Graduates Admission Prediction using Machine Learning

Jayaganesan Chandrasekaran

Date : 22/Jul/2022

"*You can have data without information, but you cannot have information without data."*

– Daniel Keys Moran

1) Problem Statement

Many Indian students want to do their higher studies abroad. They have to apply for multiple universities in order to get admitted into the best university. But preparing the list of universities to apply for becomes a difficult decision for students. Before applying for a university, if a student gets to know the chances of getting admitted to that university, it will be very useful for the student. He can skip the universities where the chances of getting admitted are low and narrow down the list of universities to apply for. This problem can be solved using a predictive model which predicts the probability of a student gettting admitted to a university.

2) Market/Customer/Business need Assessment

With ever increasing number of students opting for overseas higher education, there is huge amount of data regarding admission/rejection of students applying for various universities. With this data, predictions can be made using Machine Learning/Artificial Intelligence. These predictions can aid a student to take informed decisions when it comes to applying for universities for their higher education.

3) Target Specifications

Though these prediction models are aimed at students who wish to do their overseas higher eduction, they can be useful to eductional consultant firms too. Our models will help them make accurate suggestions to students applying for various universities.

4) External Search

https://unicreds.com/blog/acceptance-rate-of-universities-in-the-usa
https://www.educations.com/articles-and-advice/how-to-apply-for-schools-abroad-12940
https://www.quora.com/What-is-my-chance-of-getting-accepted-at-good-US-universities

The dataset for small scale implementation can be found in kaggle link given below
https://www.kaggle.com/datasets/mukeshmanral/graduates-admission-prediction

5) Benchmarking

There are some acceptance calculators available. Few are referenced below.
https://www.campusreel.org/college-acceptance-calculator
https://yocket.com/blog/yocket-admit-predictor

But these acceptance calculators do not take two important factors in SOP(Statement Of Purpose) and LOR(Letter Of Recommendation)into consideration. Since we account these two factors based on their strength, our model is bound to be more accurate than the existing predictors.

6) Applicable Patents

There are no patents applicable for this, as we want to enable open-source, academic and research community to audit the algorithms and research on the efficacy of the algorithms that we used in the model.

7) Applicable Regulations

Since the personal details like name of the student are not recorded in the data set, all privacy regulation when it comes to the students information is protected.

8) Applicable Constraints

1. Data collection from students/eduction consultant firms. Firms might not always be open to share the data outside, so need to explore some other sources to get the admissions data.
2. Continuous data collection and maintanance.
3. Few implicit variables like strength/rating of SOP(Statement Of Purpose) and LOR(Letter Of Recommendation) might not be directly available. Need to contact some alumini/student forums to ask them to rate a student's SOP and LOR if required.

9) Business Opportunity

Educational consulting firms can very much benifit from the accurate predictions that our models deliver. It can help them guide the students who come to them for consultation.

10) Concept Generation

I came up with this prediction model upon seeing various students find it a challenge to narrow down the list of universities to apply for their higher studies. Only a portion of the students have the luxury to apply for many universities, but the majority of students had financial constraints while applying for the universities. And as I was looking for information/data related to this, I stumbled across the above mentioned dataset from kaggle and found it very useful in helping me create a model to predict the probability that the student gets admitted into a university.

11) Concept Development

After acquiring the appropriate dataset from kaggle, the next thing was to choose a model to make the predictions.

The assumptions for a Linear Regression are as follows:

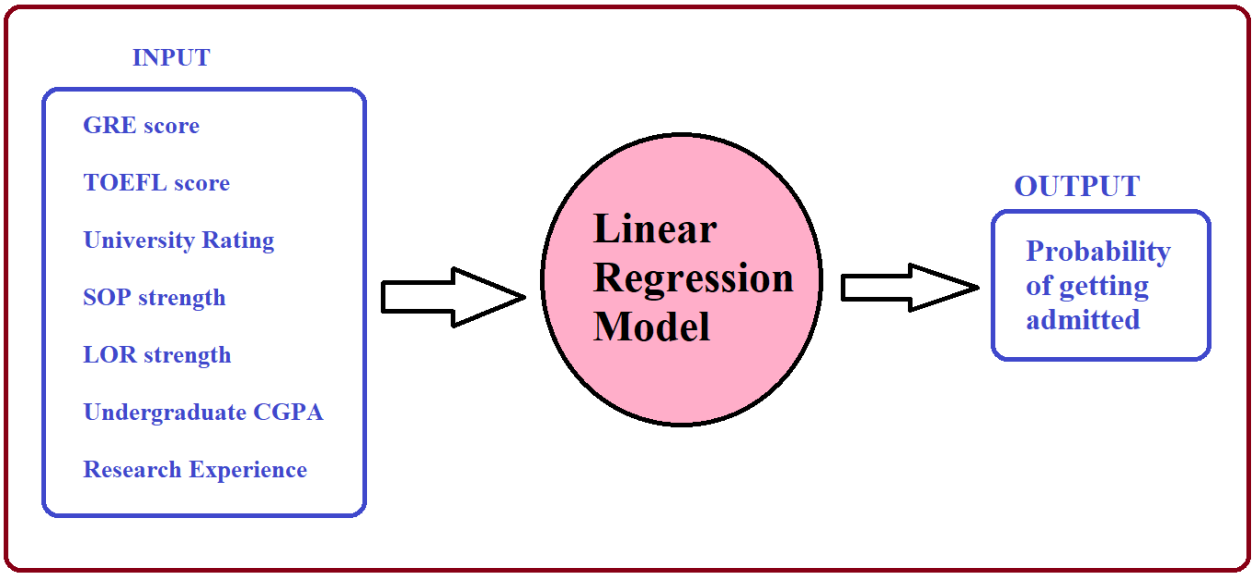
1) Linearity
2) No Homogeneity
3) Homoscedasticity
4) Normality of Error Distribution
5) No Auto-correlation
6) No Multi-Collinearity

The small scale dataset was found to satisfy all these requirements to be able to use Linear Regression model. Hence Linear Regression model was chosen and the accuracy of model was also good.

12) Final Product Prototype

The final product will take the a web application which will take in the inputs and predict the probability of getting admitted and display the same.

The schematic diagram for the Machine Learning model is shown below:



13) Product Details

13.1) How does it work?

The web application will take in these inputs, namely GRE score(out of 340), TOEFL score(out of 120), University Rating(out of 5), SOP strength(out of 5), LOR strength(out of 5), Undergraduate CGPA(out of 10) and Research Experience(0 or 1). And based on the coefficients got from the Linear Regression model, we can calculate the probability of the student getting admitted to the university. Since all the inputs are categorical data in terms of numbers, we can create a drop down box displaying the possible numbers for them to choose and on the click of the submit button the probability of getting admitted can be displayed.

13.2) Data Sources

The data for this project can be acquired directly from students or through some eductational consulting firms on a regular basis. Any other public data sources like kaggle or similar data sources can also be taken.

13.3) Algorithms and Frameworks

- Seaborn is used for visualizing the dataset.
- For predicton model LinearRegression from sklearn is used.
- For making the predictions a html can be used to feed the inputs and get the predicted probability.

13.4) Team

The team required to develop the project will be as follows:

1. Web Developer.
2. Machine Learning Engineer.

14) Code Implementation on a Small Scale

The following provides the github link:
https://github.com/jaiganesh17/Graduates-Admission-Prediction

15) Conclusion

The simple LinearRegression model created is able to accurately predict the probability of a student getting admitted to the univeristy(based on its rating and other inputs like GRE, TOEFL, SOP, LOR, CGPA & Research). These accurate predictions will help the student to decide whether or not to apply to that particular university.

These predictions can be made for other educational careers as well, which will involve collecting appropriate data and may require complex models to come up with accurate predictions.