VELALAR COLLEGE OF ENGINEERING AND TECHNOLOGY

(AUTONOMOUS)

IBM PROJECT REPORT

TEAM ID   : PNT2022TMID22911

PROJECT NAME : UNIVERSITY ADMIT ELIGIBILITY PREDICTOR

TEAM MEMBERS:
> Ahilesh J
> Akilan Y
> GokulMani S
> Mouli E

ABSTRACT

In todays 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 theuniversities 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 admissionprocedures. 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 system.

The purpose of this project is to make the prediction of eligibility of an admission to a rated university with ease using a UI with the provided user details (GRE Score, TOEFL Score, University Rating, SOP, LOR, CGPA, Research). This also eliminates the possibility of human errors.

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 studying in more than 1700 public and 2500 private universities and colleges across the USA.

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.

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 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 dont 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 to develop the system. Finally, K Nearest Neighbours 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 predicted result for the application based on the profile as output.

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.

OBJECTIVE:

College admission predictor is a boon to many students. This helps the student not only to help in filling out the application forms but also give the students an idea about their future college by calculating their cut off.

When students come from rural places , they find it hard to go along with the formal procedures. So, this website helps them a lot and eases out their fear.

• Whatever may be their scores , this application helps to find the best colleges . Hence, our proposed computer aided system will help the students to get the list of all colleges in which they could get the admission at the click of a button.

• The students only have to enter their marks of XII, AIEEE etc. With this website, the students can very easily obtain the list of colleges even branch wise and course wise. This will not only make the admission process easy but also minimizes stress for students . The main objective of our system is to make the right choice of colleges.

LITERATURE REVIEW

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. (Bibodiet 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 (Thietal. (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.

(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 tech-niques 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 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.

(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 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.

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 theapplication.(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.

EXISTING SYSTEM

Today in college's student details are entered manually. The student details in separate records are tedious task. Referring to all these records updating is needed. There is a chance for more manual errors.

1. When the student comes in college.
2. First of all,he/she takes admission form from reception.
3. Fills it and submits it into office.
4. Filled form is first checked with documents like merit list an details came from university and verified by an official person, if there is any mistake then it is corrected.
5. At the time of submission of it the fees is deposited by the candidate.
6. At the time of submission of admission form admission number is assigned to the candidate by the institute.
7. Candidate gets the receipt of fees deposition.

DISADVANTAGES OF EXISTING SYSTEM

1. Require much man power i.e. much efforts, much cost and hard to operate and maintain.
2. Since, all the work is done in papers so it is very hard to locate a particular student record when it is required.

PROPOSED SYSTEM

These problems can be resolved by using regression algorithms / classification algorithms as they can consider most of the features for prediction. Linear regression / KNN classification / Random forest Regressor can be used as the machine learning model for the model. XG boost model can also be used which performs better on small to medium scale datasets but the model giving accurate and desired results only will be selected. The aim of the proposed system is to address the limitations of the current system. The requirements for the system have been gathered from the defects recorded in the past and also based on the feedback from users of previous metrics tools.

ADVANTAGES OF PROPOSED SYSTEM:

• Reach to geographically scattered student.
• Reducing time in activities
• Paperless admission with reduced man power
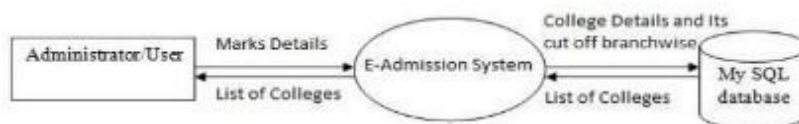• Operational efficiency.

Fig:1 Dataflow Diagram

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 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.

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.
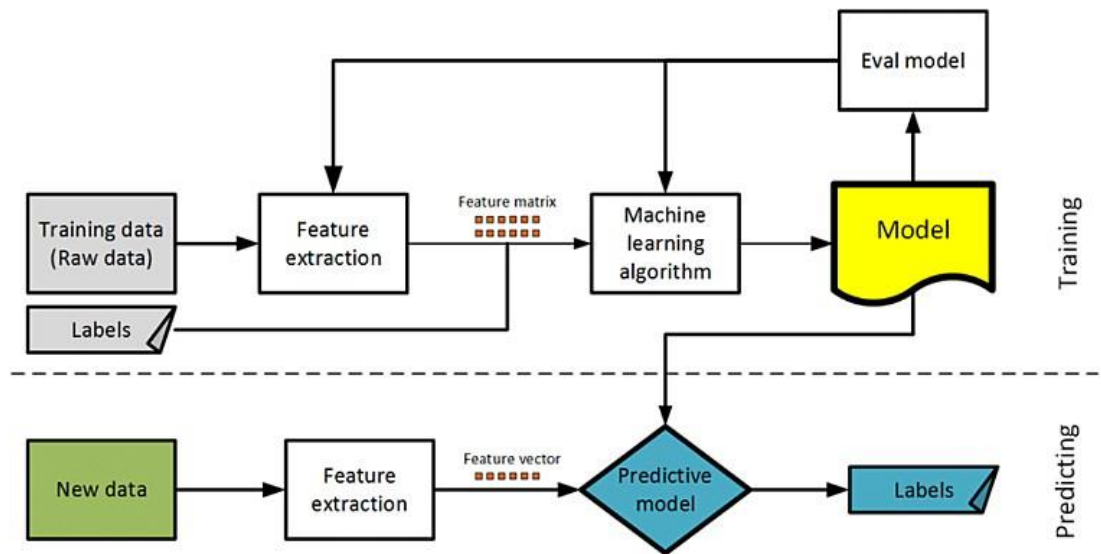
Evaluation:

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

Deployment:

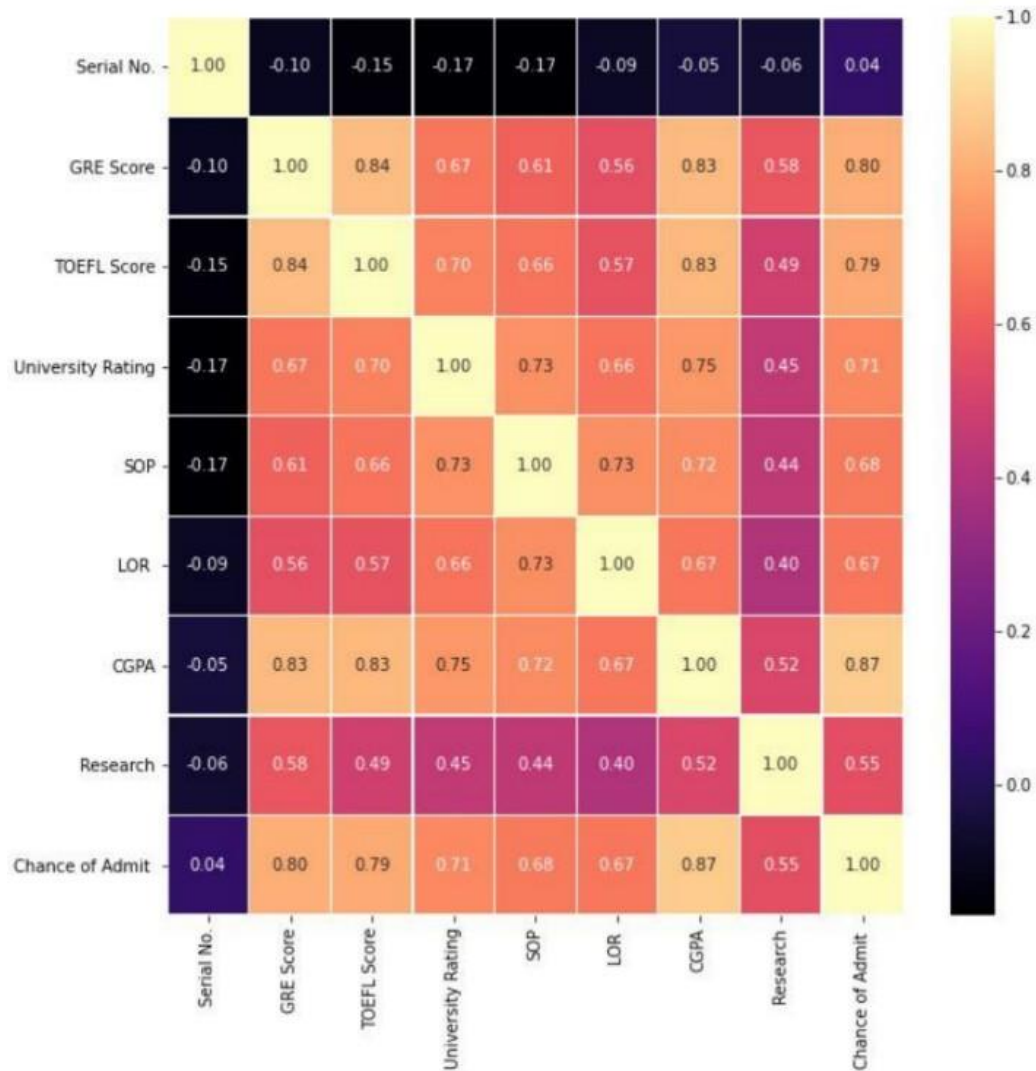Once the models were evaluated they were integrated with code developed for user interface.
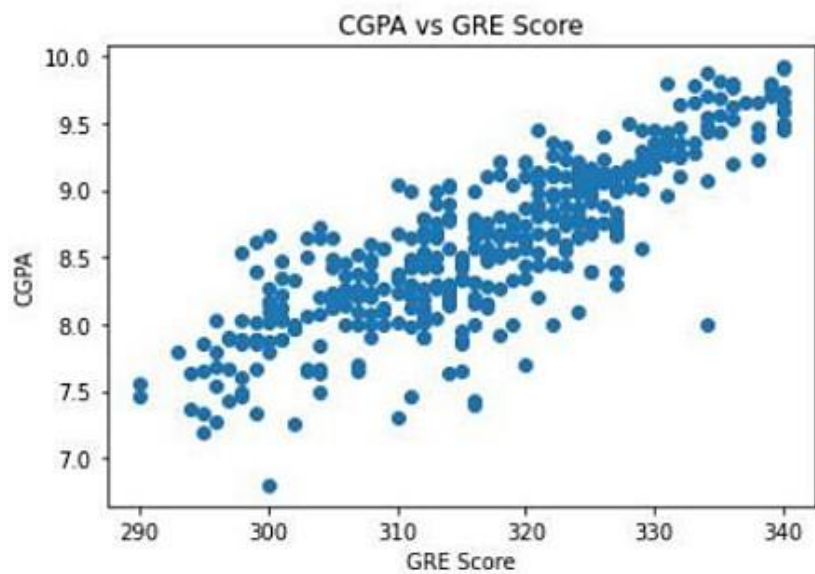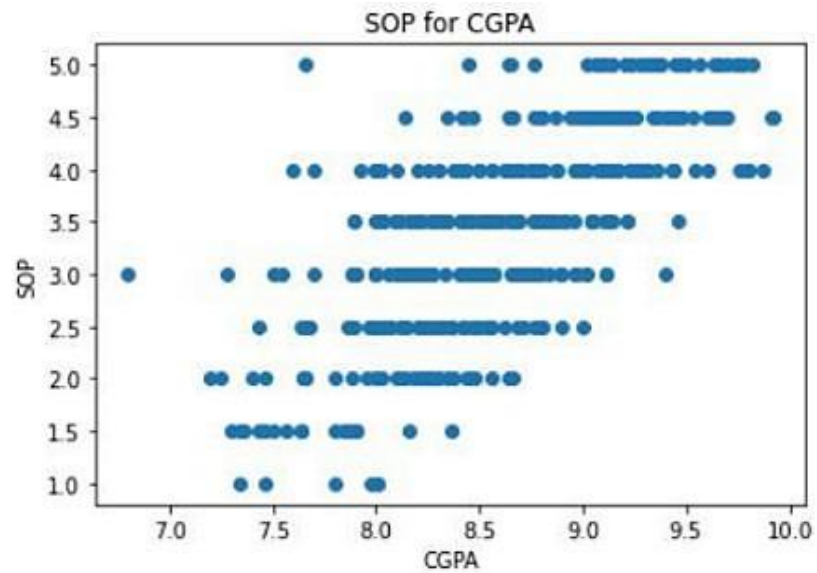
THEORETICAL ANALYSIS

BLOCK DIAGRAM:

EXPERIMENTAL ANALYSIS

HEAT MAP OF DATA:

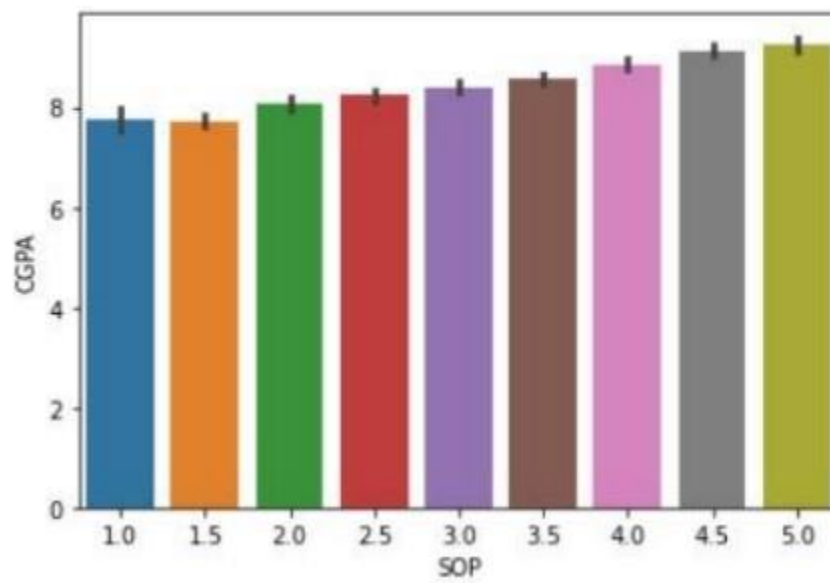|	Serial No.	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
Serial No.	1.00	-0.10	-0.15	-0.17	-0.17	-0.09	-0.05	-0.06	0.04
GRE Score	-0.10	1.00	0.84	0.67	0.61	0.56	0.83	0.58	0.80
TOEFL Score	-0.15	0.84	1.00	0.70	0.66	0.57	0.83	0.49	0.79
University Rating	-0.17	0.67	0.70	1.00	0.73	0.66	0.75	0.45	0.71
SOP	-0.17	0.61	0.66	0.73	1.00	0.73	0.72	0.44	0.68
LOR	-0.09	0.56	0.57	0.66	0.73	1.00	0.67	0.40	0.67
CGPA	-0.05	0.83	0.83	0.75	0.72	0.67	1.00	0.52	0.87
Research	-0.06	0.58	0.49	0.45	0.44	0.40	0.52	1.00	0.55
Chance of Admit	0.04	0.80	0.79	0.71	0.68	0.67	0.87	0.55	1.00

SCATTER PLOTS AND BAR GRAPHS:

SOP for CGPA

CGPA vs GRE Score

+

<AxesSubplot:xlabel='SOP', ylabel='CGPA'>

<AxesSubplot:xlabel='University Rating', ylabel='count'>

<AxesSubplot:xlabel='University Rating', ylabel='Chance of Admit '>
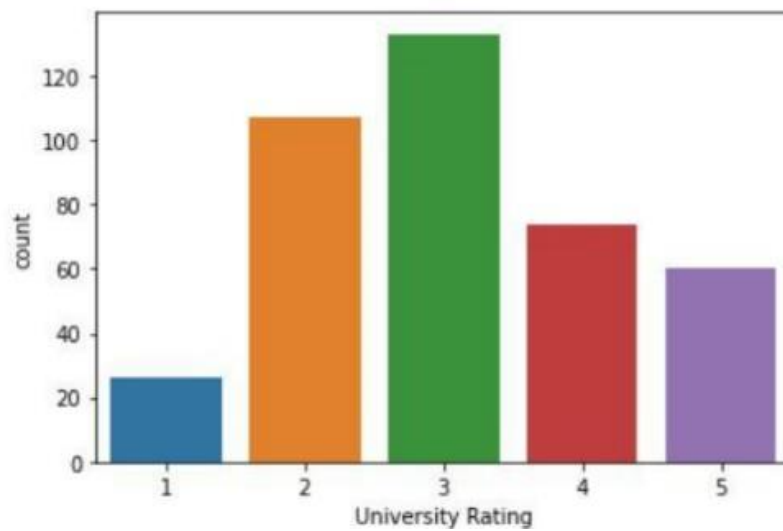

<AxesSubplot:xlabel='University Rating', ylabel='count'>
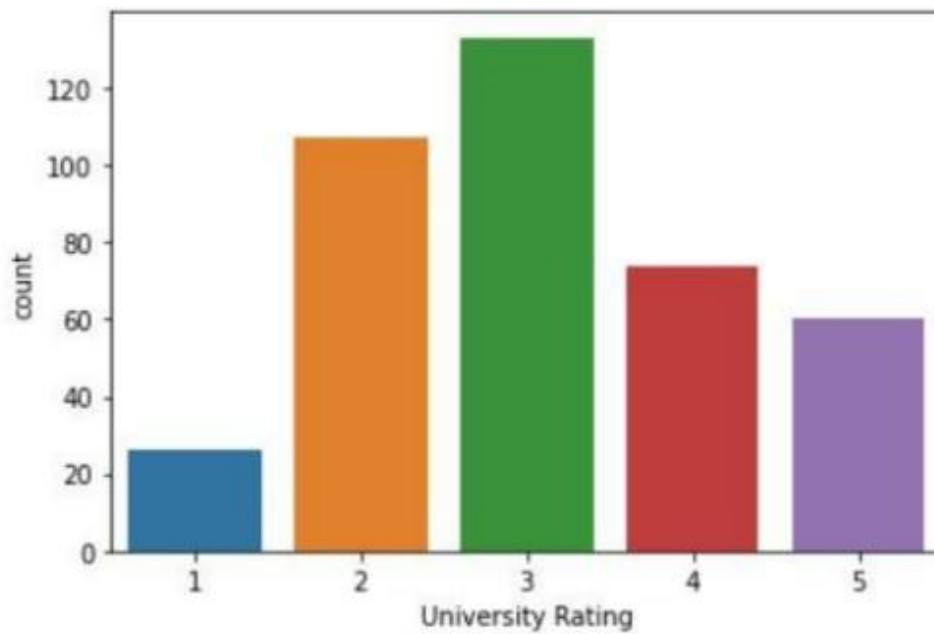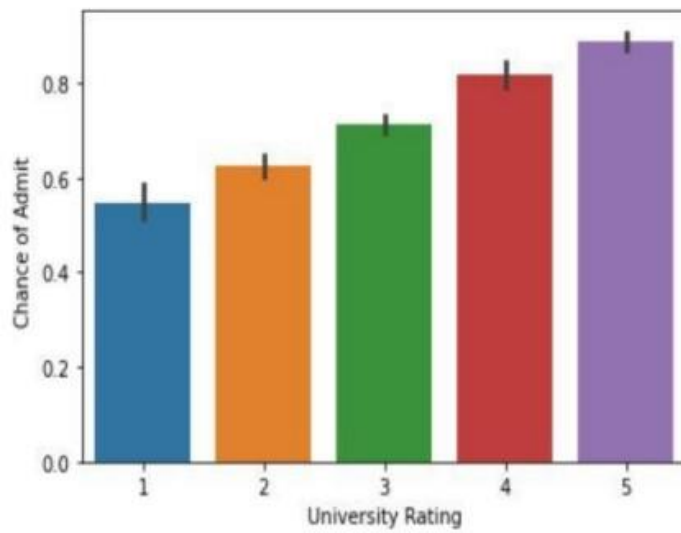
ARCHITECTURE

• The student will enter his/her profile details using the user interface developed in shiny.

• The user interface code will interact with the KNN and Decision Tree models to provide the users with the required result.

• The KNN algorithm will be used to determine the chance of the student of securing admission in a particular university based on his/her profile.

• The Decision Tree algorithm 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 models have been executed the result will be provided to the student as the output on the user interface.
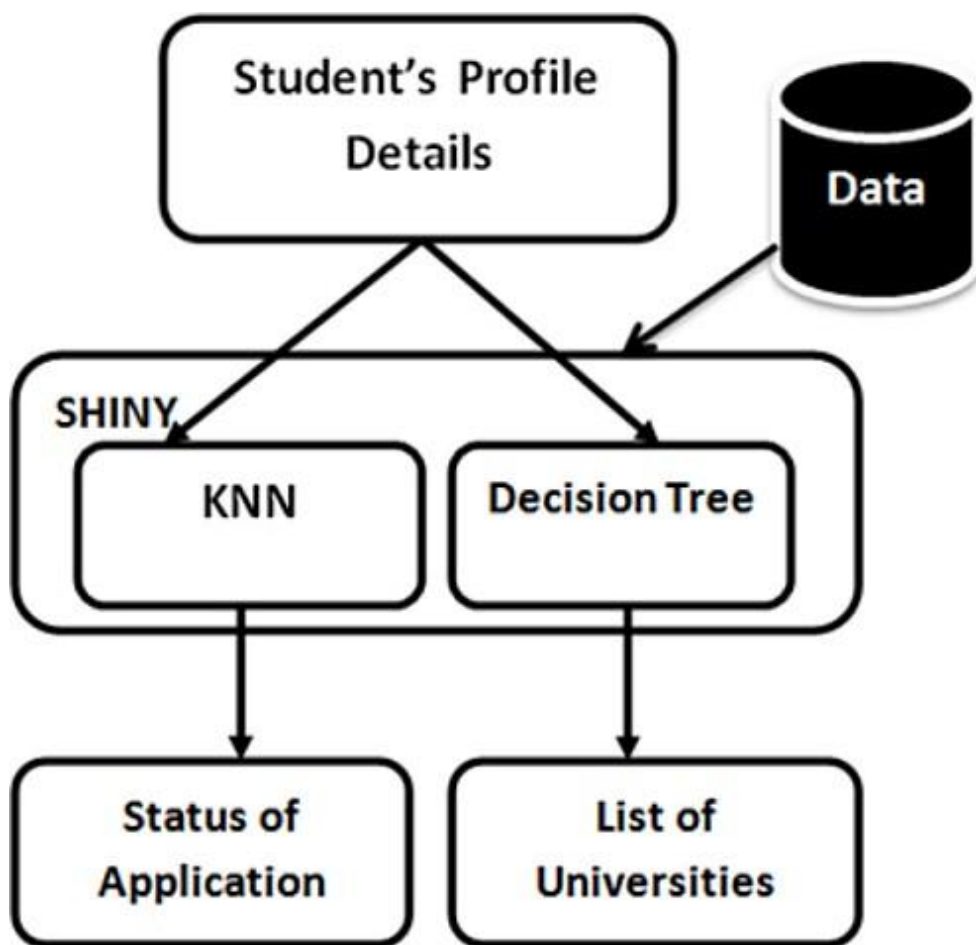
TESTING

UNIT TESTING:

Unit testing is carried out screen-wise, each screen being identified as an object. Attention is diverted to individual modules, independently to one another to locate errors .Thishas enabled the detection of errorsin coding and logic .This is the first level of testing.In this, codes are written such thatfromone module ,we can move on to the next module according to the choice we enter.

SYSTEM TESTING:

In this, the entire system was tested as a whole with all forms, code, modules and class modules .System testing is the stage of implementation, which is aimed at ensuring that the system works accurately and efficiently before live operation commences. It is a series of different tests that verifies that all system elements have been properly integrated and perform allocated functions. System testing makes logical assumptions that if all parts of the system are correct, the goal will be successfully achieved. Testing is the process of executing the program with the intent of finding errors. Testing cannot show the absence of defects, it can only show that software errors are present.

INTEGRATION TESTING:

This testing strategies combines all the modules involved in the system. After the independent modules are tested, dependent modules that use the independent modules are tested. This sequence of testing layers of dependent modules continues until the entire system is constructed .Though each module individually, they should work after linking them together. Data may be lost across interface and one module can have adverse effect on another. Subroutines, after linking, may not do the desired function expected by the main routine. Integration testing is a systematic technique for constructing program structure while at the same time ,conducting test to uncover errors associated with the interface. In the testing the programs are constructed and tested in the small segments.

WHITE BOX TESTING:

White-box testing is concerned with testing the implementation of the program .The intent of this testing is not to exercise all the different input or output conditions but to exercise the different programming structures and data structures used in the program. White box testing is also called structural testing .To test the structure of a program, structural testing aims to achieve test cases that will force the desired coverage of different structures. Various criteria have been proposed for this there are three different approaches to structural testing: control flow-based testing data flow-based testing, and mutation testing.

BLACK BOX TESTING:

In black-box testing the structure of the program is not considered. Test cases are decided solely on the basis of the requirements or specifications of the program or module ,and the internals of the module or the program are not considered for selection of test cases .In black-box testing, the tester only knows the inputs that can be given to the system and what output the system should give.This form of testing is also called functional or behavioral testing .The most obvious functional testing procedure is exhaustive testing. One criterion for generating test cases is to generate them randomly. There are no formal rules for designing test cases for functional testing. In fact, there are no precise criteria for selecting test cases.

FUTURE SCOPE

The future scope of this project is very broad.

Few of them are:

    This can be implemented in less time for proper admission process.

    This can be accessed anytime anywhere, since it is a website provided only an internet connection.

    The user had not need to travel a long distance for the admission and his/her time is also saved as a result of this automated system.

CONCLUSION

The project uses a Random forest regressor to predict the output and a website is built to make the UI more accessible and easy using various technologies such as python, HTML5, CSS, Flask, Scikit, Matplot, Numpy, Pandas, Seaborn and other libraries. After the deployment of the web application, it can be accessed from anywhere with internet connection. This project reduces the long hours of analysis to predict the eligibility of the admission to a rated university.

SOURCE CODE:

```html
<html>
  <head>
    <SCRIPT language=Javascript>
      <!--
      function check(e, value) {
        //Check Charater
        var unicode = e.charCode ? e.charCode : e.keyCode;
        if (value.indexOf(".") != -1)
          if (unicode == 46) return false;
        if (unicode != 8)
          if ((unicode < 48 || unicode > 57) && unicode != 46) return false;
      }
      //-->
    </SCRIPT>
  <meta name="viewport" content="width=device-width, initial-scale=1">
  <link rel="stylesheet" href="https://cdnjs.cloudflare.com/ajax/libs/font-awesome
  <style>
  body {
    font-family: Arial, Helvetica, sans-serif;
  }

  * {
    box-sizing: border-box;
  }

  /* style the container */
  .container {
    position: relative;
    border-radius: 5px;
    background-color: palevioletred;
    padding: 20px 0 30px 0;
  }

  /* style inputs and link buttons */
  input,
  .btn {
    width: 100%;
    padding: 12px;
    border: none;
```

```css
  border-radius: 4px;
  margin: 5px 0;
  opacity: 0.85;
  display: inline-block;
  font-size: 17px;
  line-height: 20px;
  text-decoration: none; /* remove underline from anchors */
}

input:hover,
.btn:hover {
  opacity: 1;
}



/* style the submit button */
input[type=submit] {
  background-color: #000000;
  color: white;
  cursor: pointer;
}

input[type=submit]:hover {
  background-color: #151615;
}

/* Two-column layout */
.col {
  float: left;
  width: 50%;
  margin: auto;
  padding: 0 50px;
  margin-top: 6px;
}

/* Clear floats after the columns */
.row:after {
  content: "";
  display: table;
  clear: both;
}
```

```css
/* vertical line */
.vl {
  position: absolute;
  left: 50%;
  transform: translate(-50%);
  border: 2px solid rgb(8, 8, 8);
  height: 490px;
}



/* hide some text on medium and large screens */
.hide-md-lg {
  display: none;
}


/* bottom container */
.bottom-container {
  text-align: center;
  border-radius: 0px 0px 4px 4px;
}

/* Responsive layout - when the screen is less than 650px wide, make the two col
@media screen and (max-width: 650px) {
  .col {
    width: 100%;
    margin-top: 0;
  }
  /* hide the vertical line */
  .vl {
    display: none;
  }
  /* show the hidden text on small screens */
  .hide-md-lg {
    display: block;
    text-align: center;
  }
}
</style>
</head>
<body>

<div class="container">
```

```html
<form action="{{url_for('predict')}}" method="post">
  <div class="row">
    <h2 class="bottom-container">University Admit Eligibility Predictor</h2>
    <div class="vl">
      <span class="vl-innertext"></span>
    </div>

    <div class="col">
      <p><b>A Simple Web App to Predict The Chances of Getting Admission in U
      <h3>Input Guide</h3>
    <ul>
                    <li>GRE Score (out of 340)</li><br>
                    <li>TOEFL Score (out of 120)</li><br>
                    <li>University Rating (out of 5) - the category of the targe
                    <li>Statment of Purpose {SOP} Strength (out of 5)</li><br>
                    <li>Letter of Recommendation {LOP} Strength (out of 5)</li><
                    <li>Undergraduate CGPA (out of 10)</li><br>
                    <li>Research Experience (0 for NONE and 1 for YES)</li>

      </ul>
    </div>

    <div class="col">
      <div class="hide-md-lg">
      </div>

                    <input type="number" name="GRE Score" placeholder="GRE Score
                    <input type="number" name="TOEFL Score" placeholder="TOEFL S
                    <input type="number" name="University Rating" placeholder="U
                    <input type="number" name="SOP" placeholder="SOP" required="
max="5"/>
                    <input type="number" name="LOR" placeholder="LOR" required="
max="5"/>
                    <input type="number" name="CGPA" placeholder="CGPA" required
max="10"/>
                    <input type="number" name="Research" placeholder="Research"

                    <input type="submit" value="Predict"></input>

                    <h4 style="text-align: center;">{{prediction_text}}</h4>
      </div>

  </div>
```

```
            </form>


        </div>



        </body>
        </html>
```

REFERENCE:

MasterPortal (2017). MasterPortal.

URL: http://www.mastersportal.eu/countries/82/united-states.html

Mishra, S. and Sahoo, S. (2016). A Quality Based AutomatedAdmission System for Educational Domain, pp. 221–223.

Mozenda (n.d.). Mozenda.

URL: https://www.mozenda.com/

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.

URL:

http://ezproxy.bethel.edu/login?url=http://inderscience.metapress.com/content/fr20 314q1862653k/%5Cnhttp://nandeshwar.info/wp-content/uploads/2008/11/DMWVU Project.pdf

Prescholar (2017). Yocket.

URL: https://www.prepscholar.com/toefl/blog/toefl-to-ielts-conversion

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.

URL: https://economictimes.indiatimes.com/wealth/personal-finance-news/around-3-lakh-indian-students-go-abroad-annually-33-take-property-backed-loans/articleshow/48703656.cms?intenttarget=no

Waters, A. and Miikkulainen, R. (2013). G RADE : Machine Learning Support for Graduate Admissions. Yocket (2017). Yocket.URL: https://yocket.in/