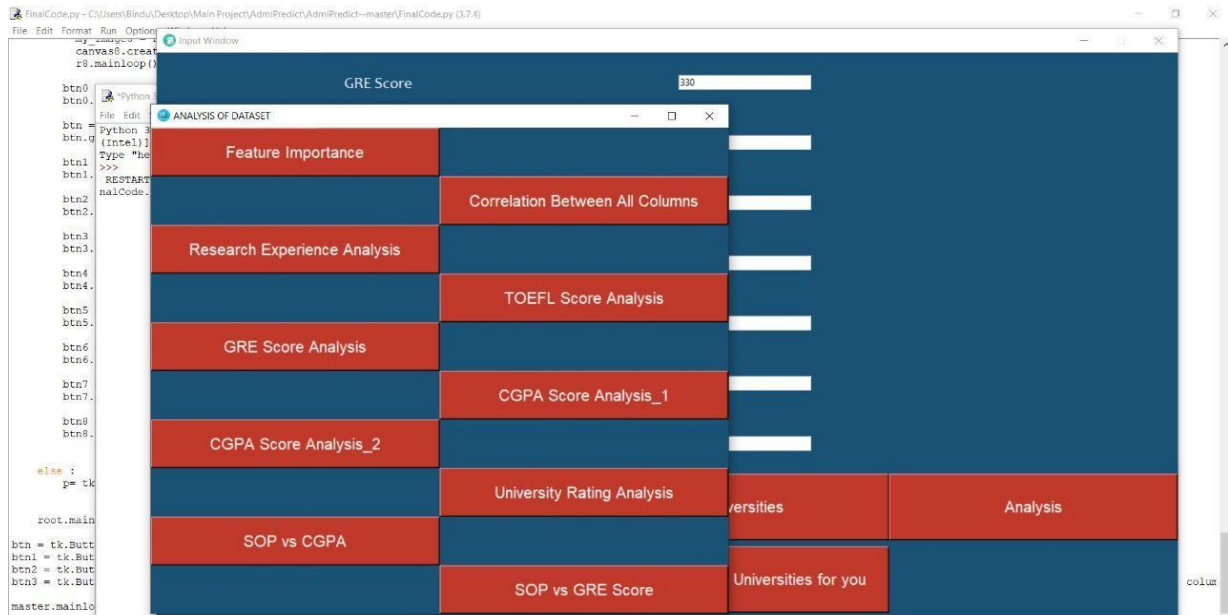


Figure 5.6: Analysis

CONCLUSION AND FUTURE WORK

The main objective of this research was to develop a prototype of the system that can be used by the students aspiring to pursue their education in the USA. Multiple machine learning algorithms were developed and used for this research. KNN proved to best-fit for development of the system when compared with the Logistic regression model. The model can be used by the students for evaluating their chances of getting shortlisted in a particular university with an average accuracy of 75%. Decision Tree algorithm was used to predict the universities which were best suitable for a student based on their profile. The decision tree algorithm proved to be 80% accurate. A simple user interface was developed to make the application interactive and easy to use for the users from the non-technical background. Shiny library from python was used to create the user interface. The overall objective of the research was achieved successfully as the system allow the students to save the extra amount of time and money that they would spend on education consultants and application fees for the universities where they have fewer chances of securing admission. Also, it will help the students to make better and faster decision regarding application to the universities.

REFERENCES

- [1]. Abdul Fatah S; M, A. H. (2012). Hybrid Recommender System for Predicting College Admission, pp. 107–113.
- [2]. Bibodi, J., Vadodaria, A., Rawat, A. and Patel, J. (n.d.). Admission Prediction System Using Machine Learning.
- [3]. Eberle, W., Simpson, E., Talbert, D., Roberts, L. and Pope, A. (n.d.). Using Machine Learning and Predictive Modeling to Assess Admission Policies and Standards.
- [4]. Jamison, J. (2017). Applying Machine Learning to Predict Davidson College ' s Admissions Yield, pp. 765–766.
- [5]. Mane, R. V. (2016). Predicting Student Admission decisions by Association Rule Mining with Pattern Growth Approach, pp. 202–207.
MasterPortal (2017). MasterPortal.
- [6]. Mishra, S. and Sahoo, S. (2016). A Quality Based Automated Admission System for Educational Domain, pp. 221–223.
Mozenda (n.d.). Mozenda.
- [7]. Nandeshwar, A., Chaudhari, S., Sampath, V., Flagel, A., Figueroa, C., Sugrue, P., Ahlburg, D. and Mcpherson, M. (2014). Predicting Higher Education Enrollment in the United States : An Evaluation of Different Modeling Approaches, *International Journal of Operational Research* **19**(26): 60–67.
- [8]. Thi, N., Hien, N. and Haddawy, P. (2007). A Decision Support System for Evaluating International Student Applications, pp. 1–6.
Times, T. E. (2015). The Economic Times.
- [9]. Acharya MS, Armaan A, Antony AS (2019) A comparison of regression models for prediction of graduate admissions. In: 2019 .
- [10]. Gupta N, Sawhney A, Roth D (2016) Will I get in? modeling the graduate admission process for American universities. In: 2016 .