

1. INTRODUCTION

1.1 Overview

The world markets are developing rapidly and continuously looking for the best knowledge and experience among people. Young workers who want to stand out in their jobs are always looking for higher degrees that can help them in improving their skills and knowledge. As a result, the number of students applying for graduate studies has increased in the last decade. This fact has motivated us to study the grades of students and the possibility of admission for master's programs that can help universities in predicting the possibility of accepting master's students submitting each year and providing the needed resources.

The dataset presented in this paper is related to the educational domain. Admission is a dataset with 400 rows that contains 7 different independent variables which are: Graduate Record Exam1 (GRE) score. Test of English as a Foreigner Language2 (TOEFL) score. University Rating (Uni Rating) that indicates the Bachelor University ranking among the other universities. Statement of purpose (SOP) which is a document written to show the candidate's life, ambition and the motivations for the chosen degree/ university. Letter of Recommendation Strength (LOR) which verifies the candidate's professional experience, builds credibility, boosts confidence and ensures your competency. Undergraduate GPA (CGPA) out of 10. Research Experience that can support the application, such as publishing research papers in conferences, working as research assistant with university professors. One dependent variable can be predicted which is the chance of admission that is according to the input given will be ranging from 0 to 1. We are developing several Regression Models which are Logistic Regression, Random forest Regression, Multiple Linear Regression using Dimensionality reduction and Random forest Regression using Dimensionality reduction to find the accuracy of those models. Out of those we use high accuracy models.

So, we believe that a predictive model generated using all the past data can be a useful resource to predict the outcome for the applicants.

In the Guided Project, our goal is to predict the outcome of applications that are filed by many students every year.

1.2 Purpose

Our project aims that this prediction algorithm could be a useful resource for the students to easily process the admission to their preferred university .

In order to predict the case status of the applicants, we will be feeding the model with the dataset which contains the required fields by which the machine can classify the case status as eligible or ineligible.

2.LITERATURE SURVEY

2.1 Existing Problem

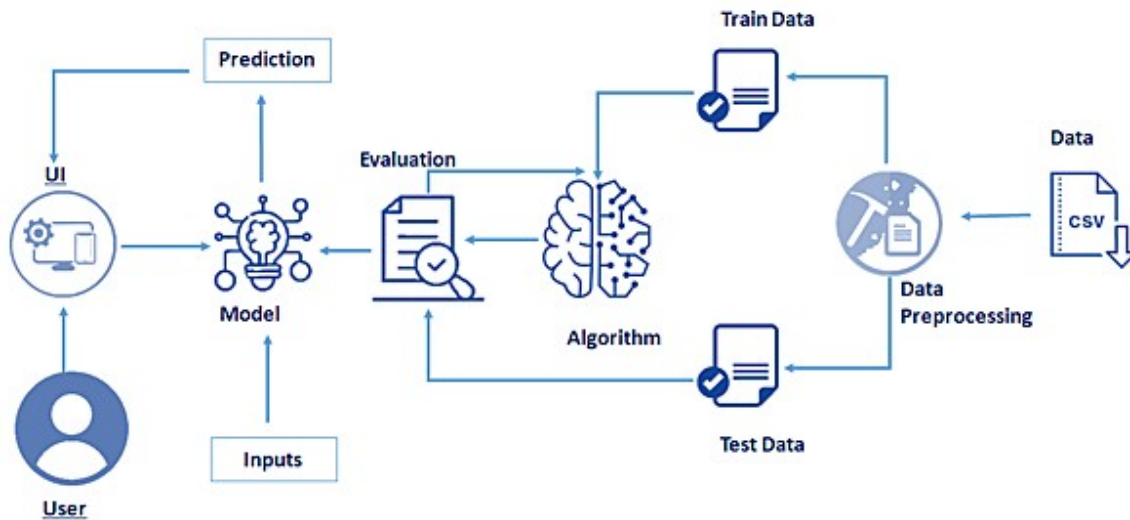
In India every year lacks students getting the graduation degree and willing to join post-graduation in other countries. Newly graduate students usually are not knowledgeable of the requirements and the procedures of the postgraduate admission and might spend a considerable amount of money to get advice from consultancy organizations to help them identify their admission chances. Human consultants and calculations might be biased and inaccurate.

2.2 Proposed Solution

This paper helps in predicting the eligibility of Indian students getting admission in the best university based on their Test attributes like GRE, TOEFL, LOR, CGPA etc. according to their scores the possibilities of chance of admit is calculated.

3.THEORITICAL ANALYSIS

3.1 Block Diagram



3.2 Hardware / Software designing

- IBM Watson Studio

Watson Studio provides you with the environment and tools to solve your business problems by collaboratively working with data. It provides a suite of tools for data scientists, application developers and subject matter experts, allowing them to collaboratively connect to data, wrangle that data and use it to build, train and deploy models at scale. Successful AI projects require a combination of algorithms + data + team, and a very powerful compute infrastructure.

- IBM Watson Machine Learning

IBM Watson Machine Learning is a full-service IBM Cloud offering that makes it easy for developers and data scientists to work together to integrate predictive capabilities with their applications. The Machine Learning service is a set of REST APIs that you can call from any programming language to develop applications that make smarter decisions, solve tough problems, and improve user outcomes.

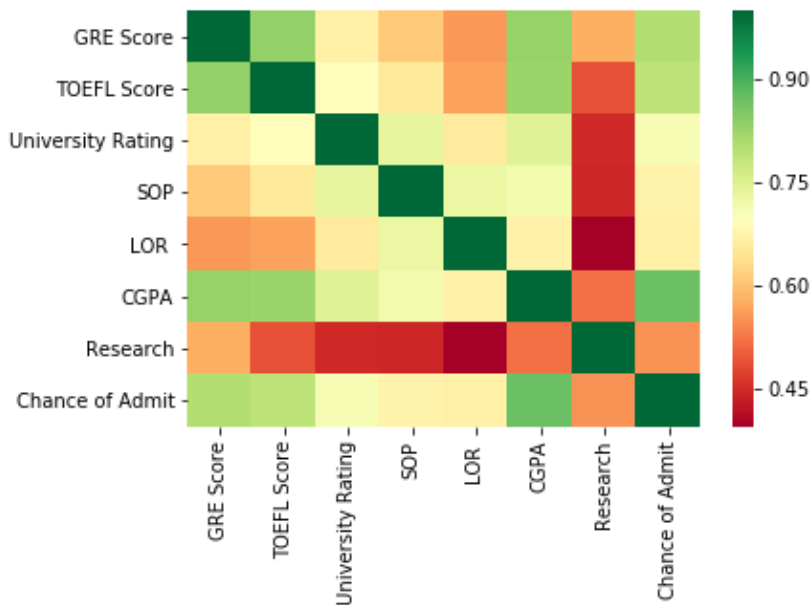
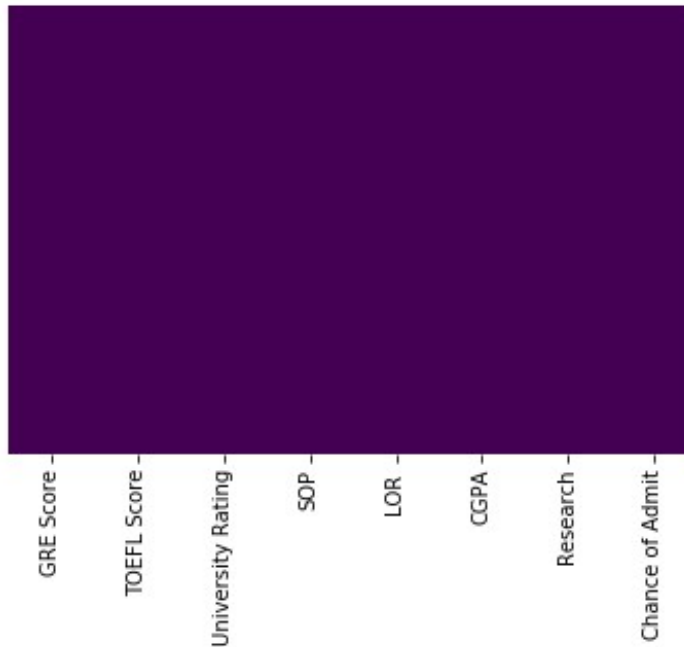
- IBM Cloud Object Storage

IBM Cloud Object Storage is an application-data archive and backup platform that offers persistent cloud storage and data encryption by default. Made for electronic records retention, it helps with regulatory compliance.

4. EXPERIMENTAL INVESTIGATIONS

The dataset [5] presented in this paper is related to the educational domain. Admission is a dataset with 500 rows that contains 7 different independent variables which are:

- Graduate Record Exam1 (GRE) score. The score will be out of 340 points.
- Test of English as a Foreigner Language2 (TOEFL) score, which will be out of 120 points.
- University Rating (Uni.Rating) that indicates the Bachelor University ranking among the other universities. The score will be out of 5
- Statement of purpose (SOP) which is a document written to show the candidate's life, ambition and the motivations for the chosen degree/ university. The score will be out of 5 points.
- Letter of Recommendation Strength (LOR) which verifies the candidate professional experience, builds credibility, boosts confidence and ensures your competency. The score is out of 5 points
- Undergraduate GPA (CGPA) out of 10
- Research Experience that can support the application, such as publishing research papers in conferences, working as research assistant with university professors (either 0 or 1). One dependent variable can be predicted which is chance of admission that according to the input given will be ranging from 0 to 1.

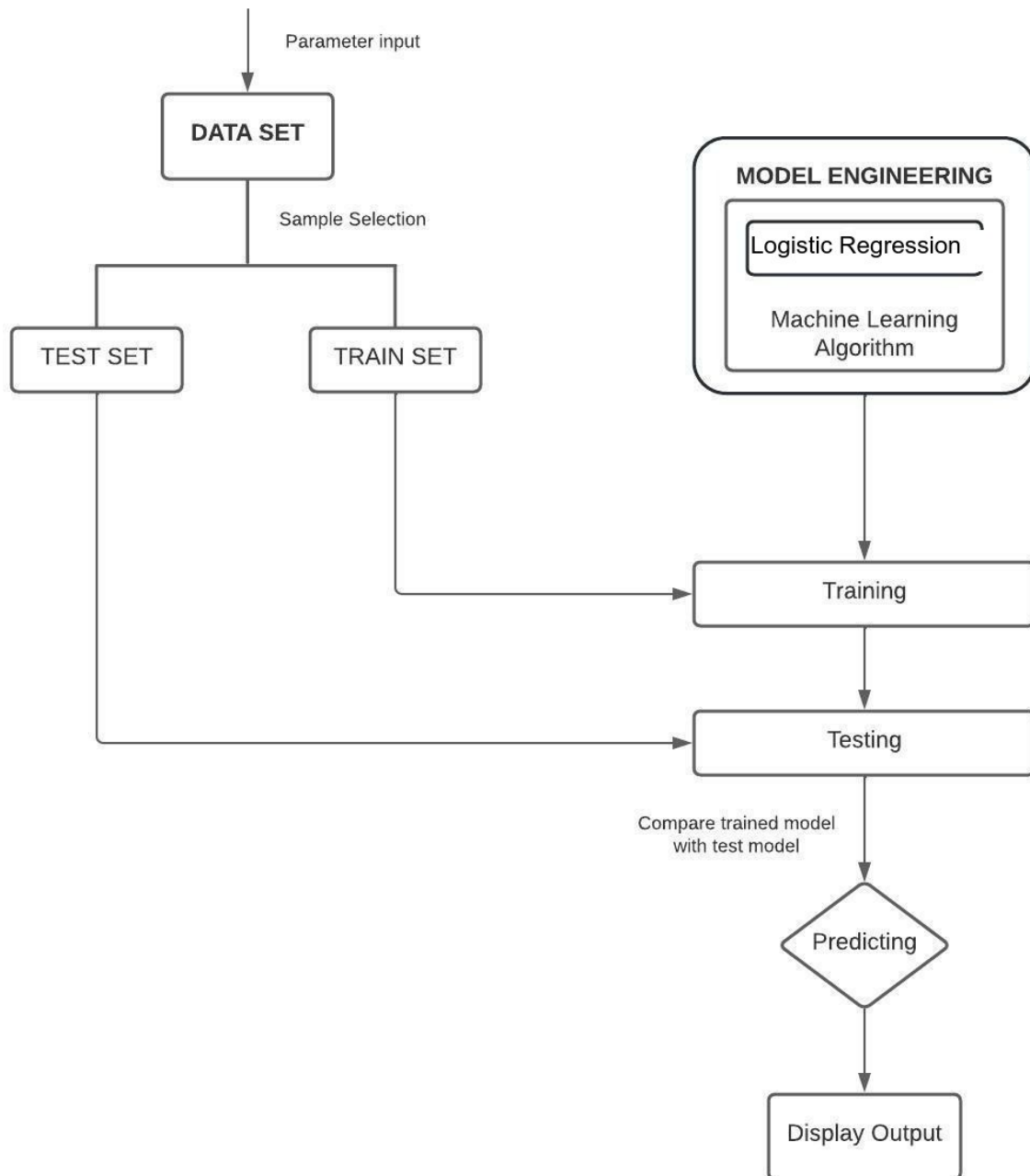


The first figure is a heat map used to represent the missing values in the dataset. We can easily identify if there are any missing values in the dataset. Since there are no missing values the heat map is empty. From this we can understand that the dataset is complete and there is no need to change anything in the dataset.

The second figure is also a heat map. We have to find the correlation of the variables using pandas and represent it in a heatmap using seaborn. So when

analyzing the heat map we can understand the chance of admission mainly depends on CGPA, GRE, TOFEL and less on statement of purpose, strength of the recommendation letter and research. Like this we have various plots and graphs that can be used to thoroughly identify the data sets and reach at conclusions about admission eligibility. these processes comes under Exploratory Data Analysis

5. FLOWCHART



6. RESULT

The final result of the project is to predict the certification status of the University admission Prediction applications.

← → ↻ 127.0.0.1:5000

Apps https://prezi.com/p... New folder Classes rejsir sendoff offici... Meet - bqz-myba-e... Meet - qqv-rcjz-ksq Meet - knw-vhbz-ran >> Other bookmarks Reading list

UNIVERSITY ADMISSION PREDICTION SYSTEM

Enter your details and get probability of your admission

Enter GRE Score

Enter TOEFL Score

Select University no

☐ 1

☐ 2

☐ 3

☐ 4

☐ 5

Enter SOP

Enter LOR

Enter CGPA

Research

☐ Research

☐ NO Research

← → ↻ 127.0.0.1:5000

Apps https://prezi.com/p... New folder Classes rejsir sendoff offici... Meet - bqz-myba-e... Meet - qqv-rcjz-ksq Meet - knw-vhbz-ran >> Other bookmarks Reading list

UNIVERSITY ADMISSION PREDICTION SYSTEM

Enter your details and get probability of your admission

Enter GRE Score

Enter TOEFL Score

Select University no

☒ 1

☐ 2

☐ 3

☐ 4

☐ 5

Enter SOP

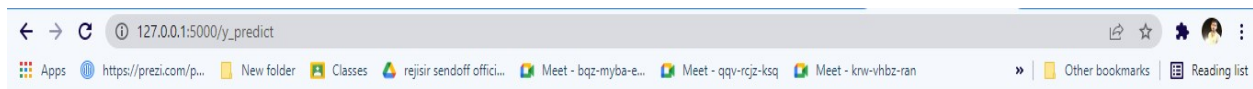
Enter LOR

Enter CGPA

Research

☒ Research

☐ NO Research



Predicting Chance of Admission

A Machine Learning Web App using Flask.

Prediction : You have a chance



7. ADVANTAGES & DISADVANTAGES

The main advantage of this proposed application is reduction of time. One can infer from these applications to know the case status of the application as eligible or ineligible. we don't need to check all data to know whether it is accepted or rejected. it takes necessary data and makes predictions. Hence it helps to reduce a huge amount of time for checking each and every application.

Disadvantages

- Need more datasets, to increase the accuracy of the algorithms.
- A large amount of data is used in the process of training and learning, so the use of data should be of good quality, unbiased.
- The proposed application can only be used by students or faculties for prediction.
- The proposed application is Web-based, hence cannot be used in Mobile devices.
- The result of the application depends upon the accuracy of the algorithms

8. APPLICATIONS

University admissions require a lot of time and effort. With our prediction application we round up the favorable universities by thoroughly analyzing the existing admission criteria. The application provides high accuracy and reliability. Therefore the applicant can take a lot out of their plates.

9.CONCLUSION

After successfully completing the project we learned about various machine learning models that were performed to predict the opportunity of a student to get admitted to a master's program. The machine learning models included are logistic regression, k-nearest neighbor, random forest, and linear regression. Experiments show that the logistic regression model surpasses other models on accuracy . As for the future work, more experiments can be conducted on models to find out which model gives the best performance.

10. FUTURE SCOPE

As the system allows students to save the lot of time and money that they would spend on educational mentors and application fees for colleges where they have less chances of getting admissions. The main limitation of this research is we developed models based solely on data from Indian Students studying Masters in Computer Science in the United States, we considered only few universities with different rankings. More information relating to new colleges and courses can be added to the curriculum in the future. The system may also be modified to a web-based application by making node-red modifications. To solve the problem, it is possible to test other classification algorithms if they have high accuracy score than the current algorithm, the framework can be easily modified to support the new algorithm by changing the server code in the Node Red. Finally students can have an open source machine Learning model which will help the students to know their chance of admission into a particular university with high accuracy.

11. BIBLIOGRAPHY

Import Necessary Libraries

Let us import necessary libraries to get started!

```
In [2]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

The Data

Let's start by reading in the Admission_Predict.csv file into a pandas dataframe.

```
In [77]: #read_csv is a pandas function to read csv files
data = pd.read_csv('Admission_Predict.csv')
```

```
In [78]: #head() method is used to return top n (5 by default) rows of a DataFrame or series.
data.head()
```

```
Out[78]:   Serial No.  GRE Score  TOEFL Score  University Rating  SOP  LOR  CGPA  Research  Chance of Admit
```

```
Out[78]:
```

	Serial No.	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
0	1	337	118	4	4.5	4.5	9.65	1	0.92
1	2	324	107	4	4.0	4.5	8.87	1	0.76
2	3	316	104	3	3.0	3.5	8.00	1	0.72
3	4	322	110	3	3.5	2.5	8.67	1	0.80
4	5	314	103	2	2.0	3.0	8.21	0	0.65

```
In [79]: #Let us drop Serial No. Column as it is not required for prediction
data.drop(["Serial No."],axis=1,inplace=True)
data.head()
```

```
Out[79]:
```

	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
0	337	118	4	4.5	4.5	9.65	1	0.92
1	324	107	4	4.0	4.5	8.87	1	0.76
2	316	104	3	3.0	3.5	8.00	1	0.72
3	322	110	3	3.5	2.5	8.67	1	0.80
4	314	103	2	2.0	3.0	8.21	0	0.65

#describe() method computes a summary of statistics like count, mean, standard deviation, min, max and quartile values.

```
In [80]: data.describe()
```

```
In [80]: data.describe()
```

```
Out[80]:
```

	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
count	400.000000	400.000000	400.000000	400.000000	400.000000	400.000000	400.000000	400.000000
mean	316.807500	107.410000	3.087500	3.400000	3.452500	8.598925	0.547500	0.724350
std	11.473646	6.069514	1.143728	1.006869	0.898478	0.596317	0.498362	0.142609
min	290.000000	92.000000	1.000000	1.000000	1.000000	6.800000	0.000000	0.340000
25%	308.000000	103.000000	2.000000	2.500000	3.000000	8.170000	0.000000	0.640000
50%	317.000000	107.000000	3.000000	3.500000	3.500000	8.610000	1.000000	0.730000
75%	325.000000	112.000000	4.000000	4.000000	4.000000	9.062500	1.000000	0.830000
max	340.000000	120.000000	5.000000	5.000000	5.000000	9.920000	1.000000	0.970000

From the data we infer that there are only decimal values and no categorical values

```
In [81]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 400 entries, 0 to 399
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   GRE Score              400 non-null   int64
1   TOEFL Score            400 non-null   int64
2   University Rating      400 non-null   int64
3   SOP                    400 non-null   float64
4   LOR                    400 non-null   float64
5   CGPA                   400 non-null   float64
6   Research               400 non-null   int64
7   Chance of Admit        400 non-null   float64
dtypes: float64(4), int64(4)
memory usage: 25.1 KB
```

```
In [82]: #Let us rename the column Chance of Admit because it has trailing space
data=data.rename(columns = {'Chance of Admit ':'Chance of Admit'})
```

Exploratory Data Analysis

Missing Data

We can use seaborn to create a simple heatmap to see where we have missing data!

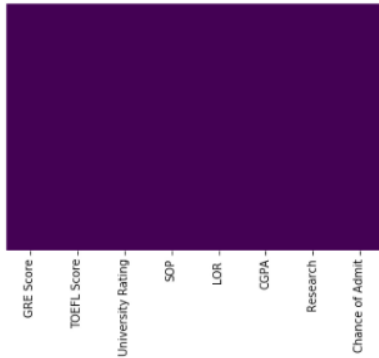
```
In [101]: data.isnull().any()
```

```
Out[101]: GRE Score          False
TOEFL Score          False
University Rating     False
SOP                   False
LOR                   False
CGPA                  False
Research              False
Chance of Admit       False
dtype: bool
```

Heatmap:It is way of representing the data in 2-D form.It gives coloured visual summary of the data

```
In [102]: sns.heatmap(data.isnull(),yticklabels=False,cbar=False,cmap='viridis')
```

```
Out[102]: <matplotlib.axes._subplots.AxesSubplot at 0x2b6e1bb0408>
```



From the heatmap,we see that there are no missing values in the dataset

```
In [103]: data.corr()
```

```
In [103]: data.corr()
```

```
Out[103]:
```

	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
GRE Score	1.000000	0.835977	0.668976	0.612831	0.557555	0.833060	0.580391	0.802610
TOEFL Score	0.835977	1.000000	0.695590	0.657981	0.567721	0.828417	0.489858	0.791594
University Rating	0.668976	0.695590	1.000000	0.734523	0.660123	0.746479	0.447783	0.711250
SOP	0.612831	0.657981	0.734523	1.000000	0.729593	0.718144	0.444029	0.675732
LOR	0.557555	0.567721	0.660123	0.729593	1.000000	0.670211	0.396859	0.669889
CGPA	0.833060	0.828417	0.746479	0.718144	0.670211	1.000000	0.521654	0.873289
Research	0.580391	0.489858	0.447783	0.444029	0.396859	0.521654	1.000000	0.553202
Chance of Admit	0.802610	0.791594	0.711250	0.675732	0.669889	0.873289	0.553202	1.000000

```
In [104]: corr = data.corr()
sns.heatmap(corr,xticklabels=corr.columns.values,
            yticklabels=corr.columns.values,cmap="RdYlGn")
```

```
Out[104]: <matplotlib.axes._subplots.AxesSubplot at 0x2b6e1c0d588>
```

