**Summer Project Report on**

**“**STUDENT ADMISSION PREDICTOR**”**

**By**

Keval Joshi (501827)

Chirag Jawale (501825)

Aditya Sawant (501860)

Sriansh Sahu (501859)





**Department of Information Technology**

**Fr. Conceicao Rodrigues Institute of Technology**

Sector 9A, Vashi, Navi Mumbai – 400703

University of Mumbai

2019-2020

TABLE OF CONTENTS

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sr. No.** |  | | **Topic** | **Page No.** |  |
|  |  | | |  |  |
|  | Introduction | | |  |  |
|  | 1.1 | Motivation | |  |  |
| 1 | 1.2 | Problem Definition | | 5 |  |
| 1.3 | Scope | |  |
|  |  |  |
|  | 1.4 | Objectives | |  |  |
|  | 1.5 | Limitations | |  |  |
|  |  | | |  |  |
| 2 | Literature Survey  2.1 Literature Survey | | | 8 |  |
|  |  | | |  |  |
| 3 | System Design  3.1 Flow of Website | | | 11 |  |
|  | 3.2 Insight on the ML Algorithms | | |  |  |
|  |  | | |  |  |
|  | Implementation | | |  |  |
| 4 | 4.1 Implementation of ML Model | | | 16 |  |
|  |  | | |  |  |
|  |  | |  |  |  |

II

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sr. No.** |  | **Topic** | **Page No.** |  |
|  |  | |  |  |
| 5 | Implementation Details   * 1. User Interface | | 21 |  |
|  |  | |  |  |
| 6 | CONCLUSION/FUTURE SCOPE | | 24 |  |
|  |  | |  |  |
| 7 | Appendix: Code Samples | | 26 |  |
|  |  | |  |  |
| 8 | Acknowledgement | | 29 |  |
|  |  | |  |  |
| 9 | Bibliography | | 31 |  |
|  |  | |  |  |
| 10 | Getting Things to work |  | 33 |  |

**Abstract**

In today’s era we see a lot of students pursuing their education away from their home countries. Majority of the international student in top Universities are from India. In the past decade the number of Indian students pursuing post graduate education from abroad has rapidly increased. With the increase in the number of international students studying abroad, each applicant has to face a tough competition to get admission in their dream university. Generally, as the students do not have much idea about the procedures, requirements and details of the universities overseas 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 a university being accepted. Also it will help them in identifying the universities which are best suitable for their profile and also provide them with the details of those universities. A simple user interface will be developed for the users to access the SAP system.

Since foreign universities are expensive it is not possible for all the students to afford its expenses. Therefore, a majority of students in India aim to get admission in top Engineering colleges of India. Since there is a lot of competition it gets difficult for students to figure out if they will end up getting admission in the desired university. This projects aims to aid them the best university according to their rank in competitive exam using Machine Learning algorithms. A simple interface will be developed for the users to access the SAP.

**CHAPTER-1**

**INTRODUCTION**

**MOTIVATION:**

A person’s 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. 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 markets and many more advantages overseas universities are the dream destination for the international students.

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 countries 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.

For those appearing for admissions in India the level of competition is enormous. According to research every year an estimated of one million students appear for the JEE exam to end up into the prestigious IITs around the country. However due to limited seats the ratio of students getting selected are very low. Therefore, our system aims to provide predictions such that students according to the rank secured can draw predictions for their dream universities across the country.

**PROBLEM DEFINITION:**

Due to numerous applications being forwarded to universities it is often difficult to find out the chances of the application being selected. As a result of waiting for long time for the response from the universities the students tend to miss the application dates of various other universities accepting admissions at the same time. Many consultancy companies and firms tend to take advantage of this and costs students a lot of money. This issue has created a need to develop a platform where students can find out the chances of their applications being selected.

**SCOPE:**

The primary objective of this project is to develop a system to solve the problems the students are facing while applying for universities. 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.

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.

**OBJECTIVES:**

* Designing an interface where students can enter their details.
* Predicting the chances of admission according to the input provided by the user.
* Suggesting the names of Universities that the student is eligible to apply and can easily get admission.
* Helping students save application fees of various universities that might reject their application.
* Helping students choose the right University before the deadline of the application.

**LIMITATIONS:**

The website currently has limited number of suggestions for the university since it is difficult to handle large amounts of data however the recommendations can be increased.

The predictions for Universities in India are restricted to the data obtained from the JEE ranks as data for other Universities was not available. Predictions based on other exams can also be added upon the availability of data.

A feature that links the student to the university website upon clicking on the recommendations can also be added. Moreover, application deadlines of the various universities can also be added depending on the availability of data.

**CHAPTER-2**

**LITERATURE SURVEY**

**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 student 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 students etc.

Bayesian Networks were used by (Thi et al. (2007)) to create a decision support system for evaluating the application submitted by international students in the university. This model was designed to predict the performance of the aspiring students by comparing 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.

(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 know version of the C4.5 algorithms, 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.

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 decisions. Logistic regression and SVM were used to create the model, both models performed equally well and the final system was developed using Logistic regression 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 score, GPA score, residency race etc. The Model was created using the Multiple Logistic regression algorithm, it was able to achieve accuracy rate of 67% only.

(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 the have enquired about of 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

**As referred from the research report by:**

Himanshu Sonawane, National college of Ireland

**CHAPTER-3**

**SYSTEM DESIGN**

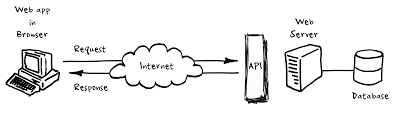
**INSIGHT ON THE ML ALGORITHMS**

**SYSTEM DESIGN:**

* **FLOW OF WEBSITE**

Universities For MS

Select University Rating



Predicting Chance of Admit

Predicting College Name

Colleges after 12th

Start

**INSIGHT ON SOME ML ALGORITHMS:**

Broadly there are three types of ML Algorithms:

1) SUPERVISED LEARNING:

**How it works:** This algorithm consists of a target / outcome variable (or dependent variable) which is to be predicted from a given set of predictors (independent variables). Using these set of variables, we generate a function that map inputs to desired outputs. The training process continues until the model achieves a desired level of accuracy on the training data. Examples of Supervised Learning: Regression, [Decision Tree](https://www.analyticsvidhya.com/blog/2015/01/decision-tree-simplified/), [Random Forest](https://www.analyticsvidhya.com/blog/2014/06/introduction-random-forest-simplified/), KNN, Logistic Regression etc.

2) UNSUPERVISED LEARNING:

**How it works:**In this algorithm, we do not have any target or outcome variable to predict / estimate. It is used for clustering population in different groups, which is widely used for segmenting customers in different groups for specific intervention. Examples of Unsupervised Learning: Apriori algorithm, K-means.

3) REINFORCEMENT LEARNING:

**How it works:**Using this algorithm, the machine is trained to make specific decisions. It works this way: the machine is exposed to an environment where it trains itself continually using trial and error. This machine learns from past experience and tries to capture the best possible knowledge to make accurate business decisions. Example of Reinforcement Learning: Markov Decision Process

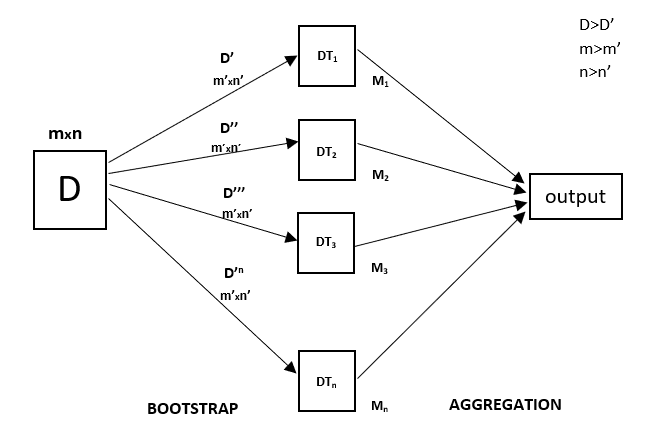
SOME COMMON ALGORITHMS:

1. Linear Regression
2. Multiple Regression
3. Decision Trees
4. Random Forest
5. kNN
6. SVM
7. Naïve Bayes

**Insight of the ML Algorithm used in the project:**

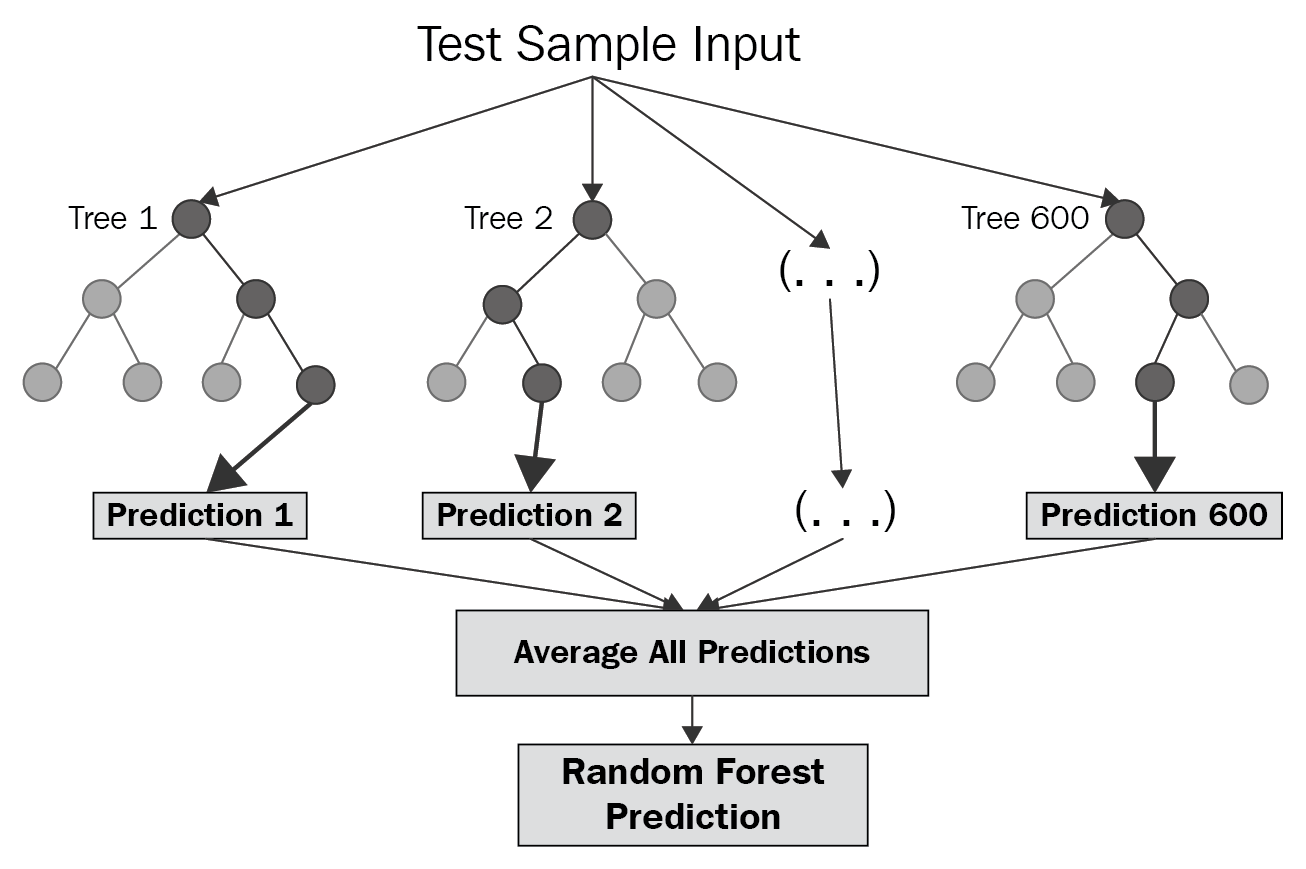
**Random Forest Regression**

Every decision tree has high variance, but when we combine all of them together in parallel then the resultant variance is low as each decision tree gets perfectly trained on that particular sample data and hence the output doesn’t depend on one decision tree but multiple decision trees.

In the case of a classification problem, the final output is taken by using the majority voting classifier. In the case of a regression problem, the final output is the mean of all the outputs. This part is Aggregation

A Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called Bootstrap and Aggregation, commonly known as **bagging**. The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees.

Random Forest has multiple decision trees as base learning models. We randomly perform row sampling and feature sampling from the dataset forming sample datasets for every model. This part is called Bootstrap.



**Random Forest over decision Trees:**

• Decision trees are sensitive to the specific data on which they are trained. If the training data is changed the resulting decision tree can be quite different and in turn the predictions can be quite different.

• Also Decision trees are computationally expensive to train, carry a big risk of overfitting, and tend to find local optima because they can’t go back after they have made a split.

• To address these weaknesses, we turn to Random Forest which illustrates the power of combining many decision trees into one model.

**CHAPTER-4**

**IMPLEMENTATION**

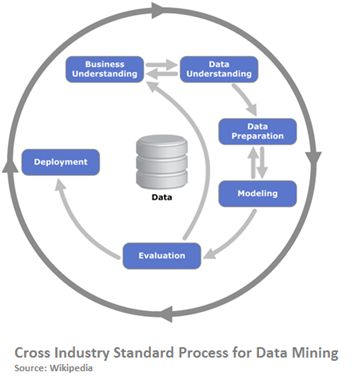
**Methodology:**

**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 analysed 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

**FIG: CRISP-DM**

**LIBRARIES USED:**

Python provides various libraries to support Machine Learning Algorithms and these libraries contain various modules to perform actions and visualise the ML models.

The libraries used in the project and their purposes are listed below:

Numpy: Used for working with arrays. It has functions for working in domain of linear algebra, Fourier transform, and matrices. It stands for Numerical Python.

Pandas: It is an open source data analysis and manipulation tool built on top of Python programming language.

Matplotlib.pyplot: used for creating static, animated and interactive visualisation.

sklearn: It is a built in machine learning library with various features such as classification, regression and clustering algorithms.

**DATASET PARAMETERS:**

The Dataset used is a CSV file that contains several parameters which are considered important during the application of Masters Programs.

The parameters included in first dataset are:

1. GRE Scores (out of 340)
2. TOEFL Scores (out of 120)
3. University Rating (out of 5)
4. Statement of Purpose and Letter of Recommendation Strength (out of 5)
5. Undergraduate GPA (out of 10)
6. Research Experience (either 0 or 1)
7. Chance of Admit (ranging from 0 to 1)

The parameters included in second dataset are:

1. JEE Rank
2. 10th Percentage
3. 12th Percentage

**Extraction of the dataset:**

The datasets were extracted to obtain the independent and dependent variables in both the cases. The independent variables(x-axis) are the ones that determines our target variable which is the dependent variable(y-axis).

**Training and Testing of the models:**

The dataset is first split into train and test data respectively in the ratio 3:1. The train data is used to train our machine learning model while the test data helps to determine the accuracy of the model.

The model is then using Random Forest Algorithm and the accuracy of the algorithm is then determined by evaluating the results.

**Evaluation of Results:**

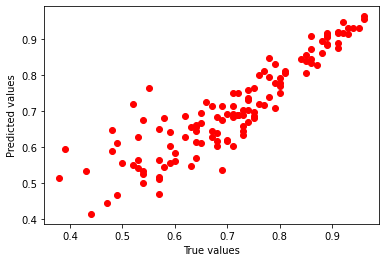
Dataset for Foreign Universities:

Mean Square Error: 0.0037444844799999987

Root Mean Square Error: 0.06119219296609657

**R2 coefficient:** 0.8013611283129357

Plot of Test values and predicted values using RFR algorithm:



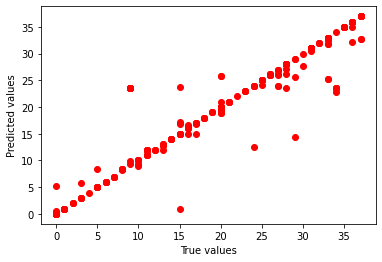
Dataset for Indian Universities:

Mean Square Error: 12.107062342041992

Root Mean Square Error: 3.4795204183970516

**R2 coefficient:** 0.9068400188124648

Plot of Test values and predicted values using RFR algorithm:

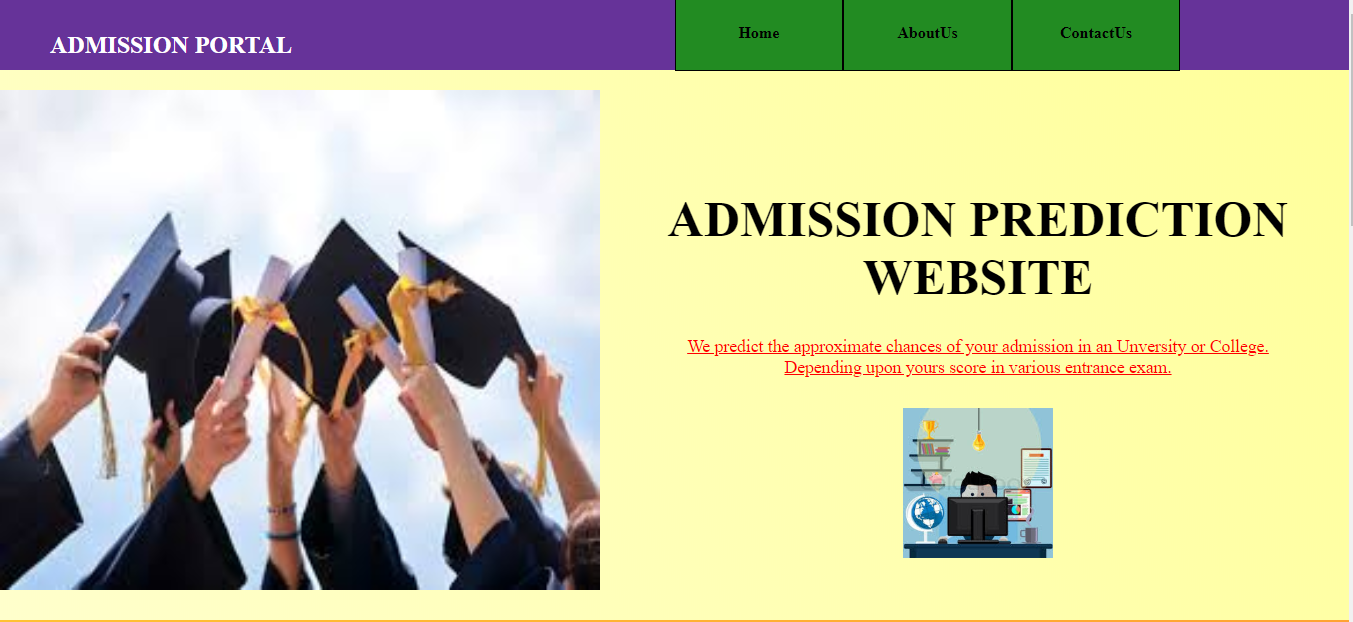


**CHAPTER-5**

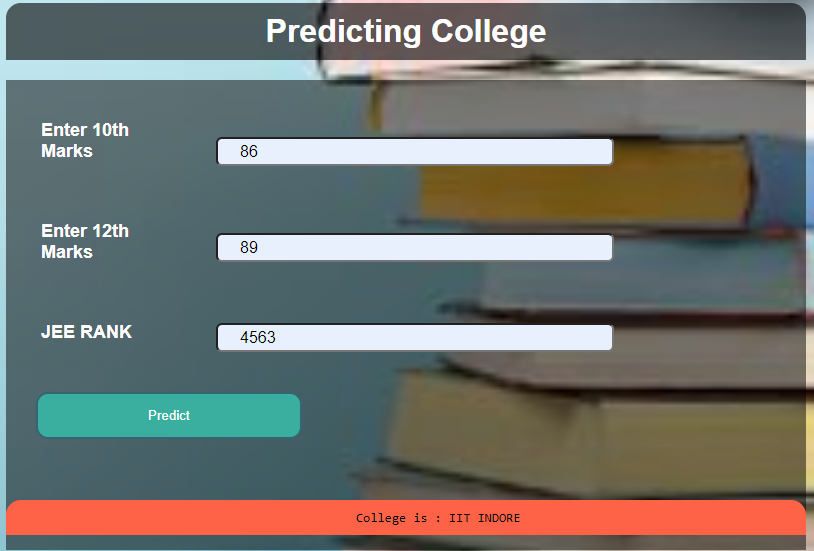
**IMPLEMENTATION DETAILS**

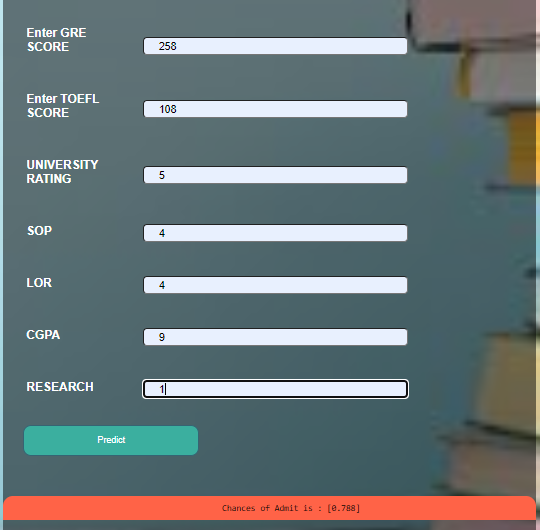
**IMPLEMENTATION DETAILS**

USER INTERFACE



The webpage users experience on navigating to the website. It contains information about the developers and their contact information along with a Home button.

The image above is the interface for predicting the Universities for the students pursuing their higher studies in India. It reads certain parameters from the user to predict the University they can apply.



The image above is the interface for predicting the Universities for the students pursuing their higher studies in foreign Universities. It reads certain parameters from the user to predict the chances of admit in the university they are trying to apply.

**CHAPTER-6**

**CONCLUSION/FUTURE SCOPE**

**Conclusion and Future Work**

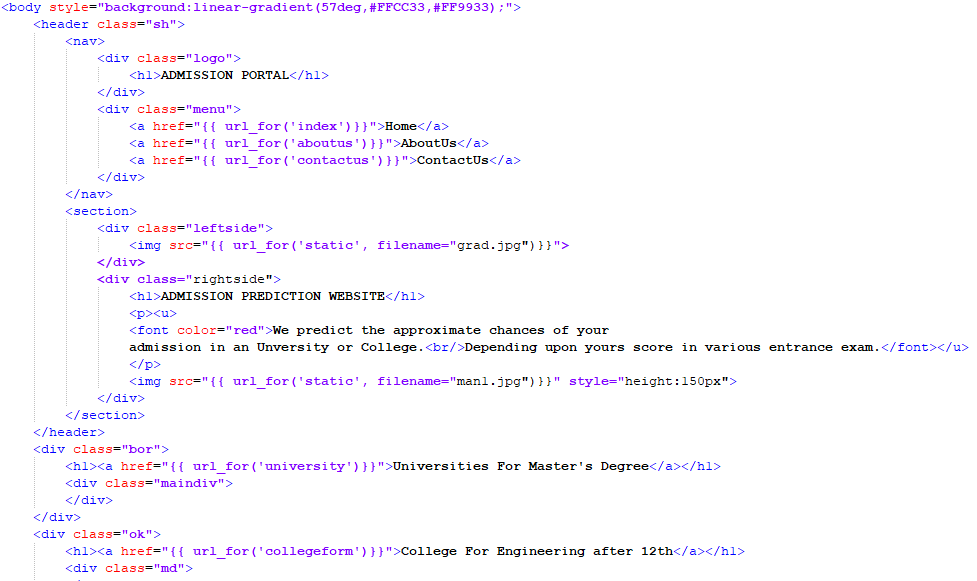
The main objective of this research was to develop a prototype of the system that can be used by the students aspiring to pursue their education in the foreign universities. Random Forest machine learning algorithm was developed and used for this research. The model can be used by the students for evaluating their chances of getting shortlisted in a particular university with an average accuracy of 80-90%. A simple user interface was developed to make the application interactive and easy to use for the users from the non-technical background. 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.

As discussed earlier in the limitation of the research we have created the models based only on the data of JEE exam and we have considered only limited universities with different rankings. In future, more data related to additional universities and courses can be added to the system. Also, the system can be enhanced to a mobile application by making changes to the code. Other classification algorithms can be evaluated to resolve the problem if they perform better than the current algorithm the system can be easily updated to support the new algorithm.

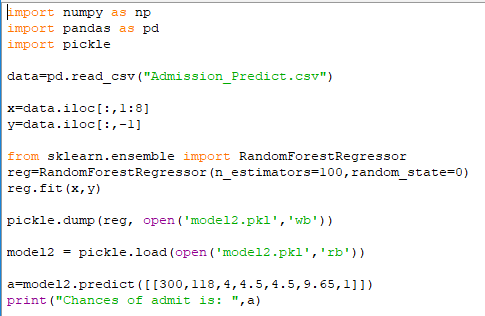
**CHAPTER-7**

**APPENDIX: CODE SAMPLES**

APPENDIX: CODE SAMPLES



HTML CODE SAMPLE



ML MODEL CODE SAMPLE



BACK-END CODE SAMPLE

**CHAPTER-8**

**ACKNOWLEDGEMENT**

**ACKNOWLEDGEMENT**

We would like to express our heartfelt gratitude to our college Fr. Conceicao Rodrigues Institute of Technology for giving us an opportunity to undertake this project. We are grateful to our H.O.D. Dhanashree Hadsul for giving us a chance to work on this project in our course of Engineering. We are thankful for and fortunate enough to get constant encouragement, support, and guidance from all our teachers of the IT Department which helped us in successfully completing our project.

We would also like to extend our gratitude to our parents for their support throughout, and the constant efforts and motivation by our teammates and friends which helped us attain successful completion of this project.

**CHAPTER-9**

**BIBLIOGRAPHY**

**REFERENCES**

1. https:// www.stackoverflow.com
2. https://www.youtube.com/programminghub
3. http://www.tutorialspoint.com
4. https://www.kaggle.com/
5. https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/
6. https://geekforgeeks.com/
7. Abdul Fatah S; M, A. H. (2012). Hybrid Recommender System for Predicting College Admission, pp. 107–113.
8. Bibodi, J., Vadodaria, A., Rawat, A. and Patel, J. (n.d.). Admission Prediction System Using Machine Learning.
9. Eberle, W., Simpson, E., Talbert, D., Roberts, L. and Pope, A. (n.d.). Using Machine Learning and Predictive Modelling to Assess Admission Policies and Standards.
10. Jamison, J. (2017). Applying Machine Learning to Predict Davidson College’s Admissions Yield, pp. 765–766. Mane, R. V. (2016). Predicting Student Admission decisions by Association Rule Mining with Pattern Growth Approach, pp. 202–207.

**CHAPTER-10**

**GETTING THINGS TO WORK**

**STEPS TO RUN THE CODE AND ACCESSING THE WEBSITE:**

1. Navigate to the folder where all the files are store from the command prompt and run the API using the command python app1.py
2. Copy the URL from command prompt and paste it into the address bar on the web browser to access the website.
3. The GUI will be displayed to the user.
4. For MS Universities click on “Universities for MS degree” link and select any University for which you want to check the chance of admit.
5. For pursing colleges in India for under graduation click on “College after 12th” link and enter the values as requested to predict the college.

**To access the ML code, follow the steps below:**

1. Run command prompt.
2. Type Jupyter Notebook in command prompt and press Enter.
3. The web browser will open and you will be able to see the different directories on your system.
4. Navigate to the folder containing the ML file and double click on it.
5. The code will open in a new tab.
6. Click on run all cells to run the code and check the output.