**1.1 Objectives:**

* To assess the scope and suitability of machine learning methods for University admission prediction.
* To develop initial machine learning models to predict the chance of admit.
* To analyze the performance of the model developed.

**1.2 Problem Statement:**

Each year, over 2 million college applications are filed, and each of them comes with a certain element of chance. Even students of the highest academic caliber apply with an element of randomness, and often face difficult situations as a result -the intended meritocracy inherent in college admissions gives way to uncertainty, doubt, and anxiety.

**2.Literature Review**

Literature survey plays a very important role in the project development. Literature survey provides the required knowledge about the project and its background. It also helps in following the best practices in project development. Literature survey also helps in understanding the risk and feasibility of the project. The feasibility of the project depends upon the risk of the project. If the resources, time and money are not available for the project development, then the risk is higher. Literature survey also gives light on various tools, platforms and operating systems suitable for project development. Once programming begins then the programmers require a lot of support and advice.

In this paper, author describes the capabilities of various algorithms in predicting several phenomena such as GRE Score, TOEFL Score, University Rating, Cgpa, LOR, Research these parameters concluded that major techniques like decision trees, artificial neural networks, clustering and regression algorithms are suitable to predict Chance of Admit. This shows that the decision trees and k-means clustering are best-suited data mining techniques for this application.

**3. Data Collection**

This dataset is originally from the Kaggle. The objective of the dataset is to predict chance of admission to a student, based on certain input measurements included in the dataset. Several constraints were placed on the selection of these instances from a larger database. The datasets consists of several predictor variables and one target variable, Outcome. Predictor variables includes the GRE Score, TOEFL Score, University Rating, Cgpa, LOR, Research Experience . Chance of Admit this is the column that you predict yourself.

**4. Methodology**

* Eighty percent for training and 20% for testing
* R language(open source) used
* R neural net package for model generation
* A three layer model generated
* Seven hidden nodes
* Correlation and mean square error calculated
* Validation done using 20% dataset
* Models generated separately for the 2 locations

Brief Description of Algorithm Used:

**Multiple linear regression:**

Machine learning, more specifically the field of predictive modeling is primarily concerned with minimizing the error of a model or making the most accurate predictions possible, at the expense of explain ability. In applied machine learning we will borrow, reuse and steal algorithms from many different fields, including statistics and use them towards these ends. As such, linear regression was developed in the field of statistics and is studied as a model for understanding the relationship between input and output numerical variables, but has been borrowed by machine learning. It is both a statistical algorithm and a machine learning algorithm.

Linear Regression Model Representation:

Linear regression is an attractive model because the representation is so simple. The representation is a linear equation that combines a specific set of input values (x) the solution to which is the predicted output for that set of input values (y). As such, both the input values (x) and the output value are numeric.

The linear equation assigns one scale factor to each input value or column, called a coefficient and represented by the capital Greek letter Beta (B). One additional coefficient is also added, giving the line an additional degree of freedom (e.g. moving up and down on a two-dimensional plot) and is often called the intercept or the bias coefficient.

For example, in a simple regression problem (a single x and a single y), the form of the model would be:

y = B0 + B1\*x

In higher dimensions when we have more than one input (x), the line is called a plane or a hyper-plane. The representation therefore is the form of the equation and the specific values used for the coefficients (e.g. B0 and B1 in the above example).

It is common to talk about the complexity of a regression model like linear regression. This refers to the number of coefficients used in the model.

When a coefficient becomes zero, it effectively removes the influence of the input variable on the model and therefore from the prediction made from the model (0 \*x = 0). This becomes relevant if you look at regularization methods that change the learning algorithm to reduce the complexity of regression models by putting pressure on the absolute size of the coefficients, driving some to zero. Now that we understand the representation used for a linear regression model, let’s review some ways that we can learn this representation from data.

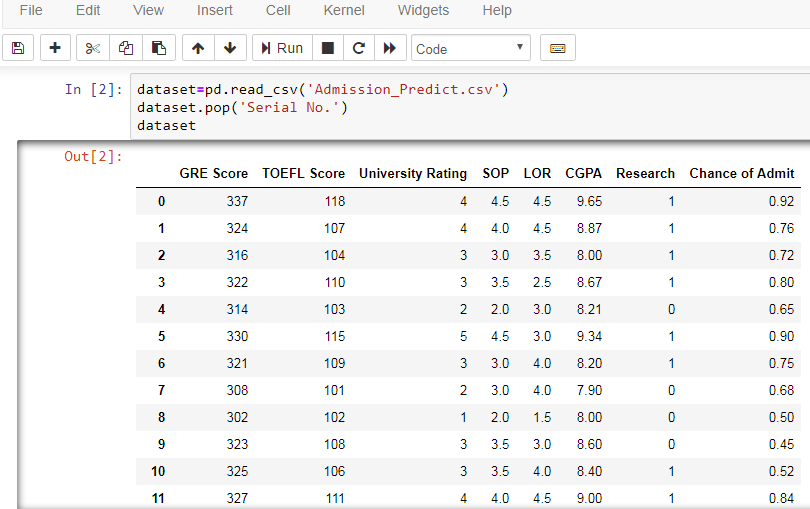
**Accuracy Measures**:

Multi Linear Regression algorithm is used in this research work. Experiment is performed by using r2\_score.

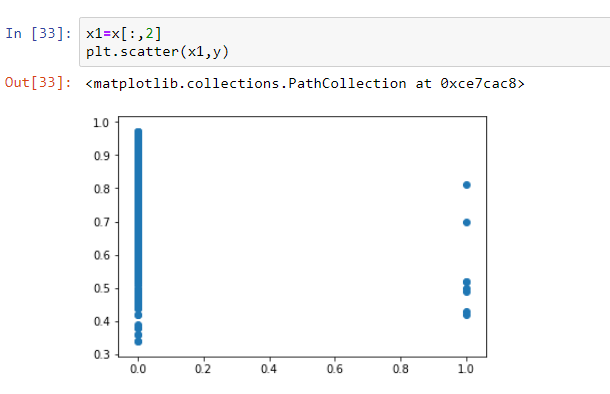
**4.1 Exploratory data analysis:**

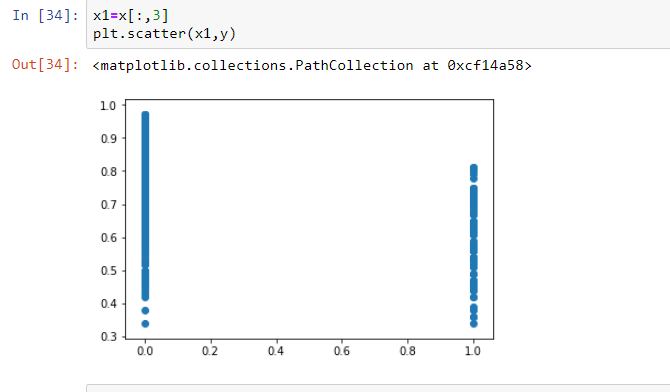
4.1.1 Figures and tables:

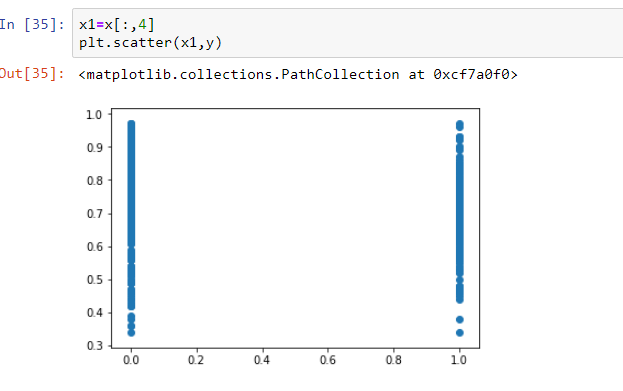
Dataset for Data Center:

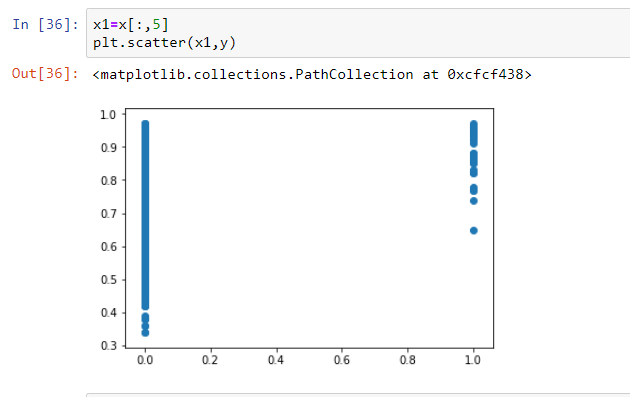


Multilinear Regression:



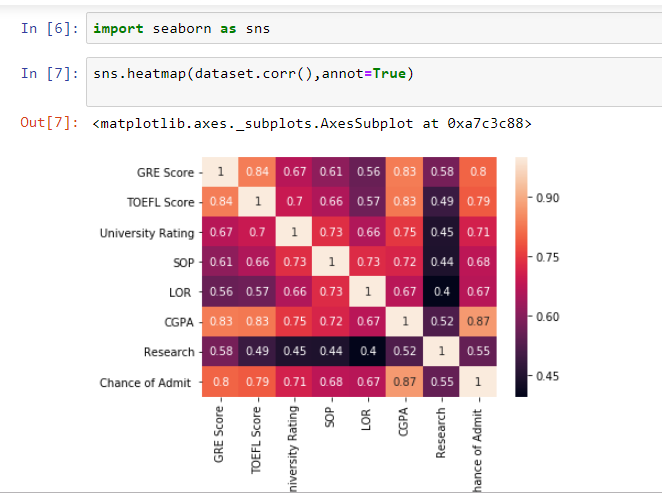






**4.2 Statistical techniques and data visualization:**

By importing matplotlib.pyplot library we have drawn graphs. We have visualized the probability. And we have used the co-relation function to demonstrate the impact of every factor on each other.

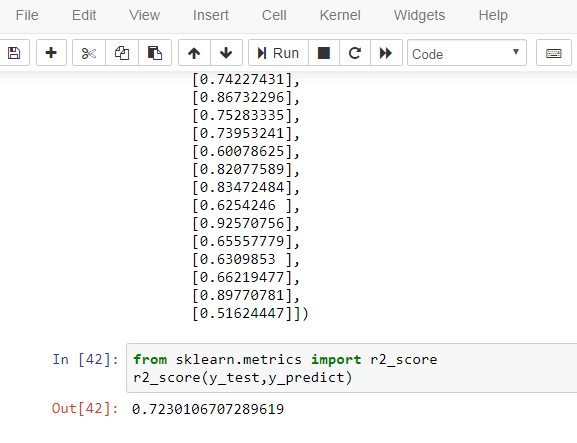


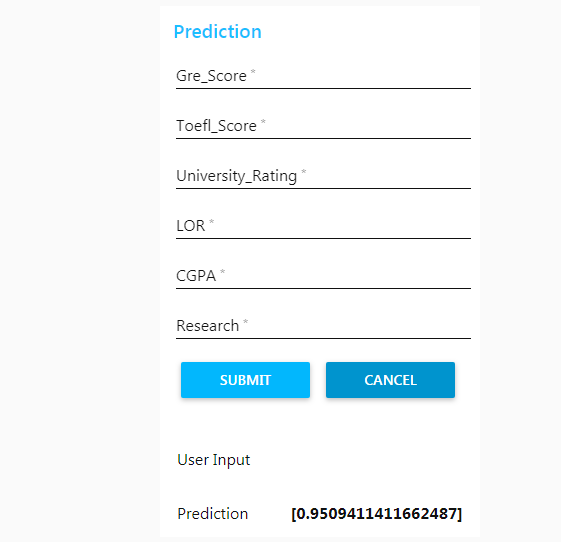
**4.3 Data modeling using supervised learning algorithms:**

In general we have two types of learning algorithms, supervised and unsupervised learning algorithms and in detail it consists of different techniques like, Multi Linear Regression.

Since our model comes under supervised learning algorithm we applied every technique, of the algorithm and based on the accuracy values we obtained we chosen Multi Linear Regression and our model has dependent variable which is Chance of Admit means the output can be the probability of Admission for a student.

**5. Findings and Suggestions**





**6. Conclusion**

According to the survey what is been done, there are still chances that the result might get improved. During the research which we have carried out it is observed that the most of the algorithm and methods which is been used by most of the authors does not use a collective approach where all the factors that decides the chance of admission can be simultaneously be used for predicting the probability. There can be further improvements because in some analysis very few amount of datasets have been used. Therefore the result can be improved by using a large dataset. Larger the datasets the better the results.