

Graduate Admission Prediction Using Linear Regression

Ms. Moram lakshmi rekha¹, Mr.M Prasad², Mrs TSN Chandreswari³, Mr.B Durga Prasad⁴

Ms Moram Lakshmi Rekha, Asst.Prof,
Department Of MCA,
BVC Institute of Technology and Science,
Amalapuram, E.G.Dt.,AP
Email:rekhamoram@gmail.com

Mr M Prasad, Prof,
Department Of CSE,
BVC Institute of Technology and Science,
Amalapuram, E.G.Dt.,AP
Email:Prasads.Maddula@gmail.com

Mrs TSN Chandrewari, Asst.Prof,
Department Of MCA,
BVC Institute of Technology and Science,
Amalapuram, E.G.Dt.,AP
Email:chanduprsadvinnu@gmail.com

Mr B Durga Prasad, Prof,
Department Of CSE,
BVC Institute of Technology and Science,
Amalapuram, E.G.Dt.,AP
Email:prasaddurga5802@gmail.com

Abstract

A person's 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. Securing admission in their dream university is one of their main concerns. Generally as the students don't have much idea about the procedures, requirements and details of the universities, they seek help from the education consultancy firms to help them successfully secure admission in the universities which are best suitable for their profile, for this they have to invest huge amount of money as consultancy fees. The aim of this paper is to develop a system using machine learning algorithms because machine learning contains statistical functions, it gives highest accurate results. It will help the students to identify the chances of their application to an university being accepted. Also it will help them in identifying the universities which are best suitable for their profile and also provide them with the details of those universities. In this, Linear Regression, Random Forest Regression and Support Vector Regression algorithms are used. Of all these algorithms, the Linear Regression algorithm gives high accuracy according to the dataset.

Key Words: prediction, Linear Regression, Vector Regression, statistical, dataset,

1. INTRODUCTION

In today's era we see a lot of students pursuing their education away from their home countries. For this, students search for the best universities which they want to admit. Then Students need some help to find the best university and they take help from the some of the educational consultancy firms. Generally

the consultancy firms tell about the universities what they know. Some of the consultancy firms can give correct opinions and some of the firms can give wrong opinions for the sake of money. Based on their opinions Students decided to join one of the university. Some of the firms spoil the student's life because of their incorrect opinions. To recover all these problems machine learning algorithms can be used to give accurate results.

The Graduate admission prediction system can help the Students to find which the best university to join is. This system can give accurate results too. The system can take input from users such as ranks of entrance examinations conducted by different universities, marks of previous degree and rating of the particular university. Based on this information machine learning algorithms such as regression algorithms can help to find which one is best. Regression analysis is a set of statistical methods used for the estimation of relationships between a dependent variable and one or more independent variables. It can be utilized to assess the strength of the relationship between variables and for modeling the future relationship between them.

In order to understand regression analysis fully, it's very important to know the following terms: Dependent Variable: This is the main factor that trying to understand or predict. Independent Variables: These are the factors that you hypothesize have an impact on your dependent variable. Simple linear regression is a model that assesses the relationship between a dependent variable and one independent variable.

The simple linear model is expressed using the following equation:

$$Y = \theta_1 + \theta_2 \cdot X$$

While training the model we are given:
x: input training data (univariate – one input variable(parameter))

y: labels to data (supervised learning)

When training the model – it fits the best line to predict the value of y for a given value of x. The model gets the best regression fit line by finding the best θ_1 and θ_2 values.

θ_1 : intercept

θ_2 : coefficient of x

Once we find the best θ_1 and θ_2 values, we get the best fit line. So when we are finally using our model for prediction, it will predict the value of y for the input value of x.

Update θ_1 and θ_2 values to get the best fit line:

CostFunction(J):

By achieving the best-fit regression line, the model aims to predict y value such that the error difference between predicted value and true value is minimum. So, it is very important to update the θ_1 and θ_2 values, to reach the best value that minimize the error between predicted y value (pred) and true y value (y).

$$\text{minimize } \frac{1}{n} \sum_{i=1}^n (\text{pred}_i - y_i)^2$$

$$J = \frac{1}{n} \sum_{i=1}^n (\text{pred}_i - y_i)^2$$

Cost function (J) of Linear Regression is the **Root Mean Squared Error (RMSE)** between predicted y value (pred) and true y value (y).

1.1 RELATED WORK:

Some education agents represent only a few institutions and will therefore not show you the full range of institutions on offer. And some give university names for the sake of money in that time the students may suffer. Every candidate has to take all the required examination and build a strong profile to secure admission in their university.

Once the candidates have made their profile ready, they apply to the universities where they aim to secure admission. The students have to shortlist the

universities which are best known for the courses they are looking for and also they should have an idea about their chances of securing admission in those universities based on their profile.

This task of short listing the universities where the student has high chances of admission is difficult, so they end up with applying to many universities in hopes of getting admission in few of them thus investing an extra amount of money in the applications. There are several portals and websites which provide information and help to students in short listing the universities, but they are not reliable. Most of the students don't take the risk of evaluating the colleges by themselves, and they seek the help of the education consultancy firms to do it for them. Again for this students have to pay a huge amount of fee to the education consultant.

The primary objective of this paper is to develop a system to solve the problems the international students are facing while applying for universities. We will be developing a Graduate Admission Predictor (GAP) system which will help the students to predict the chances of their application being selected for a particular university for which they wish to apply based on their profile. Also, the system will provide a recommendation of universities to the student to which the student has a high possibility of getting admission.

Multiple machine learning Regression 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 the predicted result for the application based on the profile as output.

This system will thus eventually help students saving the extra amount of time and money they have to spend at the education consultancy firms. And also it will help them to limit their number of application to a small number by proving them the suggestion of the universities where they have the best chance of securing admission thus saving more money on the application fees.

1.2 PREVIOUS WORK:

In the past days students take decisions for the future graduation, with help of educational agents or educational consultancies. They explain some of the details of universities which are matched to the

student's profile. Based on their feedback of the universities students will take decisions which university they want to join. In that case students blindly believe the educational consultancies; sometimes students choose wrong universities with help of educational consultancies. In that case student's life will spoil, if the student doesn't satisfy by that university.

1] The educational consultancies may be biased because institutions pay education agents a commission once the student has enrolled.

2] They may not be showing you all the options, some education agents represent only a few institutions and will therefore not show you the full range of institutions on offer.

3] Having a limited choice may lead you to choosing an institution that doesn't quite meet your needs and expectations.

4] It is a time-taking process to follow the educational consultancies and the cost of educational consultancies fee is also very high.

2. Proposed functioning:

To overcome these problems in the existing system, develop a system which is used to predict the chances of admission in their dream university. The Graduate admission prediction system can help the Students to find which the best university to join is. This system can give accurate results too. The system can take input from users such as ranks of entrance examinations conducted by different universities, marks of previous degree and rating of the particular university. Based on these details of students this will give the chance of admission as the output.

This chance of admission will give the percentage, which will obtain prediction to join on that university. Machine Learning algorithms such as Regression algorithms are used to predict the chances of admission to join the particular university. By using these algorithms this system will give accurate results.

1] The payment which we pay to the educational consultancies is decreased by using this system.

2] The Graduate Admission Prediction system will give which university is suitable for the student's profile. Based upon the output students will choose the best university, because it will give accurate results.

2.1 System Architecture:

First Extract the dataset and then convert categorical data into numerical values then refining the data of features containing null values and then feature scaling. Now split the dataset into Training Data and Test Data which can be used to Train the System. Next step is selecting the model based on refined dataset for predicting the best results.

Train the machine using selected model then predict the output for new input values as shown below

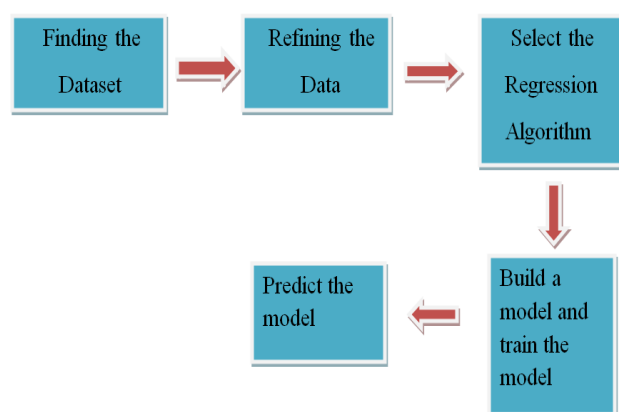


Fig: System architecture

Viewing of the dataset:

This is the first module functionality of Admin module in this system. Admin already upload the dataset which is used for prediction process. When Admin want to view the already uploaded dataset the Admin may login into page. Then the Admin can view the uploaded dataset. Admin is only one person who can modify the dataset for different operations performed in that dataset for prediction.

Count the users:

This is the second functionality of Admin module in this system. It will display how many users are used this system. For every user the count will be incremented by one. Admin can view the dataset as well as the count of user. How many users can use this system will be maintained by this count the user.

Prediction of Admission:

The second module in this system is Prediction of Admission. User had an ability to perform operations such as preprocessing techniques, feature extraction, prediction on the dataset which is uploaded by admin.

In this the user can view the predicted output after performing the algorithms on the dataset. It will display the chance of admission to join in the particular university.

Table -1: Sample user input Data

NAME	<input type="text" value="priya"/>
CGPA	<input type="text" value="9.65"/>
GRE	<input type="text" value="337"/>
TOEFL	<input type="text" value="118"/>
RATING	<input type="text" value="4"/>

THE NUMBER OF USERS VISITED ARE :: 4
1 priya 2 anusha 3 anusha 4 bhavani

Serial No.	gre score	toefl	rating	sop	lor	cgpa	research	C O	Admit
1	337	118	4	4.5	4.5	9.65	1	1	0.92
2	324	107	4	4.0	4.5	8.87	1	1	0.76
3	316	104	3	3.0	3.5	8.00	1	1	0.72
4	322	110	3	3.5	2.5	8.67	1	1	0.80
5	314	103	2	2.0	3.0	8.21	0	0	0.65
6	330	115	5	4.5	3.0	9.34	1	1	0.90
7	321	109	3	3.0	4.0	8.20	1	1	0.75
8	308	101	2	3.0	4.0	7.90	0	0	0.68
9	302	102	1	2.0	1.5	8.00	0	0	0.50
10	323	108	3	3.5	3.0	8.60	0	0	0.45
11	325	106	3	3.5	4.0	8.40	1	1	0.52
12	327	111	4	4.0	4.5	9.00	1	1	0.84
13	328	112	4	4.0	4.5	9.10	1	1	0.78
14	307	109	3	4.0	3.0	8.00	1	1	0.62
15	311	104	3	3.5	2.0	8.20	1	1	0.61

Table-2: view Dataset

THE NUMBER OF USERS VISITED ARE :: 6
1 priya 2 anusha 3 anusha 4 bhavani 5 priya 6 bunny

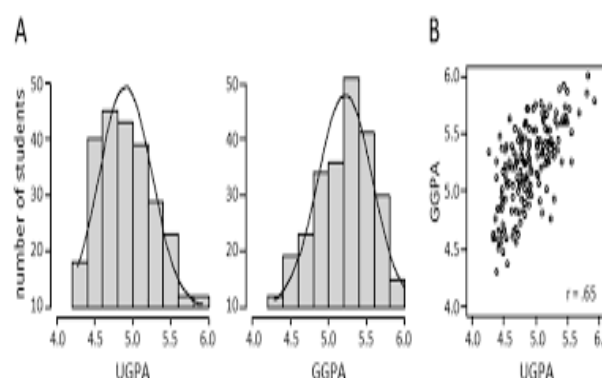
Serial No.	gre score	toefl	rating	sop	lor	cgpa	research	C O	Admit
1	337	118	4	4.5	4.5	9.65	1	1	0.92
2	324	107	4	4.0	4.5	8.87	1	1	0.76
3	316	104	3	3.0	3.5	8.00	1	1	0.72
4	322	110	3	3.5	2.5	8.67	1	1	0.80
5	314	103	2	2.0	3.0	8.21	0	0	0.65
6	330	115	5	4.5	3.0	9.34	1	1	0.90
7	321	109	3	3.0	4.0	8.20	1	1	0.75
8	308	101	2	3.0	4.0	7.90	0	0	0.68
9	302	102	1	2.0	1.5	8.00	0	0	0.50
10	323	108	3	3.5	3.0	8.60	0	0	0.45
11	325	106	3	3.5	4.0	8.40	1	1	0.52
12	327	111	4	4.0	4.5	9.00	1	1	0.84
13	328	112	4	4.0	4.5	9.10	1	1	0.78
14	307	109	3	4.0	3.0	8.00	1	1	0.62
15	311	104	3	3.5	2.0	8.20	1	1	0.61

Table-3: predicted Output

Classification Performance:

The admissions committee admitted 92 (15.6%) and rejected 546 (84.4%). Graduate predicted the correct admissions decision with 87.1% accuracy. Note that this result is not very meaningful: due to the imbalance of the classes in the test set, one could achieve 84.4% accuracy must by rejecting all applicants. Rather than looking at accuracy, the system is best understood by its precision-recall characteristics. The precision-recall curve for

Graduate's 2013 predictions. In terms of graduate admissions, precision and recall measure the system's ability to identify applicants who will be admitted. As the top left corner of the shows, Graduate is very good at finding a small number of high-quality applicants that the committee will admit. However, precision quickly drops at larger recalls. This reflects that there are many mid-scoring applicants for which the committee's decision is difficult to predict. In practice, Graduate is much better at identifying applicants that the committee will reject. The classifier's true negative rate (specificity) versus false negative rate; these are the fractions of applicants that the system would correctly and incorrectly reject at varying decision thresholds.



3. CONCLUSIONS

The main objective was to develop a system that can be used by the students aspiring to pursue their education. Multiple machine learning algorithms were developed and used for this system. Linear Regression proved to be the best fit for development of the system. The model can be used by the students for evaluating their chances of getting shortlisted in a particular university. A simple user interface was developed to make the application interactive and easy to use for the users from the non-technical background. The overall objective of this system was achieved successfully as the system allows 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 decisions regarding application to the universities.

REFERENCES

- 1] Bruggink, T. H., and Gambhir, V. 1996. Statistical models for college admission and enrollment: A case study for a selective liberal arts college. *Research in Higher Education* 37(2):221–240.
- 2]Deerwester, S.; Dumais, S.; Furnas, G.; Landauer, T.; and Harshman, R. 1990. Indexing by latent semantic analysis. *Journal of the American Society for Information Science* 41(6):391–407.
- 3]Moore, J. S. 1998. An expert system approach to graduate school admission decisions and academic performance prediction. *Omega* 26(5):659–670. Saltelli, A.; Chan, K.; Scott, E.; et al. 2000. *Sensitivity Analysis*, volume 134. Wiley New York.
- 4]Tan, M.; Wang, L.; and Tsang, I. 2010. Learning sparse svm for feature selection on very high dimensional datasets. In *International Conference on Machine Learning*.
- 5]Van Rijsbergen, C.; Robertson, S.; and Porter, M. 1980. *New Models in Probabilistic Information Retrieval*. British Library Research and Development Department.
- 6]AGBONLAHO, R. O., AND OFFOR, U. J. 2008. Predicting success in a Master of Information Science degree programme. *Education for Information* 26, 3-4, 169-190. AKAIKE, H. 1974. A new look at the statistical model identification. *IEEE Transactions*