As per UGC guidelines an electronic bar code is provided to seure your paper

International Journal for Modern Trends in Science and Technology, 6(12): 474-478, 2020
Copyright © 2020 International Journal for Modern Trends in Science and Technology
ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST061292
Available online at: http://www.ijmtst.com/vol6issue12.html

Graduate University Admission Predictor using Machine Learning

Aashish Singhal[1] | Saurabh Gautam[2]

[1]B-tech. I.T., Maharaja Agrasen Institute of Technology, GGSIPU, New Delhi, India
[2]2Assistant Professor I.T., Maharaja Agrasen Institute of Technology, GGSIPU, New Delhi, India

To Cite this Article
Aashish Singhal and Saurabh Gautam, "Graduate University Admission Predictor using Machine Learning", *International Journal for Modern Trends in Science and Technology*, 6(12): 474-478, 2020.

Article Info
Received on 16-November-2020, Revised on 09-December-2020, Accepted on 12-December-2020, Published on 21-December-2020.

ABSTRACT

With the increase in the number of graduates who wish to pursue their education, it has become more challenging to get admission for the students in their dream university. Usually, newly graduate students are not knowledgeable of the requirements and the procedures of the postgraduate admission and might spent a considerable amount of money to get advice from consultancy organisations to help them identify their admission chances. Giving the limited number of universities that can be considered by a human consultant, however, this approach might be bias and inaccurate. Higher education in abroad universities generally means we have many options like Canada, USA, UK Germany, Italy, Australia etc. But we are focusing on only the students who want to do their Masters in America. Students who want to do masters in America have to write GRE (Graduate Records Examination) and TOEFL (Test of English as a Foreign Language). Once they have attended the exams they have to prepare their SOP (statement of purpose) and LOR(letter of recommendation) which are one of the crucial factors they have to consider. These LOR and SOP plays a vital role if the student was looking for any scholarship. 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. So, with increasing demand of further education, one must not be confused in where to apply. Then the students have to choose the universities they want to study or apply, we cannot apply to all the universities that will lead to lot of application fees. Here comes the problem that the student doesn't know to which university he might get admission. There are some online blogs which help in these matters but they are not that much accurate and don't consider all the factors and there are some consultancy offices which will take lot of our money and time and sometimes they will give some false information.so our goal is to develop a model which will tell the students their chance of admission into a respective university. This model should consider all the crucial factors which plays a vital role in student admission process and should have high accuracy.

KEYWORDS: *College admission predictor; Machine Learning*

I. Introduction

Preparing for specific things plays a crucial part in your life. Thus education preparation students often have multiple questions about universities which they can get admission and scholarship and accommodation. One of the main concerns is getting admitted to their dream university. It's seen that students still choose to obtain their education from universities that are known internationally. And when it comes to international graduates, the United States of America is the first preference of the majority of them. With most world-renowned colleges, Wide variety of courses available in each discipline, highly accredited education and teaching programs, student scholarships, are available for international students. According to estimates, there are more than 10 million international students enrolled in over 4200 universities and colleges including both private and public across the United States. Most number of students studying in America are from Asian countries like India, Pakistan, Sri Lanka, Japan and China. They are choosing not only America but also UK, Germany, Italy, Australia and Canada. The number of people pursuing higher studies in these countries are rapidly increasing. The background reason for the students going to abroad universities for Masters is the no. of job opportunities present are low and number of people for those jobs are very high in their respective countries. This inspires many students in their profession to pursue postgraduate studies. It is seen that there is quite a large number of students from universities in the USA pursuing Masters in the field of computer science, the emphasis of this research will be on these students. Many colleges in the U.S. follow similar requirements for student admission. Colleges take different factors into account, such as the ranking on aptitude assessment and academic record review. The command over the English language is calculated on the basis of their performance in the English skills test, such as TOEFL. The admission committee of universities takes the decision to approve or reject a specific candidate on the basis of the overall profile of the applicant application.

II. LITERATURE SURVEY

This section includes the literature review of previous research on the assessment of student enrolment opportunities in universities. Numerous programs and studies are administered on topics concerning university admission used many machine learning models which helps the scholars within the admission process to their desired universities.

The main drawback of the previous research done on this is they didn’t consider all the factors which will contribute in the student admission process like TOEFL, SOP, LOR and under graduate score. Bayesian Networks Algorithm have been used to create a decision support network for evaluating the application submitted by foreign students of the university. This model was developed to forecast the progress of prospective students by comparing the score of students currently studying at university. The model thus predicted whether the aspiring student should be admitted to university on the basis of various scores of students. Since the comparisons are made only with students who got admission into the universities but not with students who got their admission rejected so this method will not be that much accurate.

III. METHODLOGY

Problem Understanding: Initially first we have to spend some time on what are the problems or concerns students having during their pre admission period and we should set the solutions to those problems as objectives of this research. **Data Understanding:** Data required for the research was collected from multiple data sources. Different features of the data were analyzed based on their importance and relevance.**Data Preparation:** In this phase, the data from multiple data sources were integrated into a final data-set. Further the data was cleaned by removing unwanted columns, performing transformation and cleaning activities on the data. **Building Models:** Multiple machine learning models were developed to predict the likelihood of success of the student’s application in a particular university. **Evaluation:** Models developed were evaluated based on their performance and accuracy. More information will be presented in the evaluation section of the paper.

Data Cleaning and Analysis:

Inspecting feature values that help identify what needs to be done to clean or pre-process until you see the range or distribution of values typical of each attribute. We may find missing or incomplete data such as the incorrect data form used for a column, incorrect measuring units for a particular column, or that there are not enough examples of a specific class.

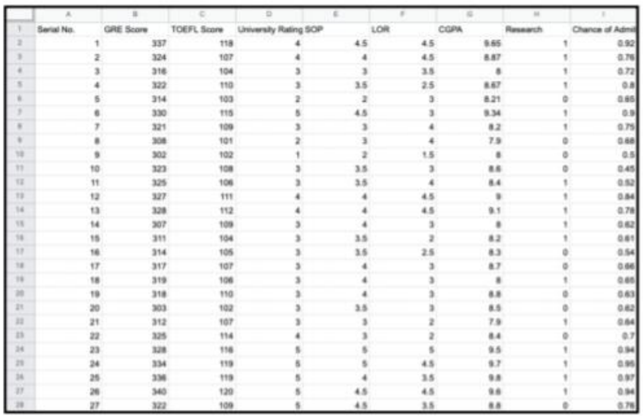
This process of data cleaning has several key benefits to it:

1. This eliminates major errors and inconsistencies which are unavoidable when dragging multiple data sources into one dataset.

2. Having data cleaning software will make everyone more effective as they will be able to get easily from the data what they need.

3. Fewer mistakes mean happy clients, and less unhappy workers.

4. The ability to chart the various functions, and what your data is supposed to do and where it comes from your data.



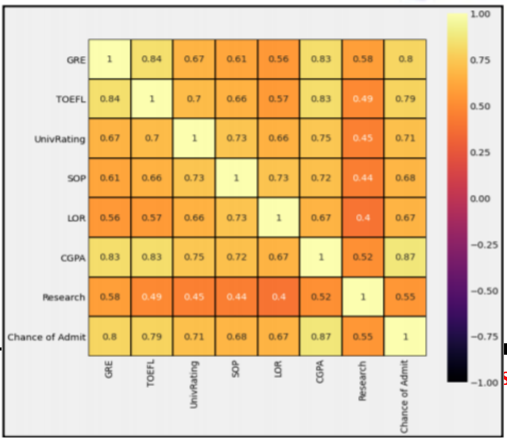
	A	B	C	D	E	F	G	H	I
1	Serial No.	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
2	1	337	118	4	4.5	4.5	9.65	1	0.92
3	2	324	107	4	4	4.5	8.87	1	0.76
4	3	316	104	3	3	3.5	8	1	0.72
5	4	322	110	3	3.5	2.5	8.67	1	0.8
6	5	314	103	2	2	3	8.21	0	0.65
7	6	330	115	5	4.5	3	9.34	1	0.9
8	7	321	109	3	3	4	8.2	1	0.75
9	8	308	101	2	3	4	7.9	0	0.68
10	9	302	102	1	2	1.5	8	0	0.5
11	10	323	108	3	3.5	3	8.6	0	0.45
12	11	325	106	3	3.5	4	8.4	1	0.52
13	12	327	111	4	4	4.5	9	1	0.84
14	13	328	112	4	4	4.5	9.1	1	0.78
15	14	307	109	3	4	3	8	1	0.62
16	15	311	104	3	3.5	2	8.2	1	0.61
17	16	314	105	3	3.5	2.5	8.3	0	0.54
18	17	317	107	3	4	3	8.7	0	0.66
19	18	319	106	3	4	3	8	1	0.65
20	19	318	110	3	4	3	8.8	0	0.63
21	20	303	102	3	3.5	3	8.5	0	0.62
22	21	312	107	3	3	2	7.9	1	0.64
23	22	325	114	4	3	2	8.4	0	0.7
24	23	328	116	5	5	5	9.5	1	0.94
25	24	334	119	5	5	4.5	9.7	1	0.95
26	25	336	119	5	4	3.5	9.8	1	0.97
27	26	340	120	5	4.5	4.5	9.6	1	0.94
28	27	322	109	5	4.5	3.5	8.8	0	0.76

Fig 1. Admission Prediction CSV Data

· There are no missing values and outliers because we analyzed the data, so for this data there is no need to fill the missing values and deal with outliers. If there are any missing values and outliers we can fill (or) drop using the fillna method and drop method and we can also standardize the data using the min-max scaler, if necessary.

Data Visualization:

· After analyzing the data, we will be able to know what the features and labels are, so from the above data, the label we have to consider is Chance of Admission and then we have to consider the parameters that influence or play a major role in Chance of Admission · We can get to know certain features that are more affected by the visualization (or) analysis or the use of feature importance method in decision tree.Heat map below shows that TOEFL score, GRE score, University Rating and Research are most important .



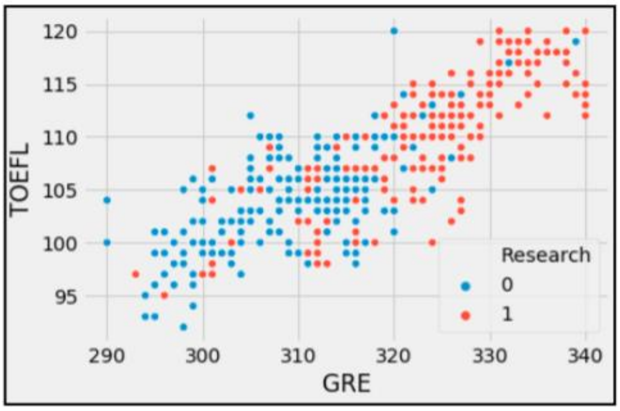


Figure 3. TOFEL Vs GRE and Research

The scatter plot above shows that students with high TOEFL and GRE score have done research.

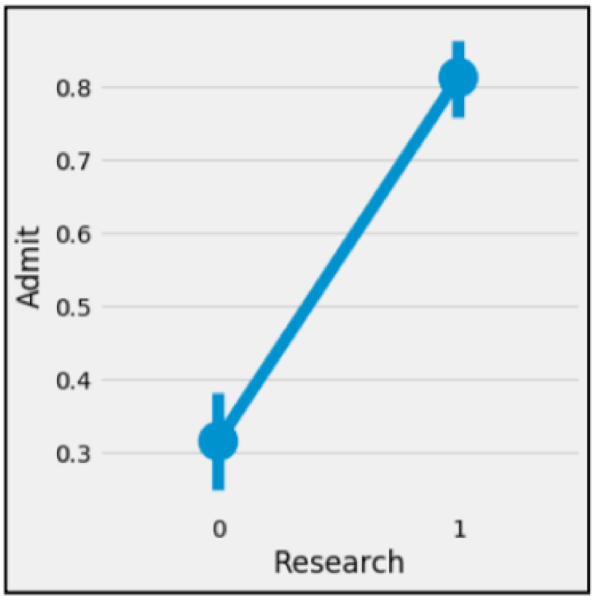


Fig 4. Research Done Vs Admit Chance

The line graph above shows that students who have done research have high chances of admission.

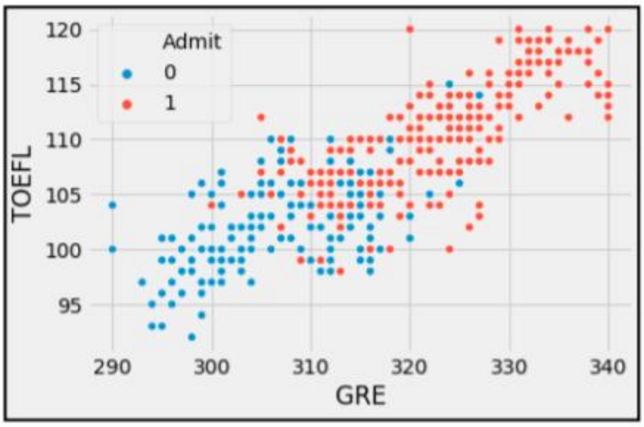
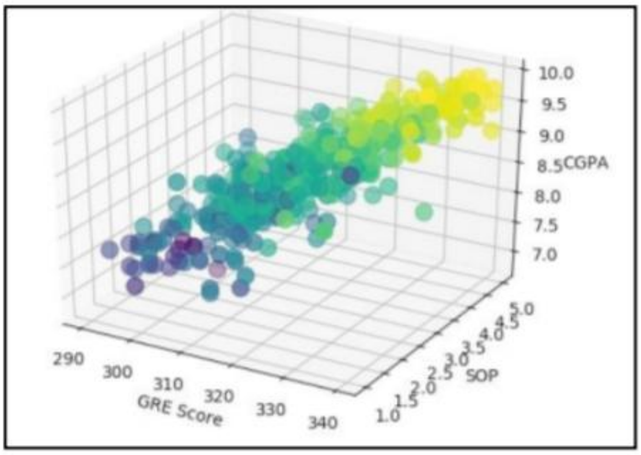


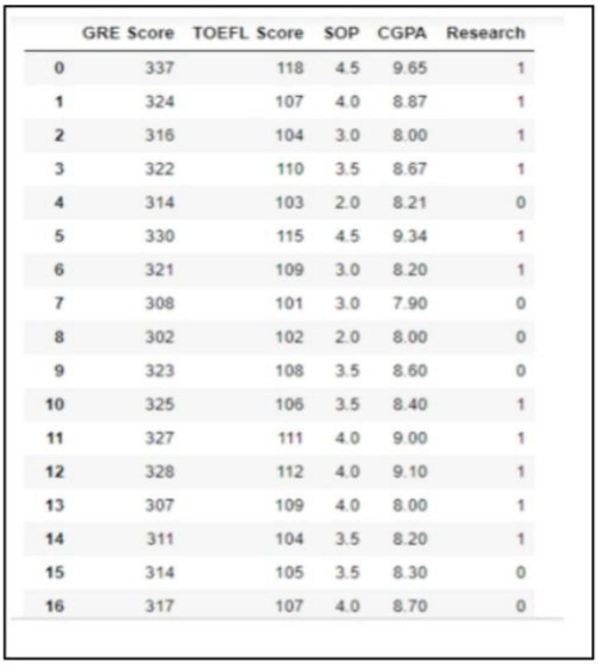
Fig 5. TOEFL Vs GRE and Chance of Admit

The scatter plot above shows that students with high TOEFL and GRE score are very likely to admit.

Fig 6. 3D Graph



Results of the visualizations and data analysis above shows that the features of the data set have a very high impact on the probability of admission, so only features are considered.



	GRE Score	TOEFL Score	SOP	CGPA	Research
0	337	118	4.5	9.65	1
1	324	107	4.0	8.87	1
2	316	104	3.0	8.00	1
3	322	110	3.5	8.67	1
4	314	103	2.0	8.21	0
5	330	115	4.5	9.34	1
6	321	109	3.0	8.20	1
7	308	101	3.0	7.90	0
8	302	102	2.0	8.00	0
9	323	108	3.5	8.60	0
10	325	106	3.5	8.40	1
11	327	111	4.0	9.00	1
12	328	112	4.0	9.10	1
13	307	109	4.0	8.00	1
14	311	104	3.5	8.20	1
15	314	105	3.5	8.30	0
16	317	107	4.0	8.70	0

Fig 7. Final Data

- Once we are done with data visualisation, predictive modelling is done. For this,first we divide the data into training set and test set.
- We will develop models using machine learning algorithms on the training data and test model accuracy on the test data part.
- We will see which algorithms giving highest accuracy according to what parameters and take that for final consideration.

IV.ALGORITHMS

For this work, several machine learning algorithms have been used like Linear Regression, K- Nearest Neighbour, Decision Tree, Random Forest are used to predict students likelihood of university admission based on their profile.

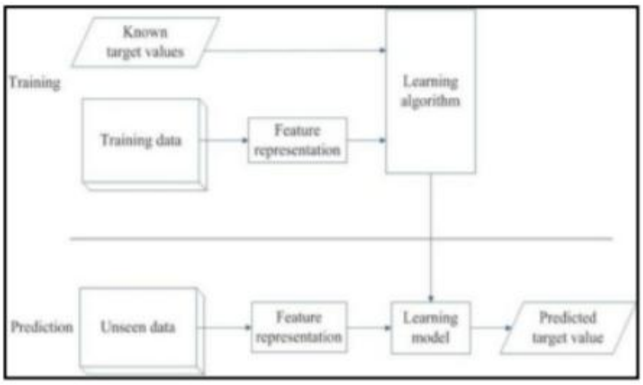
Linear Regression: It is an algorithm based on supervised learning of computers. It does the role of regression. It models a predictive goal value based on the independent variables. Mostly it is used to figure out the relation between variables and forecasting. Different regression models vary on the basis–the form of relationship between dependent and independent variables, are considered, and the number of independent variables used.

K-Nearest Neighbours: KNN algorithm is the most commonly used algorithm for classification and regression purpose. KNN stands for k nearest neighbour, here k indicates an integer value which will tell that with how many neighbours comparisons should be made. This can be used for both classification and regression purpose. Suppose if it is classification and the k value is 5 it will compare with nearest 5 neighbours and gives the mode value, if it is regression and the k value is 6 it will take the nearest six values and return its mean value.

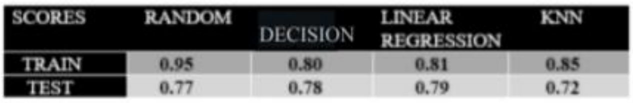
Decision Tree: It is a supervised machine learning algorithm. Due to its simple logic, effectiveness and interpretability it the most widely used classification algorithm. The model works by creating a tree-like structure by dividing the dataset into several smaller subsets based on different conditional logic.

Random Forest: This is a machine learning algorithm which has a combined effect of classification and regression and other tasks which operate by erection of decision trees at training time and outputs the class that is the mode of the classes or mean value of individual trees.

V. FLOWCHART



VI. RESULT ANALYSIS



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

Table: Comparison between Models

VII. CONCLUSION

The main goal of this work is to create a Machine Learning model which could be used by students who want to pursue their education in the US. Many machine learning algorithms were utilised for this research. Linear Regression model compared to other ones. Students can use the model to assess their chances of getting admission into a particular university with an average accuracy of 79 percent. The ultimate goal of research will be accomplished successfully, 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.

REFERENCES

1. C. Haythorhwaithe, M. de Laat, and S. Dawson, Introduction to the special issue on the learning analytics. American Behavioral Science,57(10):1371-1379,2013.
2. Liu Jinpeng. Research on the application of Data Mining Technology in Analysis of Examinee Wish, Henan University,2009.
3. Alpaydin, E. Introduction to Machine Learning,3rd;MIT press:Cambridge,MIT, USA,2010.
4. Kuncheva, LL combining pattern classifiers: Methods and wiley&sons,Inc:Hoboken,NJ,USA,2014.
5. D.M Blei, A.Y. Ng, and M. I. Jordan, Latent Dirichletallocation,Journal of Machine Learning Research,3:993-1022, 2003.
6. L. Breiman, Accuracy Predictors, Machine Learning, 24(2):123-140,1996.
7. Data Cleaning and Analytics, Machine Learninghttps://archieve.ics.uci.edu/ml/index.php
8. Data visualization and Machine Learning https://www.analyticsvidhya.com/blog/2017/09/common-machin e-learning-algorithms/
9. Jupyter Notebook, Implementing the Algorithms, MachineLearning,https://jupyter-notebook.readthedocs.io/en /stable/