UNIVERSITY ADMISSION

PREDICTION





INTRODUCTION

RELATED WORK

TYPES OF MACHINE LEARNING

STEPS INVOLVED

DATA VISUALIZATION

PREDICTIVE MODELING

LIMITATIONS

CONCLUSION

INTRODUCTIO
N

- *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.*
- ✓ *The majority of international students studying in the USA are from India and China. In the past decade*
- ✓ *It is seen that the number of students pursuing Masters in Computer Science field from universities in the USA is quite high and we will be Focusing on those students.*
- ✓ *Majority of universities in the USA follow similar guidelines for providing admission to students.*
- ✓ *Universities taking to consideration different factors like GRE,TOEFL,SOP and undergraduate degree score.*
- ✓ *Based on the overall profile of the student decision is taken by the universities admission team to admit or reject a particular candidate*

Objectives of Research

- *The primary objective is to solve the problems of students while applying for the universities in abroad.*
- *We will be developing a university admission predictor system which will help the students to predict the chances of their application being selected for a particular university they wish to apply.*
- *Multiple machine learning algorithms were evaluated to develop the system*
- *Also we will be creating a simple user interface which will help the users to input the data related to student profile and get thepredicted resultfor the application based on the profile as output.*
- *This will thus eventually help students saving the extra amount of time and money they have to spend at the education consultancy firms .*

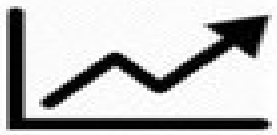
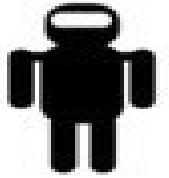
Related wORK

- *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*
- *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. 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. 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. The objective of the model was to increase the enrolment of the students in the university.*
- *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.*
- *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.*
- *Nadeshwar created a model using the Multiple Logistic regression algorithm, it was able to achieve accuracy rate of 67% only*

Types of Machine Learning

Machine Learning

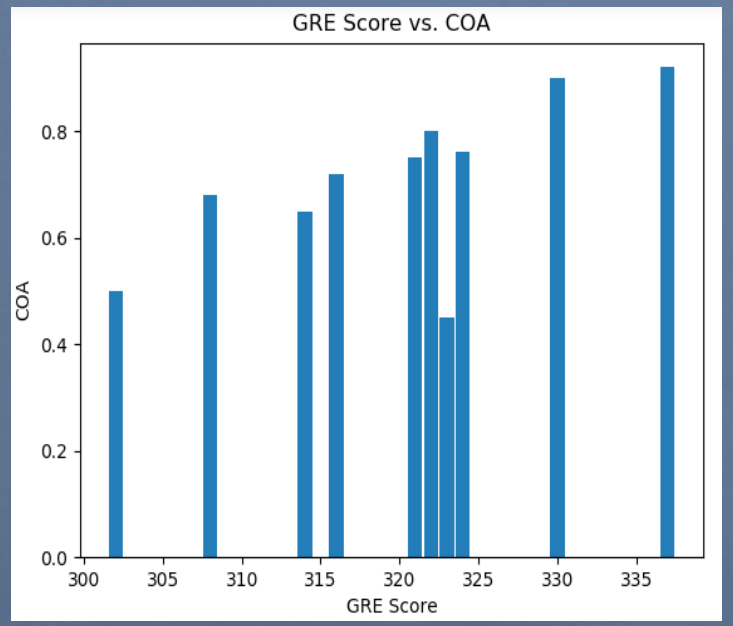
- **Supervised**
  - Task Driven (Predict next value)
  - 
- **Unsupervised**
  - Data Driven (Identify Clusters)
  - 
- **Reinforcement**
  - Learn from Mistakes
  - 

STEPS INVOLVED

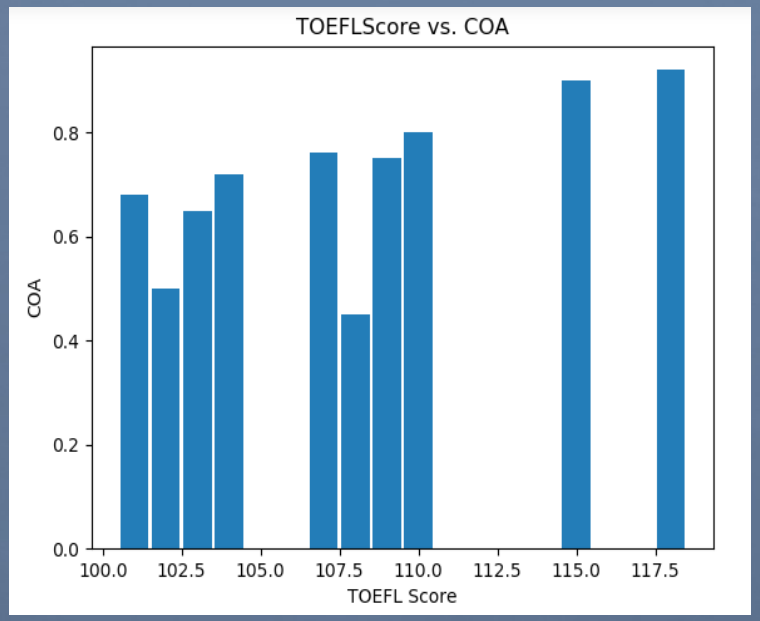
1. **Define the problem**
2. **Generate your own hypothesis**
3. **Get the Dataset**
4. **Data Cleaning**
5. **Exploratory Data Analysis**
6. **Predictive Modeling**

DATA VISUALIZATION

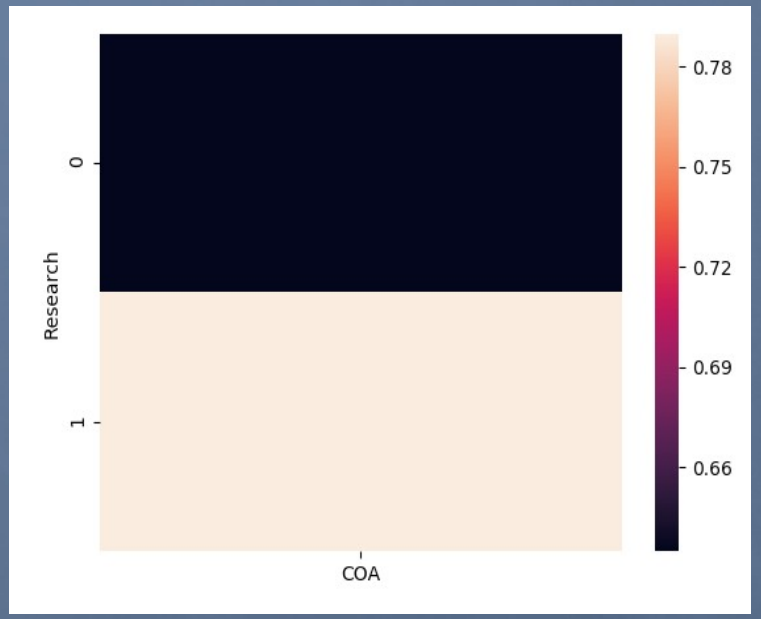
- We will visualize the data using matplotlib
- We will compare between different features of the data set to get to know which has highest effect on the targeted variable



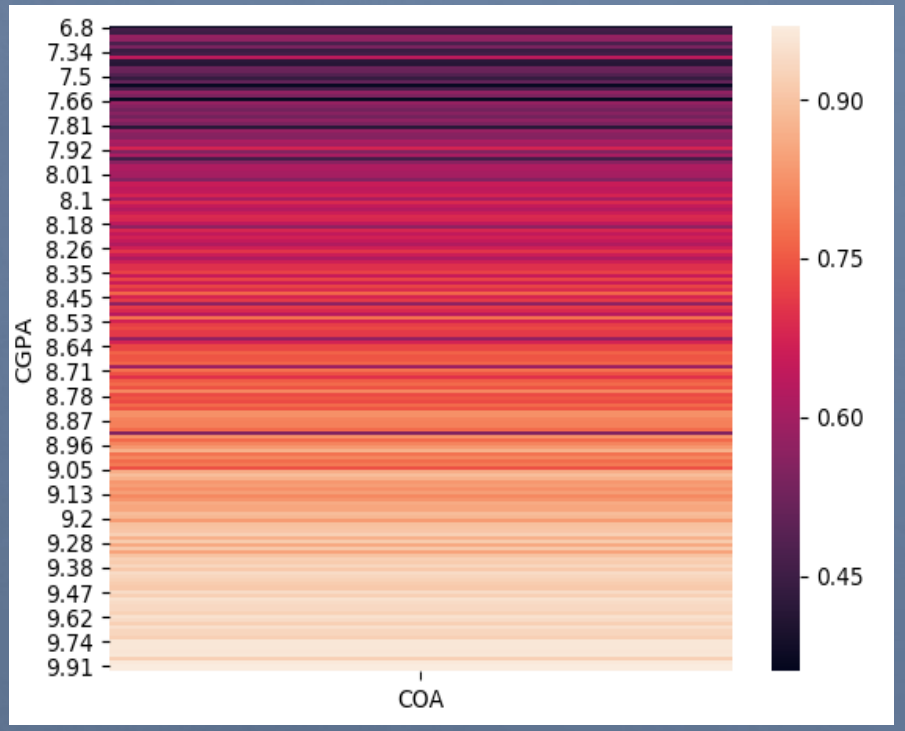
GRE Score Vs COA Bar graph



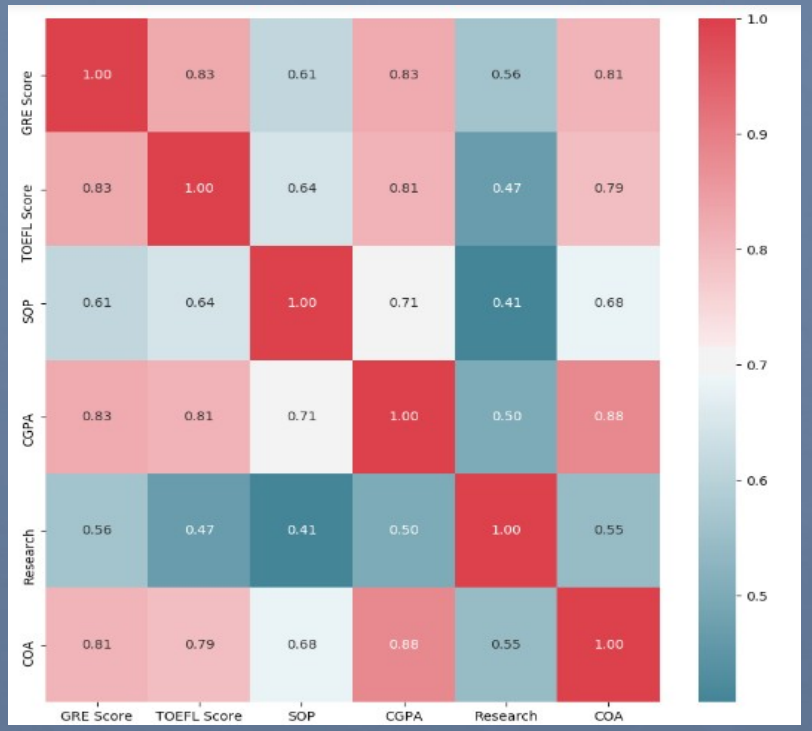
TOEFL Score Vs COA Bar graph



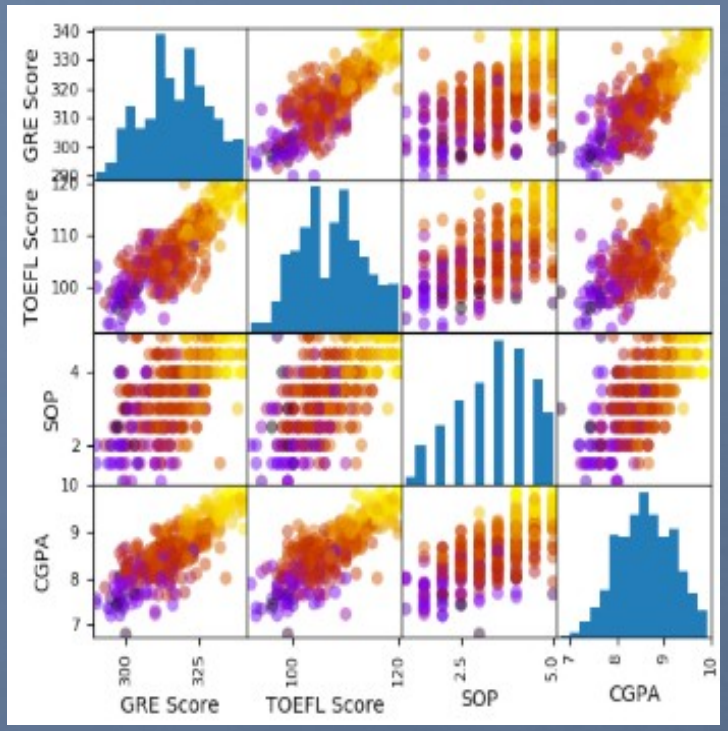
Research Vs COA heat map



CGPA VS Chance of Admit heat map



Correlation heat map



comparison between all features

IMPORTING LIBRARIES

We have to import all the required packages

```
%matplotlib notebook
import seaborn as sns
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
```

ill import a set of regressors to predict the required model and will split the data using train test sp

In [1]:
```
from sklearn.neighbors import KNeighborsRegressor
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import Ridge
from sklearn.ensemble import RandomForestRegressor
```

In [ ]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y,random_state=0)
```

PREDICTIVE MODELING

```
print(knnreg.predict(X_test))
print('R-squared test score: {:.3f}'
     .format(knnreg.score(X_test, y_test)))
```

```
[0.73  0.872 0.8   0.87  0.606 0.684 0.534 0.642 0.576 0.762 0.678 0.726
 0.638 0.708 0.906 0.468 0.816 0.638 0.804 0.664 0.526 0.726 0.532 0.926
 0.842 0.716 0.762 0.606 0.546 0.544 0.65  0.512 0.666 0.82  0.696 0.552
 0.826 0.86  0.62  0.64  0.65  0.866 0.918 0.738 0.812 0.794 0.93  0.816
 0.648 0.714 0.642 0.638 0.628 0.806 0.678 0.864 0.544 0.83  0.612 0.63
 0.442 0.878 0.626 0.87  0.752 0.638 0.784 0.86  0.494 0.798 0.912 0.74
 0.676 0.734 0.894 0.942 0.796 0.716 0.516 0.676 0.642 0.818 0.608 0.708
 0.82  0.564 0.762 0.658 0.528 0.794 0.732 0.834 0.612 0.69  0.61  0.936
 0.59  0.632 0.626 0.608 0.938 0.754 0.936 0.668 0.752 0.678 0.822 0.822
 0.64  0.728 0.504 0.624 0.618 0.592 0.608 0.86  0.586 0.942 0.824 0.916
 0.668 0.804 0.9   0.892 0.696]
R-squared test score: 0.720
```

KNN Regression Model

```
print('University data set')
print('ridge regression linear model intercept: {}'
     .format(linridge.intercept_))
print('ridge regression linear model coeff:\n{}'
     .format(linridge.coef_))
print('R-squared score (training): {:.3f}'
     .format(linridge.score(X_train, y_train)))
print('R-squared score (test): {:.3f}'
     .format(linridge.score(X_test, y_test)))
print('Number of non-zero features: {}'
     .format(np.sum(linridge.coef_ != 0)))
```

```
University data set
ridge regression linear model intercept: -1.5461959976122706
ridge regression linear model coeff:
[0.00333624 0.00415488 0.0175384  0.08102205 0.02155746]
R-squared score (training): 0.808
R-squared score (test): 0.785
Number of non-zero features: 5
```

Ridge Regression Model

```
print('linear model coeff (w): {}'
     .format(linreg.coef_))
print('linear model intercept (b): {:.3f}'
     .format(linreg.intercept_))
print('R-squared score (training): {:.3f}'
     .format(linreg.score(X_train, y_train)))
print('R-squared score (test): {:.3f}'
     .format(linreg.score(X_test, y_test)))
```

```
linear model coeff (w): [0.00203909 0.00273575 0.01150513 0.12975436 0.02593185]
linear model intercept (b): -1.384
R-squared score (training): 0.819
R-squared score (test): 0.796
```

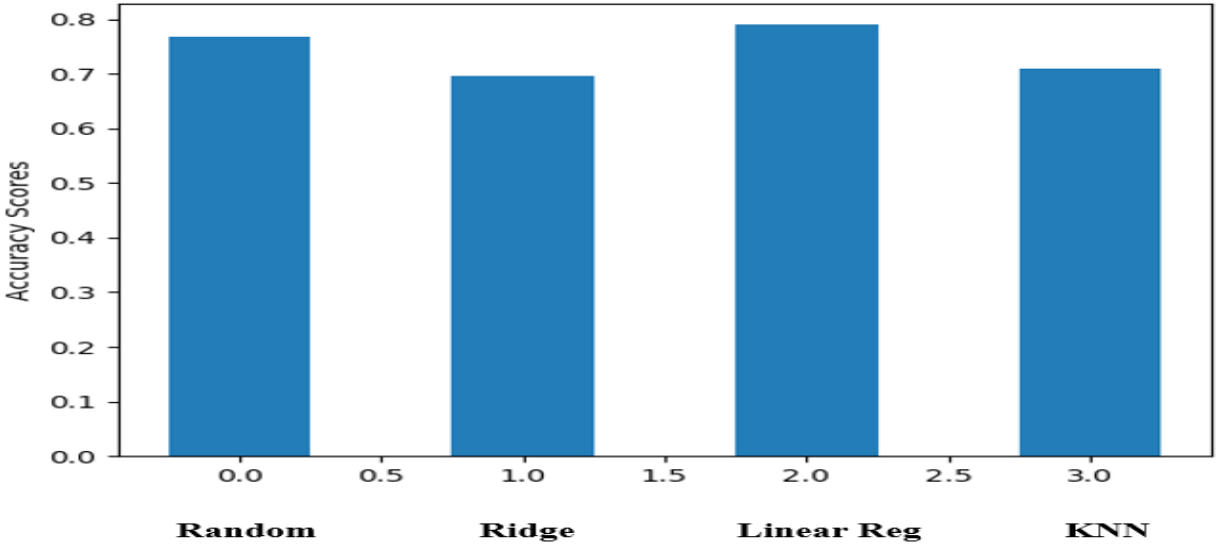
Linear Regression Model

```
print('University dataset')
print('Accuracy of RF classifier on training set: {:.2f}'
     .format(clf.score(X_train, y_train)))
print('Accuracy of RF classifier on test set: {:.2f}'
     .format(clf.score(X_test, y_test)))
```

```
University dataset
Accuracy of RF classifier on training set: 0.95
Accuracy of RF classifier on test set: 0.77
```

Random Forest Model

COMPARISON OF ACCURACY BETWEEN DIFFERENT MODELS



SCORES	RANDOM	RIDGE	LINEAR REGRESSION	KNN
TRAIN	0.95	0.80	0.81	0.85
TEST	0.77	0.78	0.79	0.72

LIMITATIONS

As discussed earlier in the limitation of the research we have created the models based only on the data of Indian Students studying Masters in Computer Science in the USA, we have considered only ten 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 web-based application by making changes to the Node-Red. 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 by changing the server code in the Node-Red.

CONCLUSION

- ✓ 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. Linear Regression model compared with other models. The model can be used by the students for evaluating their chances of getting shortlisted in a particular university with an average accuracy of 79%.
- ✓ A simple user interface was developed to make the application interactive and easy to use for the users from the non-technical background. Node-Red 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.



Presented by:

P.Rohith

M.Sahith

L.Nikhil Sai

M.Chandu Nath

N.Nikhil Chowdary