# ADMISSION PREDICTOR

***Mini Project report submitted in***

***Partial fulfillment of requirement for the award of the degree of***

## Bachelor of Technology

***in***

## Artificial Intelligence

***by***

## Anuj Wadi Aryan Meshram Harshal Bodhe Ishita Sahare Krutika Tathe

***Under the guidance of***

### Dr. Sharda A. Chhabria

Associate Professor, AI Dept



### Department of Artificial Intelligence

G H Raisoni Institute of Engineering and Technology, Nagpur

(An Autonomous Institute Affiliated to Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur)

Accredited by NAAC with A+ Grade

## 2021-22

**G H Raisoni Institute of Engineering and Technology, Nagpur**

**(An Autonomous Institute)**

**Department of Artificial Intelligence**



# Certificate

The report of project titled **ADMISSION PREDICTOR** submitted by **Anuj Wadi, Aryan Meshram, Harshal Bodhe, Ishita Sahare** and **Krutika Tathe** in the partial fulfillment of the degree of Bachelor of Technology in **Artificial Intelligence** during academic year 2021-22, has been carried out under my/our supervision at the Department of Artificial Intelligence of GH Raisoni Institute of Engineering and Technology, Nagpur. The work is comprehensive, complete and fit for evaluation.

Dr. Sharda Chhabria

##### Guide (Mini Project) Associate Professor, AI Dept

**Dr. Smita Nirkhi Dr. Vivek Kapur**

**HOD, AI Dept. Director, GHRIETN**

**G H Raisoni Institute of Engineering and Technology, Nagpur**

**(An Autonomous Institute)**

**Department of Artificial Intelligence**

## Declaration

I certify that

1. The work contained in this project has been done by me under the guidance of my supervisor(s).
2. The work has not been submitted to any other Institute for any degree or diploma
3. We have followed the guidelines provided by the Institute in preparing the project report
4. We have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.
5. Whenever we have used materials (data, theoretical analysis, figures, and text) from other sources, we have given due credit to them by citing them in the text of the report and giving their details in the references Further, We have taken permission from the copyright owners of the sources, whenever necessary.

**Signature of the Projectees**

1. **Anuj Wadi**
2. **Aryan Meshram**
3. **Harshal Bodhe**
4. **Ishita Sahare**
5. **Krutika Tathe**

**G H Raisoni Institute of Engineering and Technology, Nagpur**

**(An Autonomous Institute)**

**Department of Artificial Intelligence**

## Acknowledgement

## 

## We are extremely thankful to our guide Dr. Sharda Chhabria under whom our project took the shape of reality from mere ideas. We are grateful to our guide for enlightening us with her precious guidance and constant encouragement. We thank our guide for providing us with ample support and valuable time. We are indebted to here to her for providing a stimulus to reach our goals.

## We are grateful to Dr. Smita Nirkhi, HOD Artificial Intelligence Department, GHRIETN, for her kind cooperation and timely help.

## Lastly, we would also like to thank Dr. Vivek Kapur, Director, GHRIETN for his constant support and inquisitiveness.

## Also, we would like to thank all those who were directly or indirectly related to our project and extended their support to make the project successful.

**Name of Projectees**

1. **Anuj Wadi**
2. **Aryan Meshram**
3. **Harshal Bodhe**
4. **Ishita Sahare**
5. **Krutika Tathe**

**TABLE OF CONTENTS**

CERTIFICATE…………………………………………………………………………II

DECLARATION……………………………………………………………………….III

ACKNOWLEDGEMENT……………………………………………………………..IV

[LIST OF FIGURES VI](#_bookmark0)

LIST OF TABLES V

ABSTRACT 1

CHAPTER 1: INTRODUCTION 2

* 1. [Overview 2](#_TOC_250003)
  2. [Problem Statement 2](#_TOC_250002)
  3. Thesis Objectives 3
  4. Thesis Contributions 3

CHAPTER 2: REVIEW OF LITERATURE

[2.1 Literature Survey 3](#_TOC_250001)

CHAPTER 3: WORK DONE 4

* 1. Overview 4
  2. Methodology 4
  3. Tools… 5

CHAPTER 4: RESULTS AND DISCUSSIONS 9

* 1. Experimental Design 9
  2. Experimental Result 9

CHAPTER 5: SUMMARY AND CONCLUSIONS 10

[REFERENCES 12](#_TOC_250000)

## LIST OF FIGURES

Figure 1: Linear Regression 5

Figure 2: Neural networks 6

Figure 3: Decision Trees 6

Figure 4: Random Forests 7

**ABSTRACT**

College Admission Predictor System is a web based application system in which students can register their marks along with their personal information. This helps to predict their admissions in colleges. Administrator can add the college details and the batch details. Using this Application, the entrance seat allotment becomes easier and efficient.

The main advantage of the project is the computerization of the entrance seat allotment process. Administrator has the power for the allotment. Admin can add the allotted seats into a file and the details are saved into the system. The total time for the entrance allotment becomes lower and the allotment process becomes faster. It helps students to make right decisions for choosing their college. In which students can register with their personal information as well as marks details to prediction the admission in colleges and the administrator can allot the seats for the students.

Administrator can add the college details and the batch details. Using this Application, the entrance seat allotment became easier and can be implemented using system. The main advantage of the project is the computerization of the entrance seat allotment process. Administrator has the power for the allotment. Admin can add the allotted seats into a file and the details are saved into the system. The total time for the entrance allotment became lesser and the allotment process became faster. It helps student for making decision for choosing a college.

## CHAPTER 1

### INTRODUCTION:

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. Its design concentrates on 3 types of users: 1. Admin 2. Students 3. Account 4. Student-section.

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. The mobile application would require connecting to the database on a remote server using Wi-Fi technology.

After XII, students desiring to take admissions in professional colleges like engineering face lots of problems. Admissions in engineering colleges in the state of Maharashtra or any state is based upon common entrance test (CET) and since more than 1.5lakh seats are to be allotted in more than 200 engineering colleges and over 35 different branches of engineering, for students belonging to many categories like open, home university, outside home university, reserved category (SC, ST, OBC etc) the problem becomes more serious and students struggle to understand which colleges they are likely to get admitted in, even after going through cut-off data of previous years.

Many students fill wrong Options and fail to get admission. To minimize the stress of students we came up with the idea of a computer aided method which will help the students get the list of all colleges in which they could get the admission at the click of a button, making the admission process fast and easy.

##### Overview:

* To help students to fill their preferences at the time of option-entry process accurately.
* To ease of making better choices of college before allotment.
* To deploy a web application for college admission system.

After intermediate, students desiring to pursue engineering face lot of problem in choosing a good college and branch of their choice. Admission into engineering colleges across states in India happens generally through Common Entrance Tests (CET). The examination authority of every state carries out the admission, through a centralized admission process. This admission process happens through many rounds, depending on availability of seats. First, the students must get their documents verified by the authority. Later, the authority releases the cut-offs of every college, branch-wise and category- wise. Students will be allowed to give their preference list of colleges and branches, which is also known as the option-entry process. Then, based on rank, category and preference list given by the students, college and branch will be allotted to them by the authority.

##### Problem Statement:

Educational organizations have always played an important and vital role in society for development and growth of any individual. There are different college prediction apps and websites being maintained contemporarily, but using thesis tedious to some extent, due to the lack of articulate information regarding colleges, and the time consumed in searching the best deserving college.

The problem statement, hence being tackled, is to design a college prediction/prediction system and to provide a probabilistic insight into college administration for overall rating, cut-offs of the colleges,

admission intake and preferences of students. Also, it helps students avoid spending time and money on counselor and stressful research related to finding a suitable college. It has always been a troublesome process for students in finding the perfect university and course for their further studies. At times they do know which stream they want to get into, but it is not easy for them to find colleges based on their academic marks and other performances. We aim to develop and provide a place which would give a probabilistic output as to how likely it is to get into a university given upon their details.

* + - 1. **Thesis Objectives:**

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 application 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 MHT-CET. With this application, 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.

The main goal of the system is to automate the process carried out in the organization with improved performance and realize the vision of paperless admission. Some of the goals of the system are listed below:

* Manage large number of student details.
* Manage all details of student who registered for the course
* Create student accounts and maintain the data is effectively.
* View all the details of the students.
* Reduce the work load in interview the students for selection.
* Activities like updating, modification, deletion of records should be easier.

### PROJECT CONTRIBUTION

#### 1.4.1 TECHNICAL FEASIBILITY

Building this project is technically feasible. The hardware and the software needed are all available, it’s not much difficult to get them. Briefly it can be said that the resources needed for the development and maintenance of the system are available.

#### OPERATIONALLY FEASIBILITY

This project development is operationally feasible as there is no need for a user to have good technical knowledge before using it. The user can learn and use the system with ease, he/she just needs to just read the manual or tutorial from the developers.

#### ECONOMIC FEASIBILITY

Besides being technically feasible, developing this system is economically feasible as well. The development of the system does not require developers to spend a lot of money. The tools used for development of these projects aren’t expensive and software is open source. The system is indeed economically feasible.

## CHAPTER 2

#### 2.1 Literature Survey

S. Lai, et. al. [1] has proposed context-based information classification; LSTM is very useful. The performance is best in several datasets particularly on document level datasets. Depending on the words used in the sentences, weights are assigned to it and are pooled into minimum, average and the max pools. Here, max pooling is applied to extract the keywords from the sentences which are most important. LSTM generates high accurate results in comparatively less time.

2. A. Hassa, et. al. [9] have proposed LSTM for the structure sentence representation. This tree-like structure captures the semantics of the sentences. The text is analyzed word by word by using LSTM then the semantic of all the previous texts are preserved in a fixed size hidden layer. For the proposed system LSTM plays an important role, being a memory storage it holds the characters which helps in predicting the next word.

3. J. Y. Lee, et. al. [7] have proposed that text classification is an important task in natural language processing. Many approaches have been developed for classification such as SVM (Support Vector Machine), Naïve Bayes and so on. Usually short text appears in sequence (sentences in the document) hence using information from preceding text may improve the classification. This paper introduced LSTM(Long-Short Term Memory) based models for text classification.

4. V. Tran, et. al. [5] have proposed that n-gram is a contiguous sequence of ‘n’ items from a given sequence of text. If the given sentence is ‘S’ ,we can construct a list of n-grams from ‘S’ , by finding pairs of words that occur next to each other. The model is used to derive probability of sentences using the chain rule of unconditional probability.

5. Z. Shi, et. al. [4] have defined that recurrent neural network has input, output and hidden layer. The repeating module of ordinary LSTM has a simple structure instead LSTM uses a more complex function to replace it for more accurate results. The key element in the LSTM is the cell state which is also called the hidden layer state.

6. K. C. Arnold, et. al. [6] have proposed an approach that presents phrase suggestions instead of word predictions. It says phrases were interpreted as suggestions that affect the context of what the user writes more than then the conventional single word suggestion. The proposed system uses statistical language modeling capable of accurately predicting phrases and sentences. It pruned the n grams that repeated less than two times in the dataset, by marking the start-of-sentence token with some additional flags to indicate the start of the sentence. The work demonstrated the phrase completions were accepted by users and were interpreted as suggestions rather than the predictions.

7. College Admission Predictor, Annam Mallikharjuna Roa 1, Nagineni Dharani 2, A. Satya Raghava 3, J. Buvanambigai 4, K. Sathish 5 1, 2, 3, 4 Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, Tamil Nadu, India Journal of Network Communications and Emerging Technologies (JNCET) Volume 8, Issue 4, April (2018) Preference of system design

## CHAPTER 3

* 1. **WORKDONE**

For anyone pursuing their postgraduate studies, it would be difficult for them to find out what college they may join, based on their GPA, Quants, Verbal, TOEFL and AWA Scores. People may apply to many universities that look for candidates with a higher score set, instead of applying to universities at which they have a chance of getting into.

This would be detrimental to their future. It is very important that a candidate should apply to colleges that he/she has a good chance of getting into, instead of applying to colleges that they may never get into.

There aren’t many efficient ways to find out the colleges that one can get into, relatively quickly. The Education Based Prediction System helps a person decide what colleges they can apply to with their scores.

The dataset that is used for processing consists of the following parameters: University name, Quants and Verbal Scores (GRE) TOEFL and AWA Scores. The GRE Test (Graduate Record Examinations) is a standardized test used by many universities and graduate schools around the world as part of the graduate admissions process.

Other factors are also taken into consideration while applying to colleges, such as Letter of Recommendation (a formal document that validates someone's work, skills or academic performance), Statement of Purpose (a critical piece of a graduate school application that tells admissions committees who you are, what your academic and professional interests are, and how you'll add value to the graduate program you're applying to), Co-curricular activities and Research papers as well (research papers from journals that are not well known or have a high percentage of plagiarism are not taken into consideration for this case).

When a person has completed their undergraduate degree and wants to pursue a Postgraduate degree in a field of their choice, more often than not, it is very confusing for the person to figure out what colleges they should apply to with the scores that they have obtained in GRE and TOEFL, along with their GPA at the time of their graduation. Many candidates may apply to colleges that do not fall under their score requirements and hence waste a lot of time.Applying to many colleges with scores also increases the cost.

There are not many efficient methods that are available to help address this issue and hence an Education Predictor System has been developed. In the system proposed, a person can enter their scores in the respective fields provided. The system then processes the data entered and produces an output of the list of colleges that a person could get into, with their scores. This is relatively quick and helps conserve time and money. In order to achieve this we have proposed a novel method utilizing Machine Learning algorithms.To maximize the accuracy of our model, we have taken into consideration not one; but several machine learning algorithms.

These algorithms include Neural Networks, Linear Regression, Decision Tree and Random Forest. More about these algorithms will be covered in the Algorithms section of this paper. These Algorithms are then compared and the algorithm which has the best key performance indicators will be used to develop the Prediction System. We also look forward to incorporate clustering of universities based on a profile and then classifying them as less likely, highly likely acceptance.

### ALGORITHM USED

##### Linear Regression

##### Regression models are used to describe a relation between different variables by using the observed data into a line. Straight lines are used in linear regression models, whereas curved line is used in logistic and non-linear regression models. Linear regression model is a method used as response for only a single feature, it is based on supervised learning. Regression models always target a prediction value which is based on independent variables. It is used to calculate the relationship between two quantitative variables.

##### Regression models differ upon – the type of relation between independent and dependent variables, the number of variables being used and the ones they are considering. It is taken under assumption that these variables are linearly related. Henceforth, we try to define a linear function that predicts the response value(y) as accurately as possible as a function of the feature or independent variable(x).cut-offs of the colleges, admission intake and preferences of students. Also, it helps students avoid spending time and money on counsellor and stressful research related to finding a suitable college. B. Artificial Neural Network Neural networks are a series of algorithms which try to recognize relationships that are underlying in a dataset through a process which imitates the way human brain operates.

##### Neural networks:

They are referred as a system of neurons. Neural networks are systems that are artificial and they were inspired from biological neural networks. They learn to complete the task from different data sets and from examples without any specific rules assigned for the task. The main inspiration behind is that the system will automatically generate identification characters from the data that has already been passed through without being programmed and basic understanding of these data sets. Artificial Neural Networks have various neurons which are artificial and they are called as units.

The units arrange themselves in a series of layers which all together make up as artificial Neural Networks. Any layer is supposed to have only a millions of units or dozen units as it depends upon the complexity of the system. These units are arranged in a series of layers that together constitute the whole Artificial Neural Networks in a system. A layer can have only a dozen units or millions of units as this depends on the complexity of the system. Commonly, Artificial Neural Network has an input layer, output layer as well as hidden layers. The input layer receives data from the outside world which the neural network needs to analyze or learn about. Then this data passes through one or multiple hidden layers that transform the input into data that is valuable for the output layer. Finally, the output layer provides an output in the form of a response of the Artificial Neural Networks to input data provided.

##### Decision Trees

Classification is a two step process, learning and prediction. At first, model is developed based upon the given training data in its learning step. Then the model is used to predict response for the given data in prediction step. One of the most popular classification algorithms and easiest to learn and understand in decision tree. Decision tree algorithm is also used for solving classification and regression problems. Decision trees use a class label for predicting, for a record it starts from the root of tree. Then compare the values of the root with its record attribute. After the comparison, it follows the branch which is corresponding to the value and jump upon to the next node. There are two types of decision trees, Categorical variable and continuous variable. Categorical variable has a categorical target variable and continuous variable has a continuous target variable. Decision tree has three types of nodes, decision nodes,

chance nodes and end nodes. Decision tree assigns a class label for each leaf node. Even the non-terminal nodes, the root and internal nodes, also contain attribute test conditions to separate records that have different characteristics.

##### Random Forests

The random forest is a machine learning algorithm which is widely used in regression and classification problems. Decision trees are built upon multiple different samples and then take their majority vote for average and bifurcation in case of regression. Random forest has the ability to handle a data set which contains continuous variables in case of regression and categorical variables in case of classification. Hence, it provides good results for classification problems. In industry lingo, reason behind forest works algorithm works so well is: Any huge quantity of moderately uncorrelated trees working as a body will outperform any of the individual constituent models

## CHAPTER 4

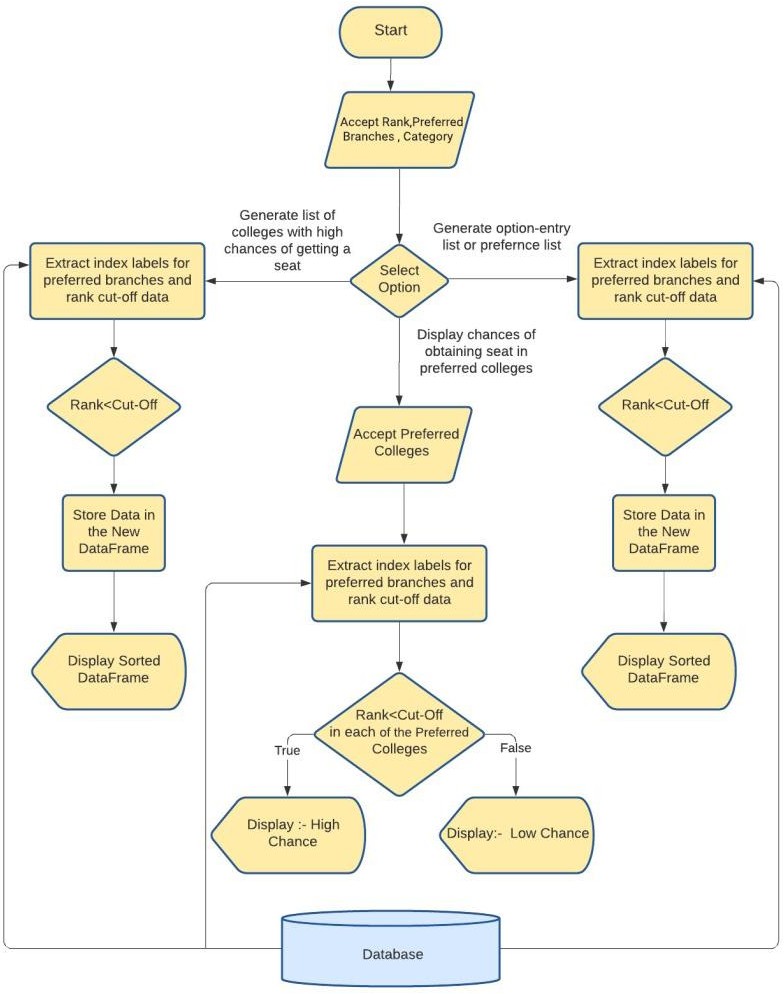
### 4.1 RESULT AND DISCUSSION

### DATASET

The data set comprises of different factors attributed towards picking the right university. It contains data of 100 different students. Data set is classified into 9 different parameters which are considered important during the application for Masters. Those parameters are: **gre** scores, **toefl** scores, university rating, statement of purpose, letter of recommendation, undergraduate **gpa**, research paper, chance of admit.

##### 4.1.5 METHODOLOGY AND IMPLEMENTATION:

Fig -4.1 : Block diagram



The primitive step to building a model for our use case is choosing the right dataset. For our predictions, we chose a dataset which contains all the important attributes that would affect the chances of admit. This is followed by data cleaning where we handle missing values present in various fields. Once the data is ready to be analyzed, we use various tools and libraries to visualize the data and perform analysis. This includes visualizing bar graphs and the correlation matrix. Once the data is ready to be processed, we split it into training and testing data. For this, we will be using 3 machine learning algorithms; linear regression, random forest and neural network. Once these models are built over the dataset, we compare them using key performance indicators. These indicators help us choose the right model for predicting whether an applicant has chances of admission.

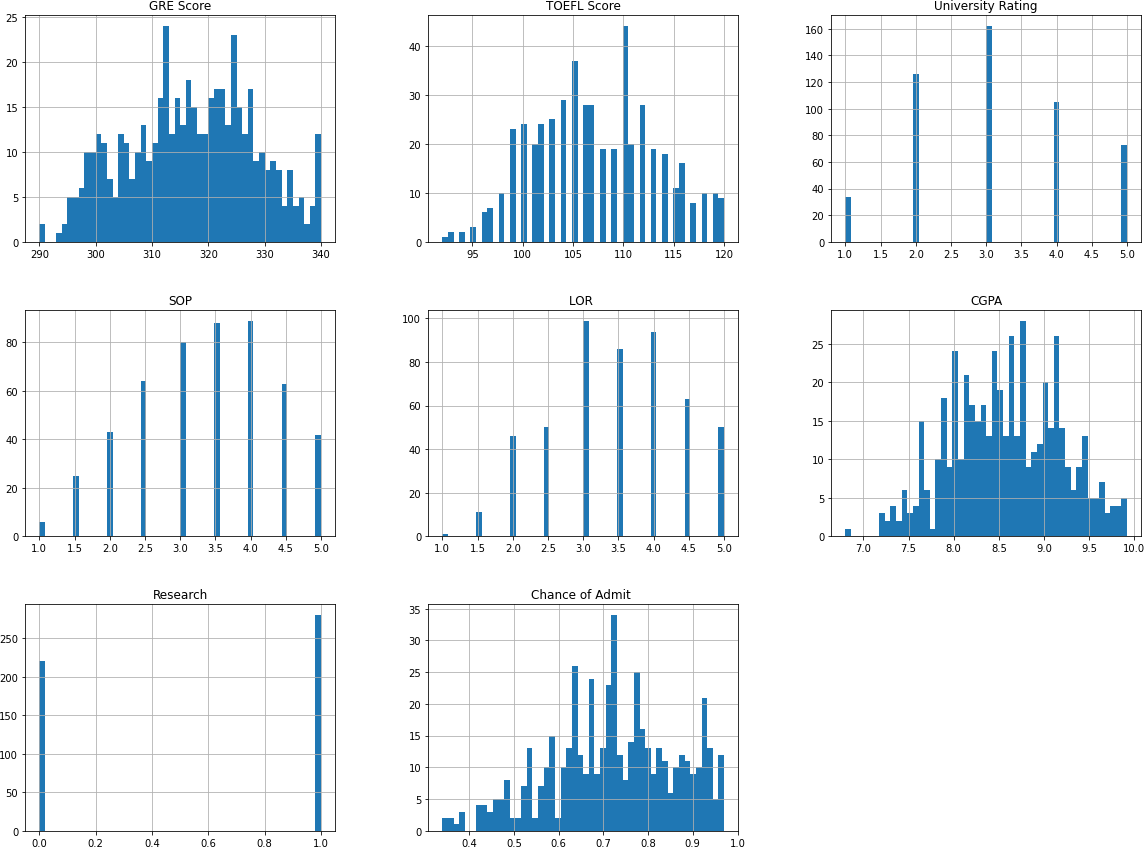
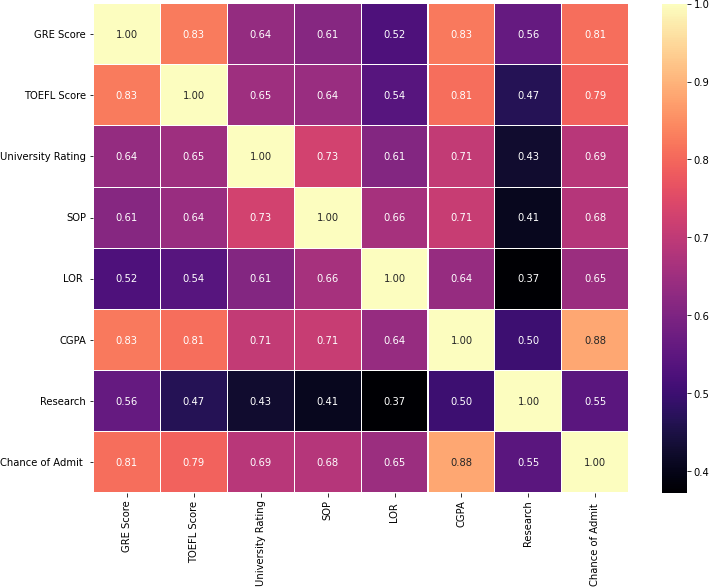


Fig - 3.2 : Data set containing SOP, LOR, CGPA, research



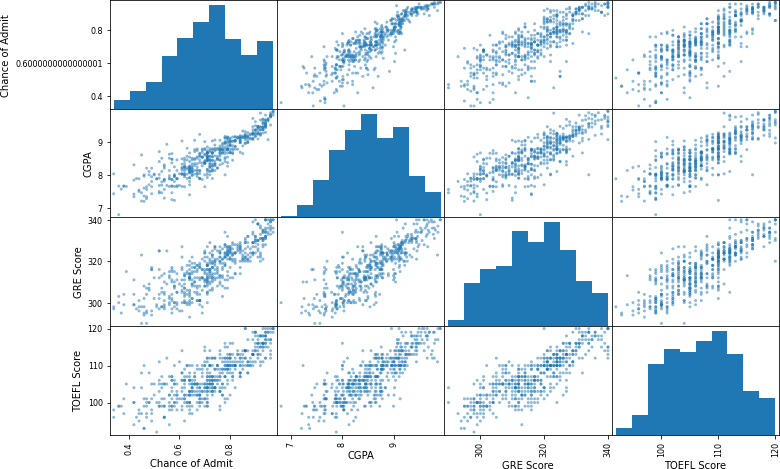


Fig -3.3: Data set containing CGPS, GRE Score, TOEFL Score.

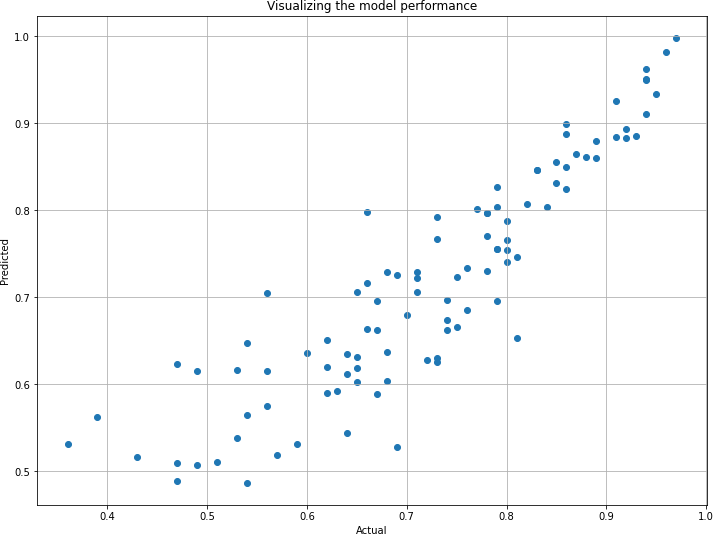


Fig -3.4: Chance to admit

The linear relationship between CGPA and chance of admit is clearly visible. Its relation seems to be the strongest, followed by GRE score and finally TOEFL score. These visual results validate the correlation matrix.

#### The Three Features for Linear Regression

* The first results for Linear Regression (7 features): r\_square score: 0.821208259148699
* The results for Linear Regression now (3 features): r\_square score: 0.8212241793299223
* The two results are very close. If these 3 features (CGPA, GRE SCORE, and TOEFL SCORE) are used instead of all 7 features together, the result is not bad and performance is increased because less calculation is required.

**Accuracy received is 88.88%**

## CHAPTER 5

### 5.1 SUMMARY AND CONCLUSIONS:

Generate list of colleges with higher chances of obtaining a seat. ;- This functionality displays a data-frame with a list consisting of branch, college name, location, and cut-off rank for the preferred branches in the colleges where the candidate has high chances of obtaining a seat, in the inputted category.

Display chances of obtaining a seat in the preferred colleges :- The user may prefer to get a seat in some college. So, the user can input preferred college/colleges. Based on the input given, his/her chances (low/high) of getting a seat in the preferred colleges will be displayed branch- wise according to the preferred branches.

Generating option-entry or preference list ;-This functionality displays a data-frame containing the final option entry list consisting of branch, college name, location, and cut-off rank for the branch in that college, in the inputted category.

The graphs illustrate the chances of getting an admit against the predicted chances of getting an admit using different algorithms. Comparison in each of these models is done by evaluating Key Performance Indicators (KPI). With the help of this, it provides a cleaner output and helps in comparing the indicators like Root Mean Square Error, Mean Square Error, Mean Absolute Error and Adjusted R-Squared of an algorithm. At times accuracy provides ambiguous results if there are unequal observations or multiple classes are present in the dataset. Results show us that the highest accuracy is achieved through the linear regression model and the decision tree has the lowest accuracy.

This information can be visualized through the abovementioned graphs. Graphs that are not spread out and are closer to the line x=y have higher accuracy. The linear regression line graph is surrounded around this line and hence has the highest accuracy. The decision tree model has the lowest accuracy and is spread out with a lot of outliers, hence depicting that getting an accurate result using this model would be inaccurate

## CHAPTER 6

**6.1 LITERATURE CITED**

##### REFERENCES

1] [1] Subba Reddy.Y and Prof. P. Govindarajulu,” A survey on data mining and machine learning techniques for internet voting and product/service selection”, IJCSNS International Journal of Computer Science and Network Security, VOL.17 No.9 September [2017].

1. Zhibo Wang, Jilong Liao, Qing Cao, Hairong Qi, and Zhi Wang, “Friend book: A Semanticbased Friend Recommendation System for Social Networks IEEE Transactions on Mobile Computing, [2018].
2. J. Bobadilla et al. “Knowledge-Based System” Elsevier B.V. [2017]
3. Hector Nunez, Miquel sanchez-Marre, Ulises Cortes, Joaquim Comas, Montse Martinez, Ignasi Rodriguez Roda, Manel Poch, “A Comaprative study on the use of similarity measure in case based reasoning to improve the classification of environmental system situations,”, ELSEVIER, Environmental Modeling and Software (2018).
4. DINO IENCO, RUGGERO G. PENSA and ROSA MEO, “From Context to Distance: Learning Dissimilarity for categorical Data Clustering,” Journal Vol. X. 10 2017, pages 1- 10.[2017].
5. Duc Thang Nguyen, Lihui Chen, Chee keong Chan, “Clustering with Multi viewpoint Based Similarity Measure,” IEEE Transactions on Knowledge and Data Engineering. Vol. 24. No. 6. June [2012].
6. Deokate monali, Gholave Dhanashri, Jarad Dipali, Khomane Tejaswini. College Recommendation System for Admission. International Research Journal of Engineering and Technology,[2018]