

PROPOSED SOLUTION:

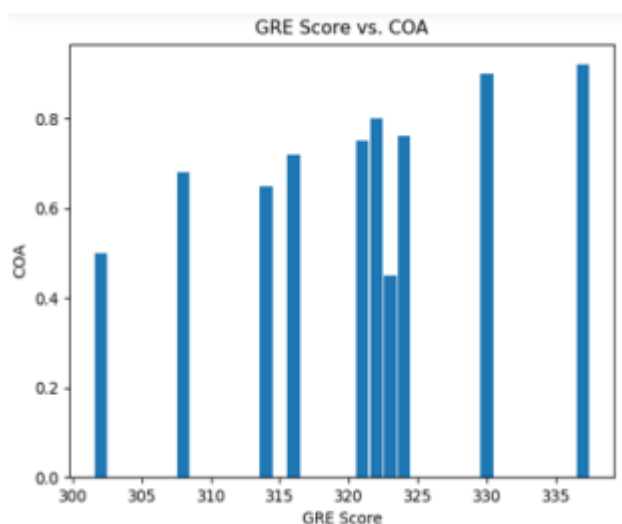
The main goal of this work is to create a Machine Learning model which could be used by students who want to pursue their education in the US. Many machine learning algorithms were utilized for this research. Linear Regression model compared to other ones. Students can use the model to assess their chances of getting admission into a particular university.

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 should be collected from multiple sources and also consider all the factors including which will play a tiny role in student admission process.

Data should be cleaned that is removing the noise in the data and filling the missing values or extreme values and finalising the attributes/factors which will have crucial importance in student admission process.

Novelty:

Several ML models have to be developed using various machine learning algorithms for admission to a particular university and the user interface has to be developed to access those models. Developed models are evaluated according to their accuracy scores. Once the model is finalised that model will be merged with node red for final deployment.



Data Cleaning and Analysis:

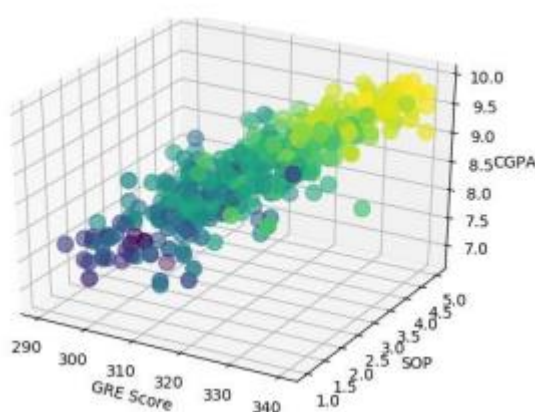
- Inspecting feature values that help identify what needs to be done to clean or pre-process until the range or distribution of values typical of each attribute is seen.
- There may be missing or noisy data, or anomalies such as the incorrect data form used for a column, incorrect measuring units for a particular column, or that there are not enough examples of a specific class.
- There are no missing values and outliers because we analysed 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.

- You can know that without machine learning, the problem is actually solvable. The data cleaning process has several key benefits to it.

Data Visualization:

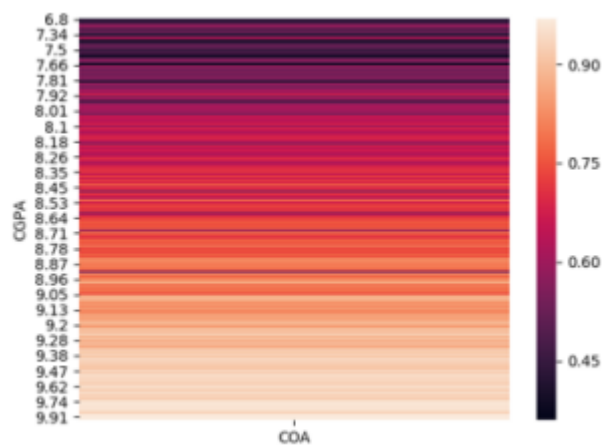
- After analysing the data, we will be able to know what the features and labels are, so from the above data, the label we have to consider is Chance of Admission and then we have to consider the parameters that influence or play a major role in Chance of Admission.



- 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.

The features given have a high impact on the probability of admission, so these features are considered only.

- Once the data visualization is done, we have to do predictive modelling for this purpose first we divide the data into train part and test part.
- We will develop model using machine learning algorithms on the train 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



ALGORITHMS:

For this work, several machine learning algorithms have been used, K- Nearest Neighbour and Linear Regression, Ridge Regression, Random Forest[4] are used to predict students ' likelihood of university admission based on their profile.

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 a integer value which will tell that with how many neighbours comparisons should be made. It 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.

Linear Regression:

It is an algorithm based on supervised learning of computers. It does the role of regression. Regression 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.

Ridge Regression:

Ridge regression is a regression method that is quite similar to unadorned minus squares. Linear regression: simply adding a penalty on parameters β to the linear regression objective function gives the ridge regression objective function. Ridge regression is an example of a shrinkage method: it shrinks the parameter estimates in the hopes of reducing uncertainty, increasing prediction accuracy, and aiding interpretation relative to the least squares.

Random Forest:

Random forest is a machine learning algorithm which is 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.

SOCIAL IMPACT:

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.