**GRADUATE ADMISSION PREDICTION**

**CHAPTER 1**

**Introduction**

**1.1** **Introduction:**

In today’s era we see a lot of students pursuing their education away from their home countries. The main country targeted by these international students is The United States of America. Majority of the international students in the United States of America are from India and China. In the past decade the number of Indian students pursuing post graduate education from the USA has rapidly increased. With the increase in the number of international students studying in the USA, each applicant has to face a tough competition to get admission in their dream university. Generally, as the students don’t have much idea about the procedures, requirements and details of the universities in the USA 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. Apart from these the education consultancy firms there are few websites and blogs that guide the students on the admission procedures. The drawback of the currently available resources is that they are very limited and also they are not truly dependable taking into consideration of their accuracy and reliability. The aim of this research is to develop a system using machine learning algorithms, we will name it as Student Admission Predictor (SAP). 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. A simple user interface will be developed for the users to access the SAP system.

Keywords: Student Admission Predictor; Machine Learning; Regression.

**1.2 Objective of Research:**

Majority of universities in the USA follow similar guidelines for providing admission to students. Universities take into consideration different factors like score on aptitude based examination like the General Record Examination (GRE), command over the English language is judged based on their score in English competency test like Test Of English as a Foreign Language (TOEFL) OR International English Language Testing System (IELTS), their work experience in same or other fields, the quality of the Letters Of Recommendation (LOR) and the Statement Of Purpose documents provided by the student etc. Based on the overall profile of the student decision is taken by the universities admission team to admit or reject a particular candidate.

Every candidate has to take all the required examination and build a strong profile to secure admission in their dream universities in the USA. 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 shortlisting the universities where the student has high chances of admission is difficult for mainly for the international students, 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 shortlisting 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.

**1.3 Problem Statement:**

Graduate Admission Prediction helps the students to know about what is the chance for students to get admission in reputed Universities. In order to do This we are using the regression methodology and the algorithm what we are going to use is Multilinear Regression. So, by using this model we are going to obtain a perfect model to predict the chance of getting admission in reputed university.

**1.4 Industry Profile:**

Our Project Graduation Admission Prediction mostly helpful for Educational society as well as students. For educational institutes it helps to maintain online records without maintain in books with employees. For students this model gives the prediction that weather we are able to get seat in reputed university or not without paying huge amounts to consultancies.

**CHAPTER 2**

**Review of Literature**

This section provides the literature review of the work that has previously done on predicting the chances of student’s enrolment in universities. There have been several project and studies performed on topics related to student’s admission into universities. used multiple machine learning models to create a system that would help the students to shortlist the universities suitable for them also a second model was created to help the colleges to decide on enrolment of the student. Nave Bayes algorithm was used to predict the likelihood of success of an application, and multiple classification algorithms like Decision Tree, Random Forest, Nave Bayes and SVM were compared and evaluated based on their accuracy to select the best candidates for the college. Limitation of this research as that it did only relied on the GRE score only.

There are several attempts recently by researchers and administrators to mine student’s data. These

There are several attempts recently by researchers and administrators to mine student’s data. These

There are several attempts are to either predict student’s enrollment rates, failure rates, dropout rates and/or graduation rates. Some of these approaches include, factors analysis, path analysis, and discriminant analysis. These techniques have shown different levels of success. However, in the case of detecting patterns and prediction they are not as effective as artificial neural networks. The drawback of this algorithm is it is not producing the output accurately and not much efficient result.

The second method they followed is naive Bayesian classification in this methodology they also used the technique of giving input parameters and predicting the output but it was taking only less number of parameters for less number of parameters only we are obtaining accurate result.

K Nearest Neighbor’s and Decision Tree algorithms were used as they were found to be the best fit for the system developed. 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. But it is un able to give the percentage of chance to predict it only tells us whether we can get admitted or not.

So, now we moved to multi linear regression model.

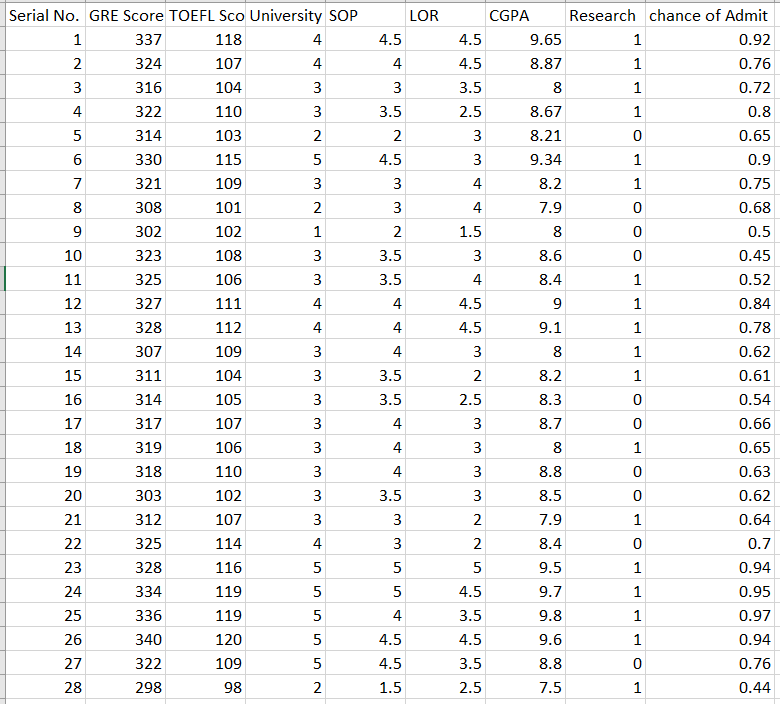
**CHAPTER 3**

**Data Collection**

This section describes, in brief, the data that has been used for the project. Data from multiple sources was used in this project, the major amount of data was extracted from public website kaggle, data regarding the rankings, fees and enrolment in colleges was obtained from a leading educational consultancy firm The Mentors Circle in India. Data from both the sources was integrated together to form a staging data-set. For predicting the chance of a student getting shortlisted in universities the final data-set was divided into multiple data-sets each representing a particular university. For predicting the list of universities suitable for students based on their profile data of all the students the staging data-set was updated only to have records of students who had successfully secured admission in the universities. Below table shows the different features of the data-sets. The parameters used as independent variable are listed below:

1. GRE: General Record Examination
2. TOFEL: Test of English as a Foreign Language
3. CGPA: Cumulative Grade Point Average

There are several attempts recently by researchers and administrators to mine student’s data. These



**Fig 3.1: Dataset Table**

**CHAPTER 4**

**Methodology**

**4.1Exploratory Data Analysis:**

Can machine learning algorithms improve the accuracy in predicting the application status of students aspiring to enroll for M Tech course at universities in the USA?

The answer is obviously yes, since as now a days machine learning has been playing an important roll to improve the efficiency of any machine to produce the accurate result. some of the methodologies we follow are like:

**Business Understanding:**

Initially good amount of time was spent on understanding the problem statement by understanding the concerns of students regarding the current application process, the objectives of the research were defined in this process.

**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-set would be explained in more detail further.

**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.

**Modelling:**

Multiple machine learning models were developed to predict the likelihood of success of the student’s application in a particular university. The user interface was developed to allow the users to access these models.

**Evaluation:**

Models developed were evaluated based on their performance and accuracy. More information will be presented in the evaluation section of the paper.

**Deployment:**

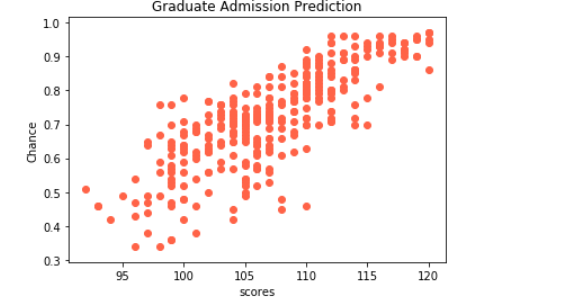
Once the models were evaluated they were integrated with code developed for user interface using the Node red



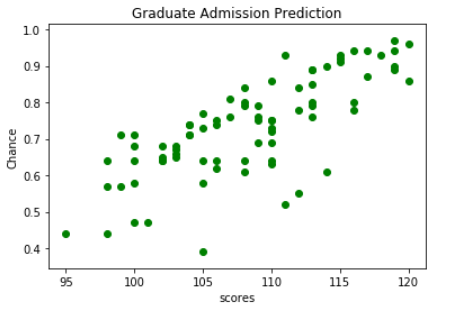
**Fig 4.1 methodology**

*4.1.1 Figures and table:*

The graph that shows the x\_train and y\_train data set



**Fig: 4.1.1 Graph for train data**

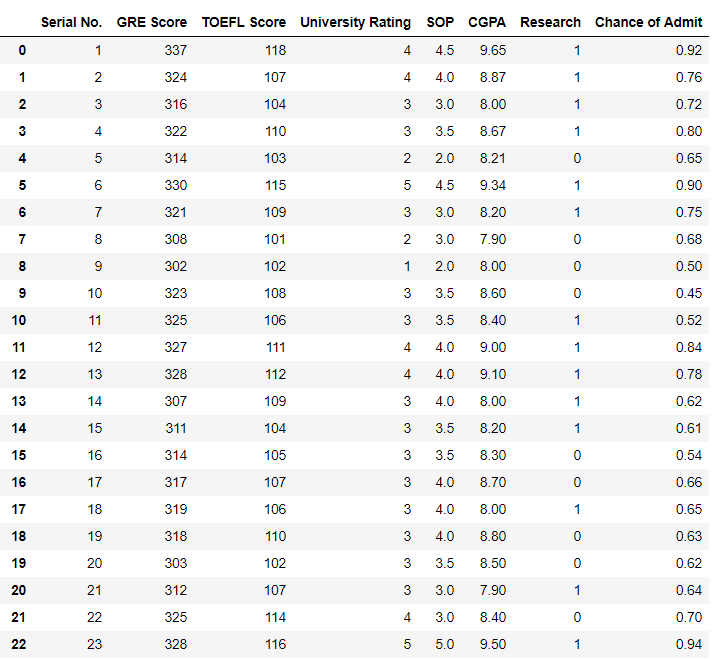


**Fi: 4.1.2 Graph for test data**

The above graphs are drawn with the help of Matplotlib library. This contains all the methods to plot the values in the form of graphs.Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter , wxPython, Qt, or GTK+. ... SciPy makes use of Matplotlib.

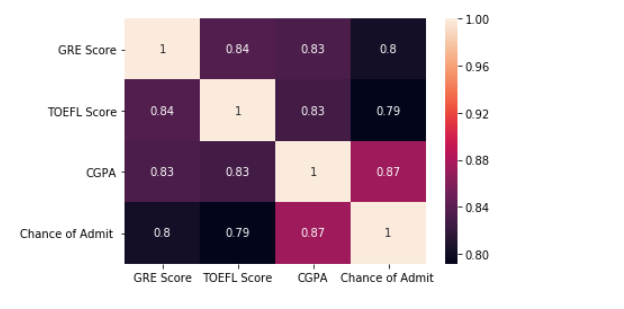
**Tables:**

The dataset what we have used contains several rows and we have to identify the rows which are important to do prediction before doing any analysis the data that contains columns like as follows



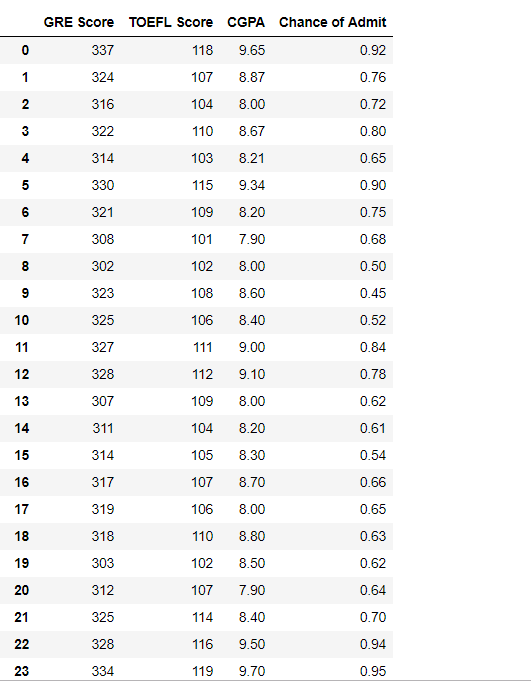
**Fig: 4.1.3 dataset without preprocessing**

For the above dataset we have to do correlation and have to find out the most accurate independent variables the correlation what we obtained is as follows:



**Fig: 4.1.4 Correlated Graph**

From the above data analysis, we are going to take GRE Score, TOFEL Score and CGPA as the most independent variables. So by applying slicing and data preprocessing we are going to remove the unwanted columns. Data prepossessing includes the steps like removing the unwanted columns, if any null values are available then filling them with different techniques, doing one hot encoding etc. Thus this will be as follows:



**Fig : 4.1.5 Dataset after preprocessing**

**4.2 Statistical techniques and visualization:**

The basic examples where Multiple Regression can be used are as follows: The selling price of a house can depend on the desirability of the location, the number of bedrooms, the number of bathrooms, the year the house was built, the square footage of the lot and a number of other factors. The height of a child can depend on the height of the mother, the height of the father, nutrition, and environmental factors.

It is the most common form of Linear Regression. Multiple Linear Regression basically describes how a single response variable Y depends linearly on a number of predictor variables. The have used are regression. techniques we Regression models are used to predict a continuous value. Predicting prices of a house given the features of house like size, price etc. is one of the common examples of Regression. It is a supervised technique.

The goal of this blog post is to equip beginners with the basics of the Linear Regression algorithm with multiple variables predicting the outcome of the target variable. This is also known as Multiple Linear Regression. Simple linear regression model has a continuous outcome and one predictor, whereas a multiple linear regression model has a continuous outcome and multiple predictors (continuous or categorical). A simple linear regression model would have the form:

Multiple linear regression can model more complex relationship which comes from various features together. They should be used in cases where one particular variable is not evident enough to map the relationship between the independent and the dependent variable.

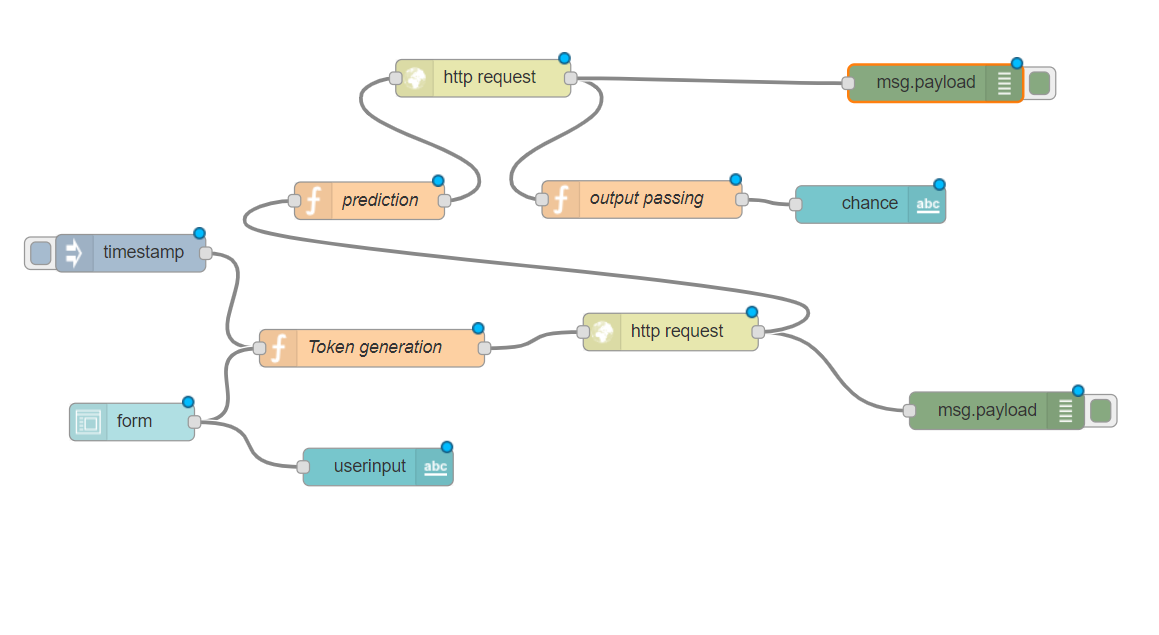
**4.3 Data Modeling and visualization:**

For data modelling we have followed Multi linear regression model and we have obtained an accuracy of 75%. Thus my model is able to give perfect result to the seventyFive percent effectively.

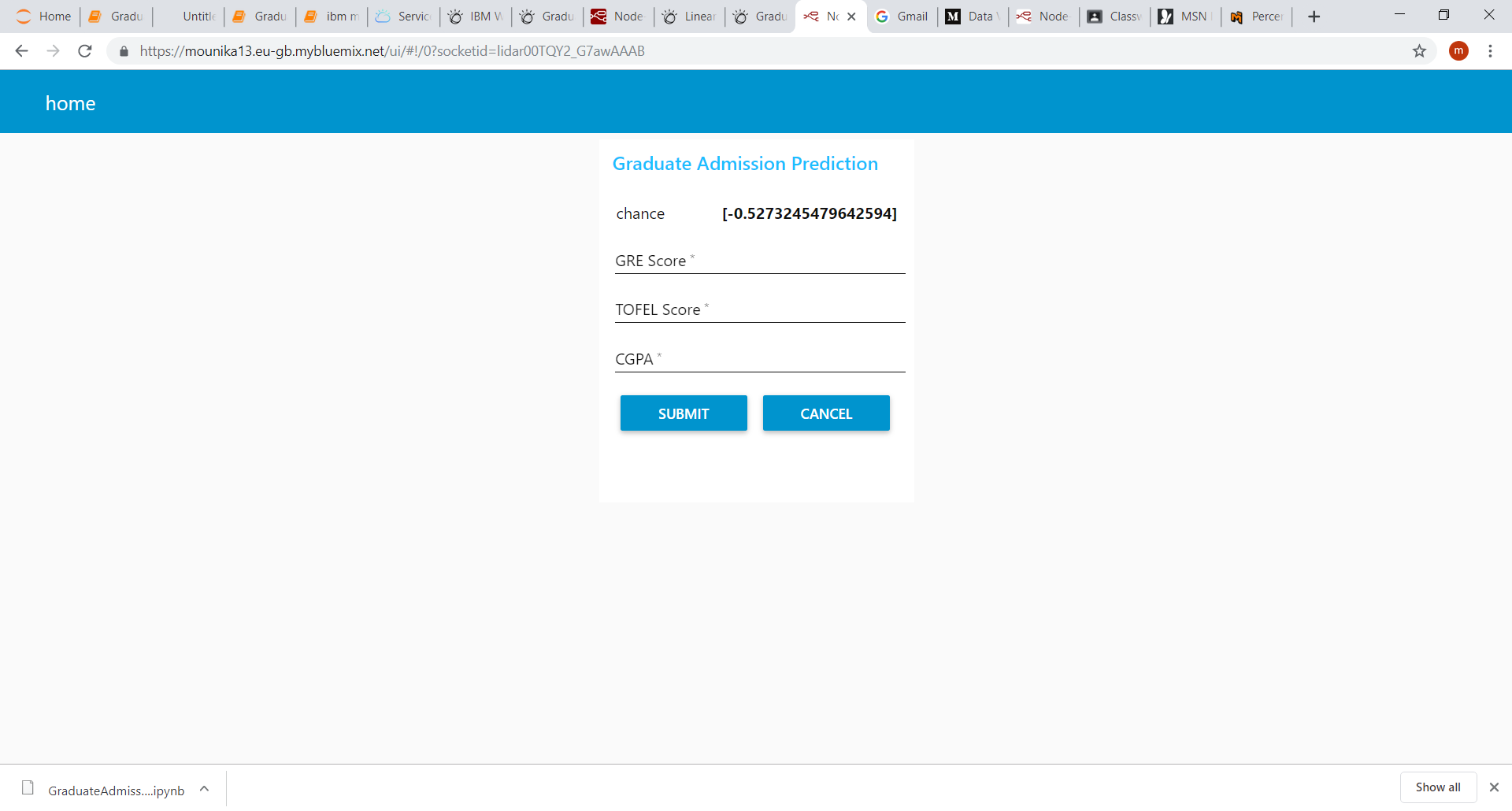
The visualization technique we have used is node red. Node-RED is a programming tool for wiring together hardware devices, APIs and online services in new and interesting ways. It provides a browser-based editor.

Since early 2014, the mobile has overtaken the personal computer (desktop/laptop) as the leading device used to navigate the Net. Along with the mobile, a number of other portable devices that connect to the Internet have also started proliferating at a very quick rate. Nowadays, most of us carry or possess at least one Internet based device and a mobile. So, the Internet of Things now doesn’t only mean different ‘things’, but has evolved into ‘intelligent things’ which have on-board computation and network connections. Most importantly, they have the capability to sense the environment around us and, accordingly, act intelligently. These devices are now referred to as connected devices, smart objects or the Web of Things.

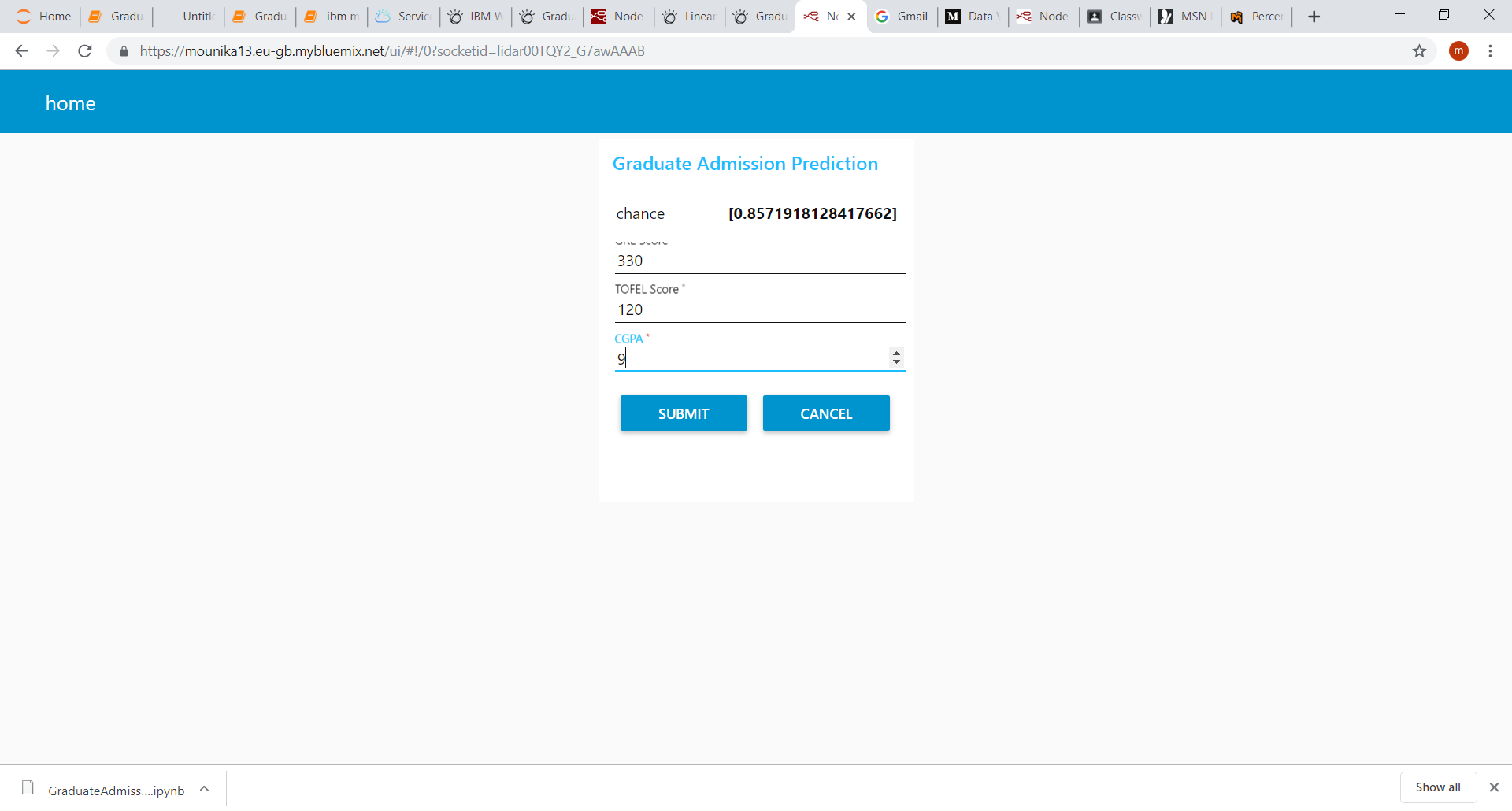
This adaptation of technology has led to many developments in the visual programming and visual coding arenas. As a result, Node-RED has evolved into a visual programming/coding tool based on the already popular Node.js (a server side Java Scripting platform), mainly targeting the Internet of Things space.



**Fig 4.3.1 Node Red Implementation**



**Fig 4.3.1 User Interface**



**Fig 4.3.1 User Interface**

**CHAPTER 5**

**Findings and Sugessions**

The main objective of this research was to develop a prototype of the system that can be used by the students aspiring to pursue their education in the USA. Multiple machine learning algorithms were developed and used for this research. Multi Linear proved to best-fit for development of the system when compared with the Logistic regression model. The model can be used by the students for evaluating their chances of getting shortlisted in a particular university with an average accuracy of 75%.

This Prediction model gives us an accurate result only if we have a small amount of data thus this is a drawback. So, in future you can implement a better than this. The additional thing what we can do is we can give the list of universities of which he have chance to get admit

<http://trap.ncirl.ie/3102/1/himanshumahadevsonawane.pdf>

<https://www.google.co.in/search>?

<https://www.kaggle.com/mohansacharya/graduate-admissions>

<https://opensourceforu.com/2017/09/node-red/>

**CHAPTER 6**

**Conclusion**

The main objective of this research was to develop a prototype of the system that can be used by the students aspiring to pursue their education in the USA. Multiple machine learning algorithms were developed and used for this research. Multi linear proved to best-fit for development of the system when compared with the Logistic regression model. The model can be used by the students for evaluating their chances of getting shortlisted in a particular university with an average accuracy of 75%.