

Prediction of Graduate Admissions with Artificial Neural Network

I. Introduction

Prospective graduate students always face a dilemma deciding universities of their choice while applying to master's programs. While there are a good number of predictors and consultancies that guide a student, they aren't always reliable since decision is made on the basis of select past admissions

This dataset is created for prediction of Graduate Admissions from an Indian perspective. The dataset contains several parameters which are considered important during the application for Masters Programs. This dataset was built with the purpose of helping students in shortlisting universities with their profiles. The predicted output gives them a fair idea about their chances for a particular university.

II. Dataset

The dataset is owned by Mohan S Acharya.

Dataset Source: <https://www.kaggle.com/mohansacharya/graduate-admissions>

Dataset Citation: Mohan S Acharya, Asfia Armaan, Aneeta S Antony : A Comparison of Regression Models for Prediction of Graduate Admissions, IEEE International Conference on Computational Intelligence in Data Science 2019

The parameters included are:

GRE Scores	out of 340
TOEFL Scores	out of 120
University Rating	out of 5
Statement of Purpose and Letter of Recommendation Strength	out of 5
Undergraduate GPA	out of 10
Research Experience	either 0 or 1
Chance of Admit	ranging from 0 to 1

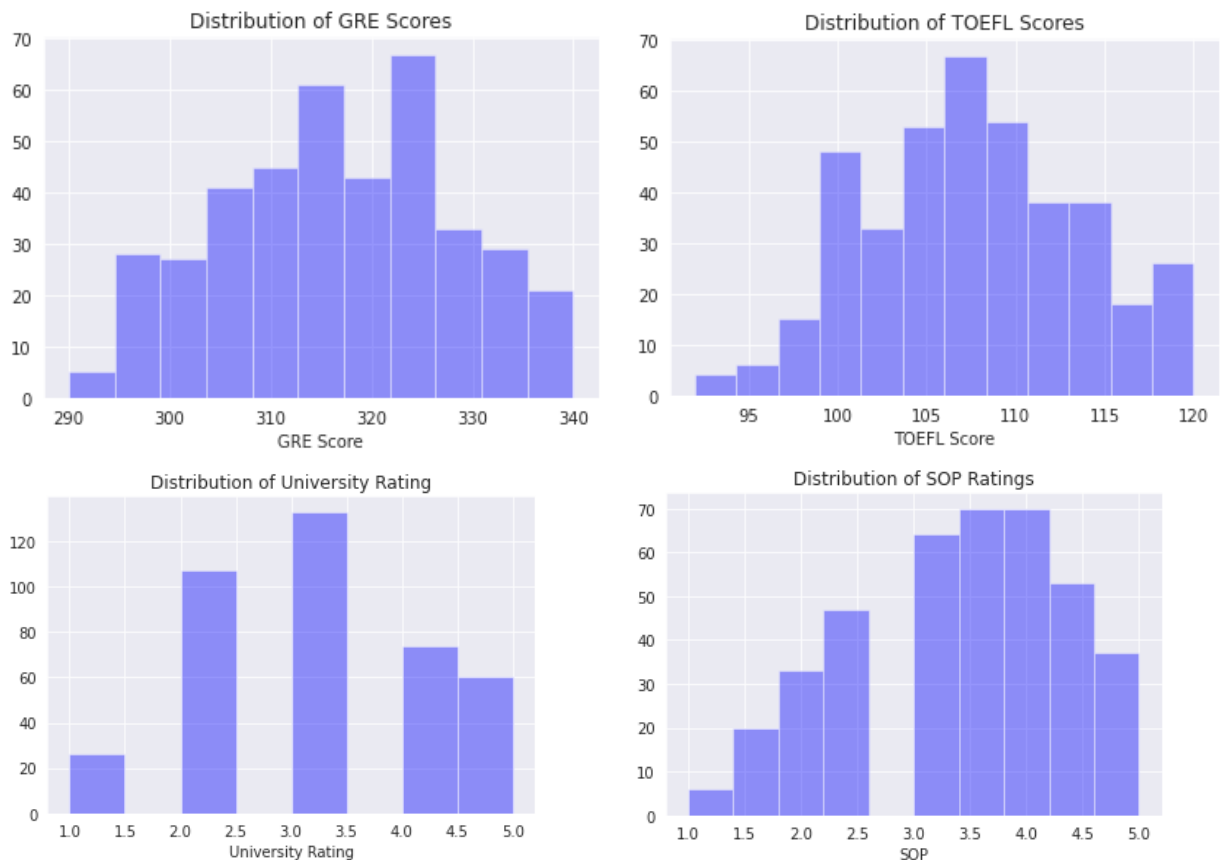
	Serial No.	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
0	1	337	118	4	4.5	4.5	9.65	1	0.92
1	2	324	107	4	4.0	4.5	8.87	1	0.76
2	3	316	104	3	3.0	3.5	8.00	1	0.72
3	4	322	110	3	3.5	2.5	8.67	1	0.80
4	5	314	103	2	2.0	3.0	8.21	0	0.65

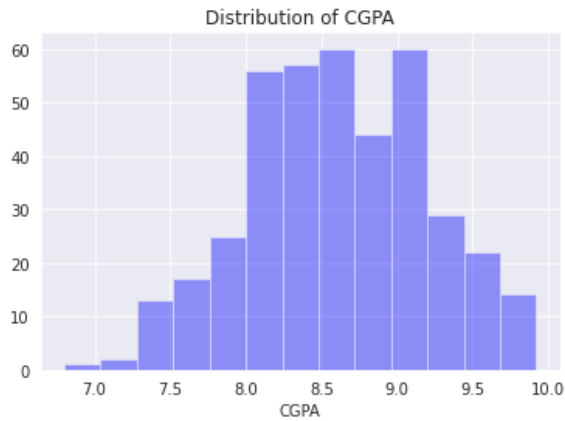
III. Exploratory data analysis

As we can see that the Serial number is not beneficial for the prediction as it just acts like an Identification for each data point. Therefore, we will remove the columns by dropping it. And check for any null data.

```
GRE Score      0
TOEFL Score    0
University Rating 0
SOP            0
LOR            0
CGPA           0
Research       0
Chance of Admit 0
dtype: int64
```

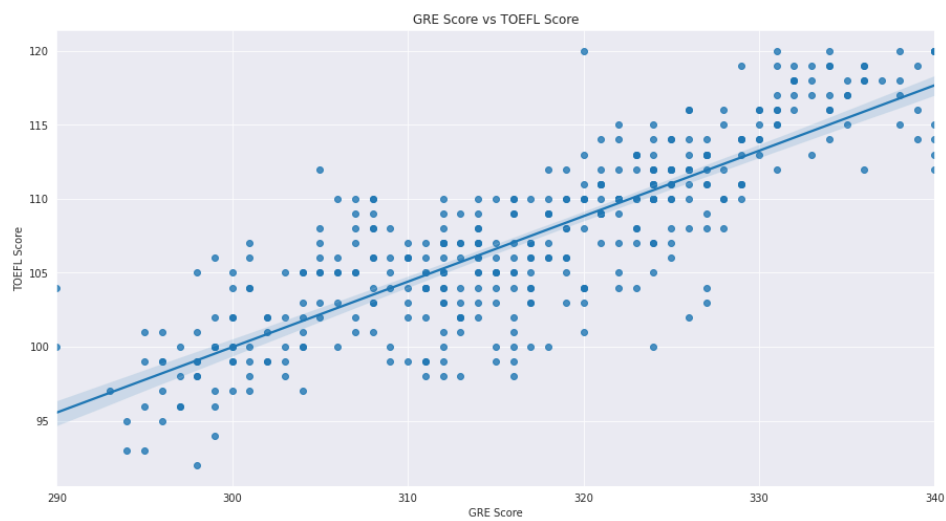
Now checking the distribution of each feature in the dataset.



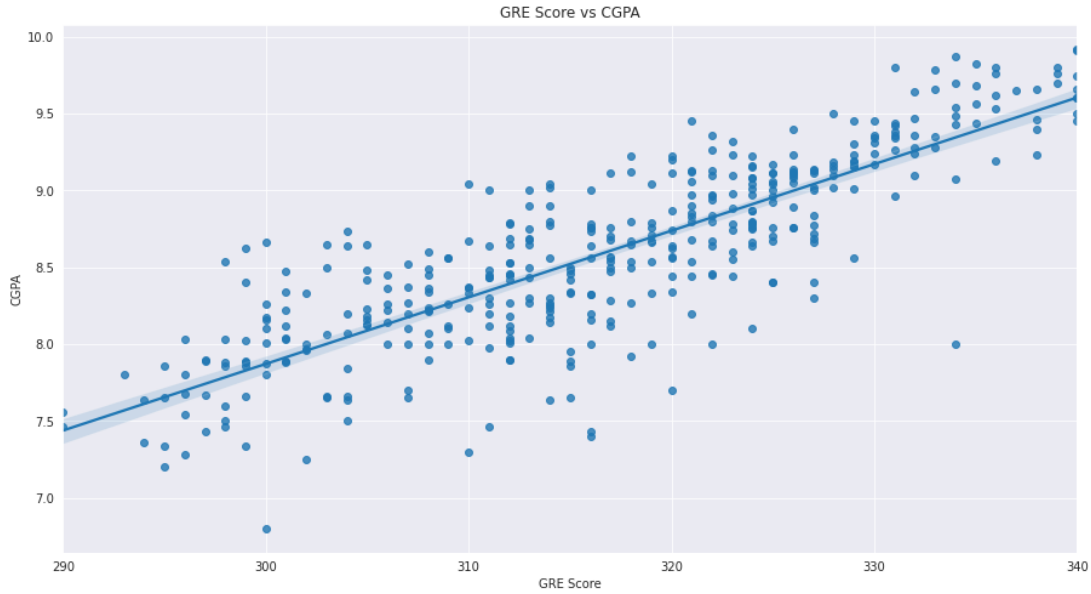


We can see that our data set is pretty balanced regarding the Research feature. All other features somewhat follow the normal distribution, in which the feature has high entries in midrange and lower data points in the low and high range of the distributions

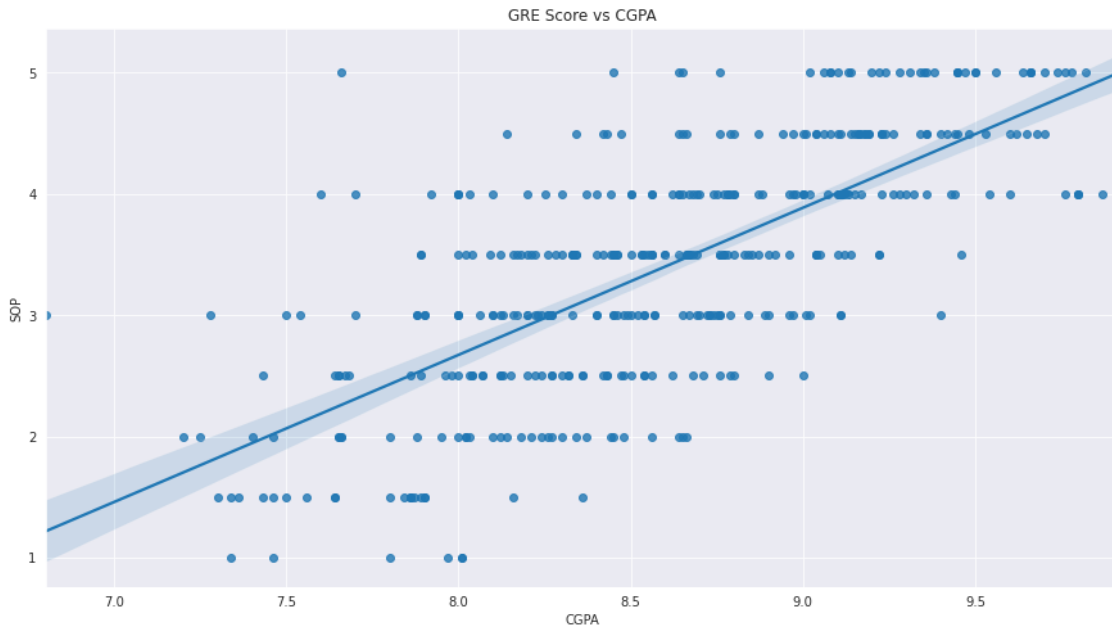
Next, we will try to understand better the relation between different factors responsible for graduate admissions.



People with higher GRE Scores also have higher TOEFL Scores which indicates that students with high intellectual proficiency also possess great language ability



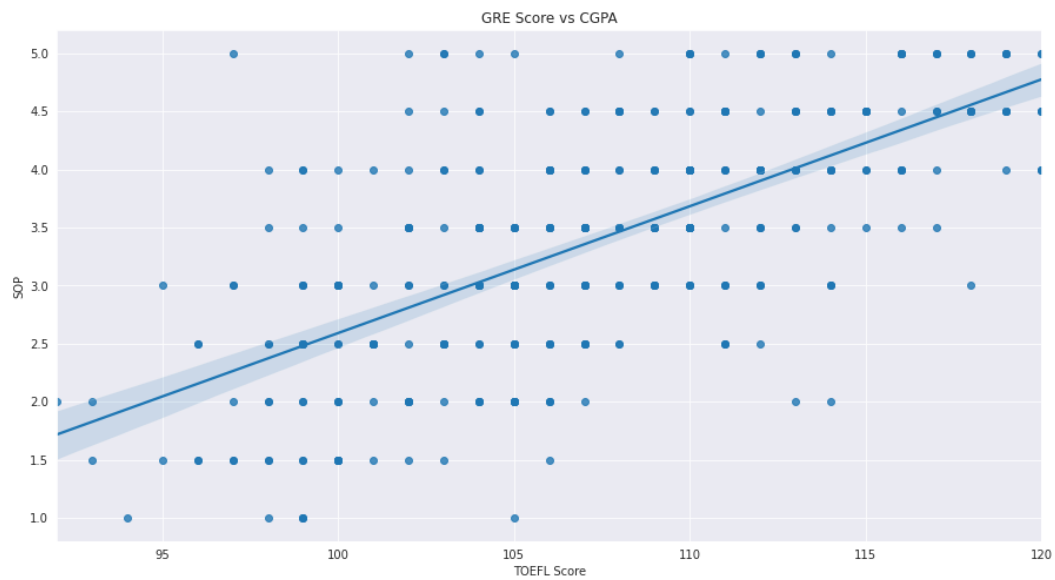
Although there are exceptions, people with higher CGPA usually have higher GRE scores maybe because they are smart or hard working



CGPA and SOP are not that related because Statement of Purpose is related to academic performance, but since people with good CGPA tend to be more hard working so they have good things to say in their SOP which might explain the slight move towards higher CGPA as along with good SOPs.



Similarly, GRE Score and CGPA is only slightly related



Applicants with different kinds of SOP have different kinds of TOEFL Score. So the quality of SOP is not always related to the applicants English skills.

Correlation table among features.

	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

We can see that there are some high couple of features with high correlative relationship of over 0.8 such as:

- TOEFL vs GRE
- CGPA vs GRE
- TOEFL vs CGPA

IV. Results and Discussion

I use two model for the prediction namely Linear Regression and Deep Neural Network Using Keras Library. All the parameters are normalized before training to ensure that values lie between the specified range, and the dataset is split with 20% for the test set.

The Neural Network has the following parameters:

- Three hidden layers, each of 10 nodes
- ReLU activation function
- Adam optimizer function
- Loss functions is Mean Square Error

And the Results

	Linear Regression	Deep Neural Network
Mean Squared Error	0.2443111527031876	0.23688311874866486

As we can see the MSE of the Neural Net Model is 0.236, which is just slightly better than the classic Linear Regression model with MSE of 0.2443

V. Conclusion

As we can see, sometimes just a simple algorithm can solve the problem just as good as a very advance ones, and also save you time and computing resources. In conclusion, understand your data, the business insights together with a practical knowledge of which algorithm of models work for which kind of problems will make you work smart and efficient.