

# YUVA VARDHAN - College Admission Predictor

<sup>1</sup>Parnnika Bhavsar and <sup>2</sup>Aryan Patil

Dept. Computer Engineering, Guru Gobind Singh Polytechnic Nashik, India

[bhavsarparneeka@gmail.com](mailto:bhavsarparneeka@gmail.com), [aryandpatil1704@gmail.com](mailto:aryandpatil1704@gmail.com)

**Abstract**—This project Engineering Admission Predictor System is web based application in which students can register with their personal as well as marks details for prediction the admission in colleges and the administrator can allot the seats for the students. Administrator can add the college details and he batch details. Using this software, the entrance seat allotment became easier and can be implemented using system. The main advantage of the project is the computerization of the entrance seat allotment process. Administrator has the power for the allotment. He can add the allotted seats into a file and the details are saved into the system. The total time for the entrance allotment became lesser and the allotment process became faster.

**Keywords:** conferences, accreditation, admission system, prediction system, convergence

## I. INTRODUCTION

For anyone pursuing their postgraduate studies, it would be difficult for them to find out what college they may join, based on their GPA, Quants, Verbal, TOEFL and AWA Scores. People may apply to many universities that look for candidates with a higher score set, instead of applying to universities at which they have a chance of getting into. This would be detrimental to their future. It is very important that a candidate should apply to colleges that he/she has a good chance of getting into, instead of applying to colleges that they may never get into. There aren't many efficient ways to find out the colleges that one can get into, relatively quickly. The Education Based Prediction System helps a person decide what colleges they can apply to with their scores. The dataset that is used for processing consists of the following parameters: University name, Quants and Verbal Scores (GRE) TOEFL and AWA Scores. The GRE Test (Graduate Record Examinations) is a standardized test used by many universities and graduate schools around the world as part of the graduate admissions process. Other factors are also taken into consideration while applying to colleges, such as Letter of Recommendation (a formal document that application that tells admissions committees who you are, what your academic and professional interests are, and how you'll add value to the graduate program you're applying to), Co-curricular activities and Research papers as well (research papers from journals that are not well known or have a high percentage of plagiarism are not taken into consideration for this case). When a person has completed their undergraduate degree and wants to pursue a Postgraduate degree in a field of their choice, more often than not, it is very confusing for the person to figure out what colleges they should apply to with the scores that they have obtained in GRE and TOEFL, along with their GPA at the time of their graduation. Many candidates may apply to colleges that do not fall under their score requirements and hence waste a lot of time. Applying to many colleges with scores also increases the cost. There are not many efficient methods that are available to help address this issue and hence an Education Predictor System

has been developed. In the system proposed, a person can enter their scores in the respective fields provided. The system then processes the data entered and produces an output of the list of colleges that a person could get into, with their scores. This is relatively quick and helps conserve time and money. In order to achieve this we have proposed a novel method utilizing Machine Learning algorithms. To maximize the accuracy of our model, we have taken into consideration not one; but several machine learning algorithms. These algorithms include Neural Networks, Linear Regression, Decision Tree and Random Forest. More about these algorithms will be covered in the Algorithms section of this paper. These Algorithms are then compared and the algorithm which has the best key performance indicators will be used to develop the Prediction System. We also look forward to incorporate clustering of universities based on a profile and then classifying them as less likely, highly likely acceptance etc.

## II. PROBLEM STATEMENT

Educational organizations have always played an important and vital role in society for development and growth of any individual. There are different college prediction apps and websites being maintained contemporarily, but using them is tedious to some extent, due to the lack of articulate information regarding colleges, and the time consumed in searching the best deserving college. The problem statement, hence being tackled, is to design a college prediction/prediction system and to provide a probabilistic insight into college administration for overall rating, cut-offs of the colleges, admission intake and preferences of students. Also, it helps students avoid spending time and money on counsellor and stressful research related to finding a suitable college.

It has always been a troublesome process for students in finding the perfect university and course for their further studies. At times they do not know which stream they want to get into, but it is not easy for them to find colleges based on their academic marks and other performances. We aim to develop and provide a place which would give a probabilistic output as to how likely it is to get into a university given upon their paper.

## III. DATASET

The data set comprises of different factors attributed towards picking the right university. It contains data of 100 different students. Data set is classified into 9 different parameters which are considered important during the application for Masters. Those parameters are: gre scores, toefl scores, university rating, statement of purpose, letter of recommendation, undergraduate gpa, research paper, chance of admit.

## IV. CLASSIFICATION ALGORITHM:

### A. Linear Regression

These models are used to describe a relation between different variables by using the observed data into a line. Straight lines are

used in linear regression models, whereas curved line is used in logistic and non-linear regression models.

Linear regression model is a method used as response for only a single feature, it is based on supervised learning. Regression models always target a prediction value which is based on independent variables. It is used to calculate the relationship between two quantitative variables. Regression models differ upon – the type of relation between independent and dependent variables, the number of variables being used and the ones they are considering. It is taken under assumption that these variables are linearly related. Henceforth, we try to define a linear function that predicts the response value(y) as accurately as possible as a function of the feature or independent variable(x).cut-offs of the colleges, admission intake and preferences of students. Also, it helps students avoid spending time and money on counsellor and stressful research related to finding a suitable college. Table Type Styles

**B. Artificial Neural Network Neural (ANN)**

ANN networks are a series of algorithms which try to recognize relationships that are underlying in a dataset through a process which imitates the way human brain operates. Neural networks are referred as a system of neurons.

Neural networks are systems that are artificial and they were inspired from biological neural networks. They learnt to complete the task from different data sets and from examples without any specific rules assigned for the task. The main inspiration behind is that the system will automatically generate identification characters from the data that has already been passed through without being programmed and basic understanding of these data sets. Artificial Neural Networks have various neurons which are artificial and they are called as units. The units arrange themselves in a series of layers which all together make up as artificial Neural Networks. Any layer is supposed to have only a

**C. Random Forests**

The random forest is a machine learning algorithm which is widely used in regression and classification problems. Decision trees are built upon multiple different samples and then take their majority vote for average and bifurcation in case of regression. Random forest has the ability to handle a data set which contains continuous variables in case of regression and categorical variables in case of classification. Hence, it provides good results for classification problems. In industry lingo, reason behind forest works algorithm works so well is: Any huge quantity of moderately uncorrelated trees working as a body will outperform any of the individual constituent models

**V. METHODOLOGY AND IMPLEMENTATION:**

millions of units or dozen units as it depends upon the complexity of the system.

These units are arranged in a series of layers that together constitute the whole Artificial Neural Networks in a system. A layer can have only a dozen units or millions of units as this depends on the complexity of the system. Commonly, Artificial Neural Network has an input layer, output layer as well as hidden layers. The input layer receives data from the outside world which the neural network needs to analyze or learn about. Then this data passes through one or multiple hidden layers that transform the input into data that is valuable for the output layer. Finally, the output layer provides an output in the form of a response of the Artificial Neural Networks to input data provided.

**D. Decision Trees**

Classification is a two step process, learning and prediction. At first, model is developed based upon the given training data in its learning step. Then the model is used to predict response for the given data in prediction step. One of the most popular classification algorithms and easiest to learn and understand in decision tree. Decision tree algorithm is also used for solving classification and regression problems. Decision trees use a class label for predicting, for a record starts from the root of tree.

Then compare the values of the root with its record attribute. After the comparison, it follows the branch which is corresponding to the value and jump upon to the next node. There are two types of decision trees, Categorical variable and continuous variable. Categorical variable has a categorical target variable and continuous variable has a continuous target variable. Decision tree has three types of nodes, decision nodes, chance nodes and end nodes. Decision tree assigns a class label for each leaf node. Even the non-terminal nodes, the root and internal nodes, also contain attribute test conditions to separate records that have different characteristics

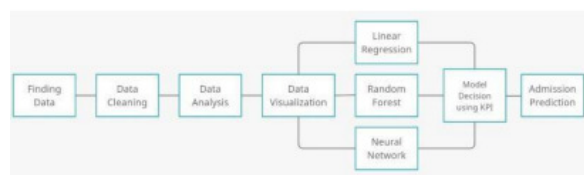


Figure 1. Classification Models

The primitive step to building a model for our use case is choosing the right dataset. For our predictions, we chose a dataset which contains all the important attributes that would affect the chances of admit. This is followed by data cleaning where we handle missing values present in various fields. Once the data is ready to be analyzed, we use various tools and libraries to visualize the data and perform analysis. This includes visualizing bar graphs and the correlation matrix. Once the data is ready to be processed, we split it into training and testing data. For this, we will be using 3 machine learning algorithms; linear regression, random forest and neural network. Once these models are built over the dataset, we compare them using key performance indicators. These indicators help us choose the right model for predicting whether an applicant has chances of admission.

**Advantages:**

- It helps student for making decision for choosing a right college.
- Here the chance of occurrence of error is less when compared with the existing system.
- It is fast, efficient and reliable
- Avoids data redundancy and inconsistency
- Very user-friendly
- Easy accessibility of data
- Number of personnel required is considerably less
- Provides more security and integrity to data.

**Disadvantages:**

- Required active internet connection.
- System will provide inaccurate results if data entered incorrectly.

**VI. CONCLUSION:**

Every year millions of students apply to universities to begin their educational life. Most of them don't have proper resources, prior knowledge and are not cautious, which in turn creates a lot of problems as applying to the wrong university/college, which further wastes their time, money and energy. With the help of our project, we have tried to help out such students who are finding difficulty in finding the right university for them.

It is very important that a candidate should apply to colleges that he/she has a good chance of getting into, instead of applying to colleges that they may never get into. This will help in reduction of cost as students will be applying to only those universities that they are highly likely to get into. Our prepared models work to a satisfactory level of accuracy and may be of great assistance to such people. This is a project with good future scope, especially for students of our age group who want to pursue their higher education in their dream college.

The graphs illustrate the chances of getting an admit against the predicted chances of getting an admit using different algorithms. Comparison in each of these models is done by evaluating Key Performance Indicators (KPI). With the help of this, it provides a cleaner output and helps in comparing the indicators like Root Mean Square Error, Mean Square Error, Mean Absolute Error and Adjusted R-Squared of an algorithm. At times accuracy provides ambiguous results if there are unequal observations or multiple classes are present in the dataset. Results show us that the highest accuracy is achieved through the linear regression model and the decision tree has the lowest accuracy.

**REFERENCES**

[1] Subba Reddy, Y and Prof. P. Govindarajulu, "A survey on data mining and machine learning techniques for internet voting and product/service selection", *IJCSNS International Journal of Computer Science and Network Security*, VOL.17 No.9 September 2017.

[2] Zhibo Wang, Jilong Liao, Qing Cao, Hairong Qi, and Zhi Wang, "Friend book: A Semantic based Friend

Recommendation System for Social Networks *IEEE Transactions on Mobile Computing*.

[3] J. Bobadilla et al. "Knowledge-Based System" Elsevier B.V.

[4] Hector Nunez, Miquel Sanchez-Marre, Ulises Cortes, Joaquim Comas, Montse Martinez, Ignasi RodriguezRoda, Manel Poch, "A Comparative study on the use of similarity measure in case based reasoning to improve the classification of environmental system situations," *ELSEVIER, Environmental Modeling and Software* (2003)

[5] DINO IENCO, RUGGERO G. PENSA and ROSA MEO, "From Context to Distance: Learning Dissimilarity for categorical Data Clustering," *Journal Vol. X. 10 2009, pages 1- 10.*

[6] Duc Thang Nguyen, Lihui Chen, Chee keong Chan, "Clustering with Multi viewpoint Based Similarity Measure," *IEEE Transactions on Knowledge and Data Engineering. Vol. 24. No. 6. June 2012.*

[7] Elham S.Khorasani, Zhao Zhenge, and John Champaign. *AMarkov Chain Collaborative Filtering Model for Course*

[8] Hana Bydžovská. *Course Enrollment Recommender System: Proceeding of the 9th International Conference on Educational Data Mining, P. 312 – 317.*

[9] Jamil Itmazi and Miguel Megias (2008), *Using recommendation Systems in Course Management Systems to Recommend Learning Objects, P. 234 – 240.*

[10] Queen Esther Booker (2009). *A Student Program Recommendation System Prototype: Issues in Information Systems, P. 544 - 551.*

[11] Akrivi Vlachou, Christos Doulkeridis, Kjetil Norvag, and Yannis Kotidis, "Identifying the Most Influential Data Objects with Reverse Top-k Queries," *Proceedings of the VLDB Endowment, Vol. 3, No. 1, Copy right 2010 VLDB Endowment 2150-8097/10/09.*