**SMARTBRIDGE SUMMER INTERNSHIP**

**PROGRAM-2019**

GROUP NAME: **LAZY GEEKS**

GROUP MEMBERS:

**1)SURRAM RAMYA**

**2**)**KANNOJU SHRUTHI**

**3)REVURI MANIKANTA**

**4)MULUKA SAHASRA**

**5)TUNUGURI SANTOSH**

**ADVERTISING BASED ON USAGE**

**INTRODUCTION:**

Advertisement is an efficient and effective technique to promote goods, services ,and ideas. It is a paid form of non-personal communication wherein business information is made available for potential customers.

Advertisements were chosen as the tool for analysis for many reasons.

Advertisement is derived from Latin word “Advertise” which literally means “to turn the minds-of-towards”. Advertisement promotes and supplements selling of products, services , and ideas to a great factual information with fascinating emotional appeal. Hence without an proper advertisement no business can prosper.

Advertising messages are usually paid for by sponsors and viewed via various traditional media including mass media such as newspaper, magazines, television commercial, radio advertisement ,outdoor advertising or direct mail or new media such as blogs, websites or text messages.

Paid communication

Creative text

A technique of success

Advertisement is..

Execution creative idea

Marketing technique

Media dependent

**OBJECTIVES OF RESEARCH:**

* The fundamental idea behind advertisement is to increase the business by selling goods/services .Besides ,there are other objectives advertisement, significant of them are-
* To promote newly launched products among the potential customers
* To promote personal selling program.
* To create awareness among maximum people about your business in a short period of time.
* To enter national or even international market and motivate new group of customers.
* To enhance the goodwill and build credibility among the customers by promising to provide better quality of products and services.
* To study and understand the importance of online advertising.
* To study which type of online advertisement is preferred by consumers.

**PROBLEM STATEMENT:**

Developing a machine learning model to predict what type of advertisements and user is interested in based on their previous data, this predictions is used to make further decisions and to predict the probability of clicking on that advertisement by the user.

**Review of literature:**

**The Indian advertising industry ,with an estimated value of INR 13,200 crore, has shown an astonishing growth and creativity in the recent times .The industry is growing at an average rate of 10-12 percent per annum. some of the commonly used media for advertising are TV, radio, websites, newspapers, magazines and out-of-home(ooh).**

**This project includes the number of users reviewing the advertisements based on usage in their daily life.**

**In this project we used classification algorithms. In order to train the model, we make use of dataset with reviews about advertisements clicked by the users.**

**From these review, we can predict which type of advertisement is being clicked by users according to their age groups.**

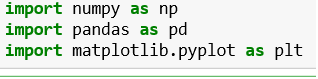
**DATA COLLECTION:**

The dataset has been used for prediction of number of users reviewing about advertisements consists of 10 attributes and 1000 instances.

* The attributes present in the dataset are Daily time spent on site, Age, Income, Daily Internet usage, Ad Topic Line, City, Male, Country, TimeStamp, Clicked on Ad
* Before building the model, preprocessing of dataset is required.

Steps involved in Data preprocessing are as follows:

* **Importing libraries such as numpy, pandas and matplot.**



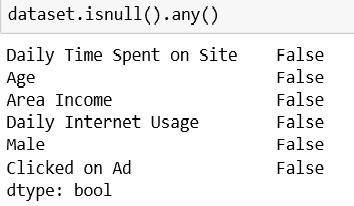
* **Importing dataset.**



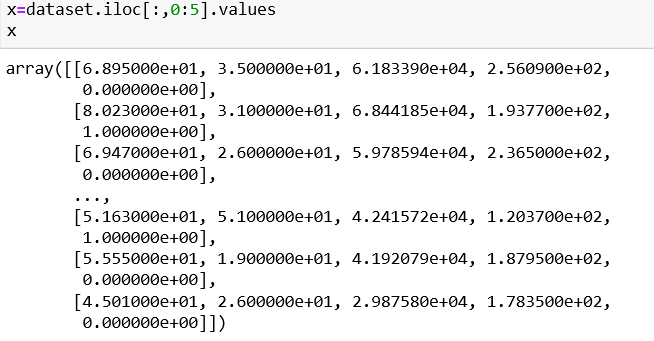
* **Handling missing values.**

The missing data can be handled in three ways:

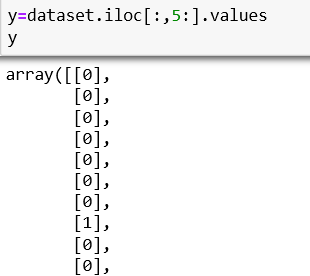
* Deleting the rows which are empty(missing values).
* Find the mean of attribute and fill the cell with obtained mean value.
* Find the mode of the attribute and fill the cell with obtained more value.



* **Separating independent and dependent variables.**

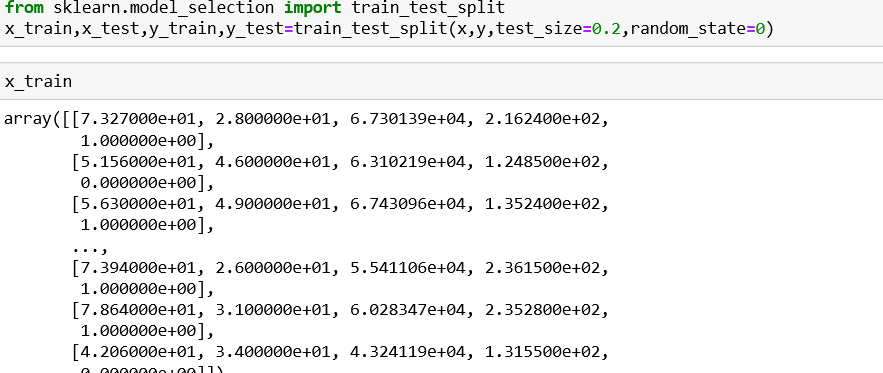


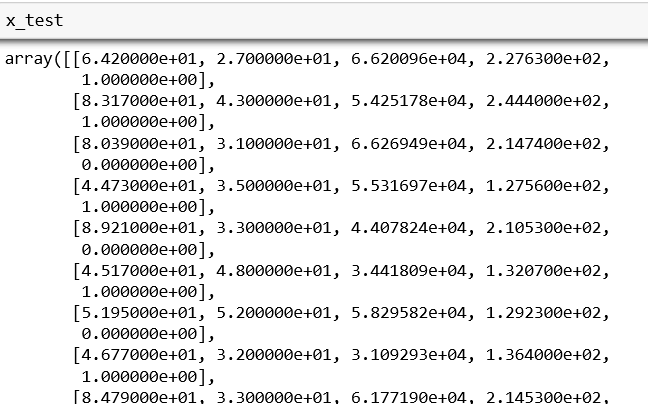
X is dependent variable.

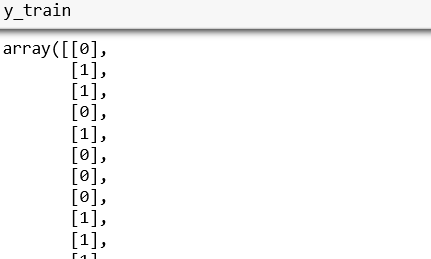


Y is independent variable.

* **Splitting training and testing data**







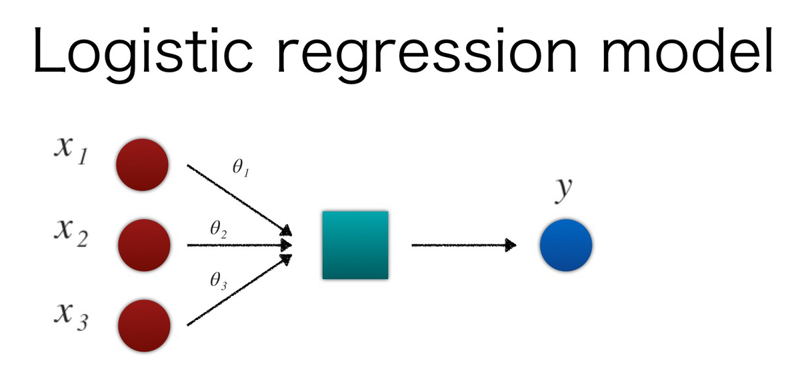
**METHODOLOGIES:**

**In machine learning, classification is a supervised learning approach in which the computer program learns from the data input given to it and then uses these learning to classify new observation.**

**Classification Algorithms:**

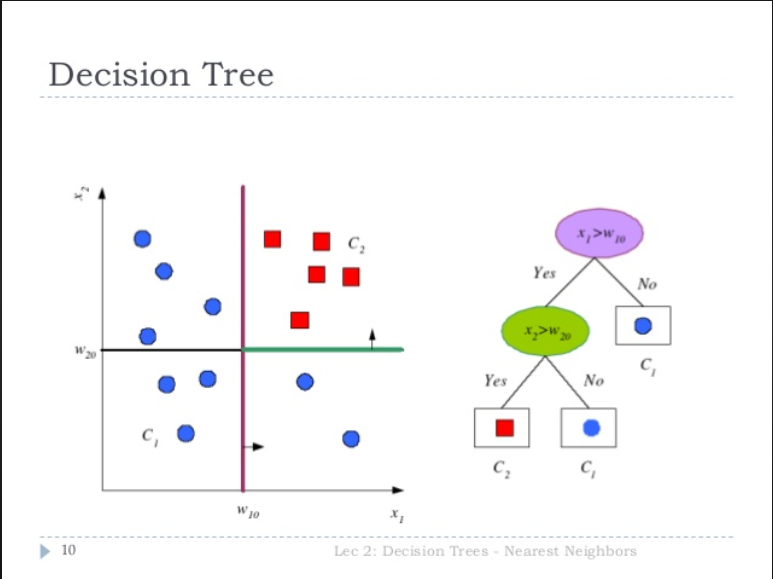
* **Logistic Regression**
* **Decision Tree**
* **K Nearest Neighbour**
* **Random Forest**
* **Logistic Regression:**

**Logistic Regression can be used to predict a dependent variable on the basis of continuous and/or categorical independents and to determine the percent of variance in the dependent variable explained by the independents; to rank the relative importance of independents. The impact of predictor variables is usually explained interms of the odd ratio. The odd ratio is the ratio of odd of the odds of an event occurring in one group to the odds of it occurring in another group, or to a sample-based estimate of that ratio.**

****

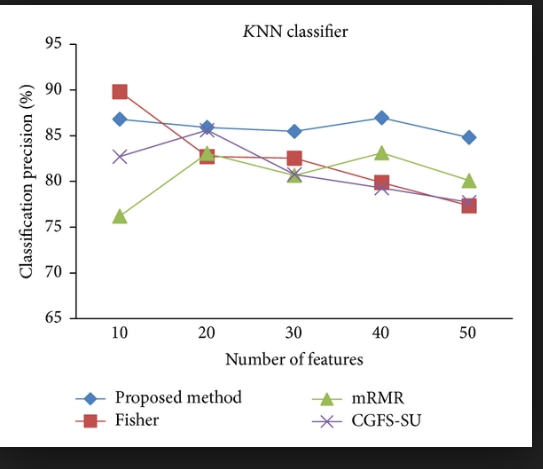
* **Decision Tree:**

**Decision Tree builds classification or regression models in the form of tree structure .It breaks down dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree** . The final result is a tree with decision nodes and leaf nodes. A decision node has two or more branches and a leaf node represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor called root  node. Decision trees can handle both categorical and numerical data.



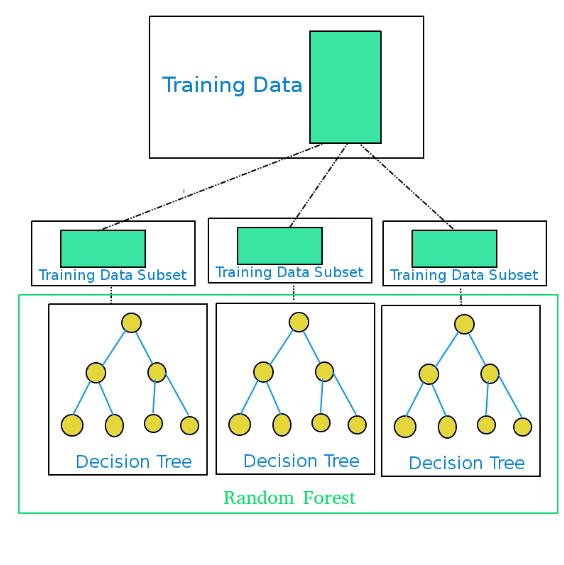
* **K Nearest Neighbor :**

 It takes a bunch of labelled points and uses them to learn how to label other points. To label a new point, it looks at the labelled points closest to that new point (those are its nearest neighbors), and has those neighbors vote, so whichever label the most of the neighbors have is the label for the new point (the “k” is the number of neighbors it checks).

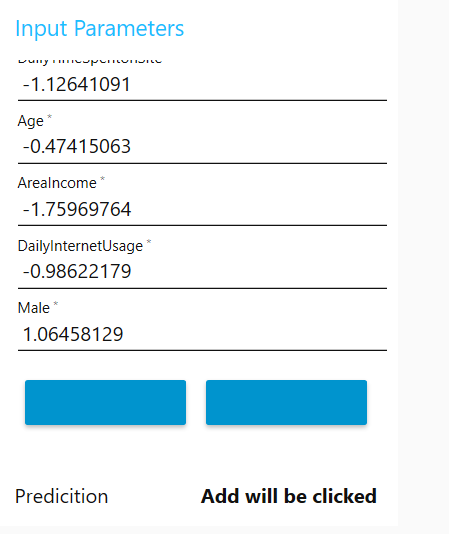
****

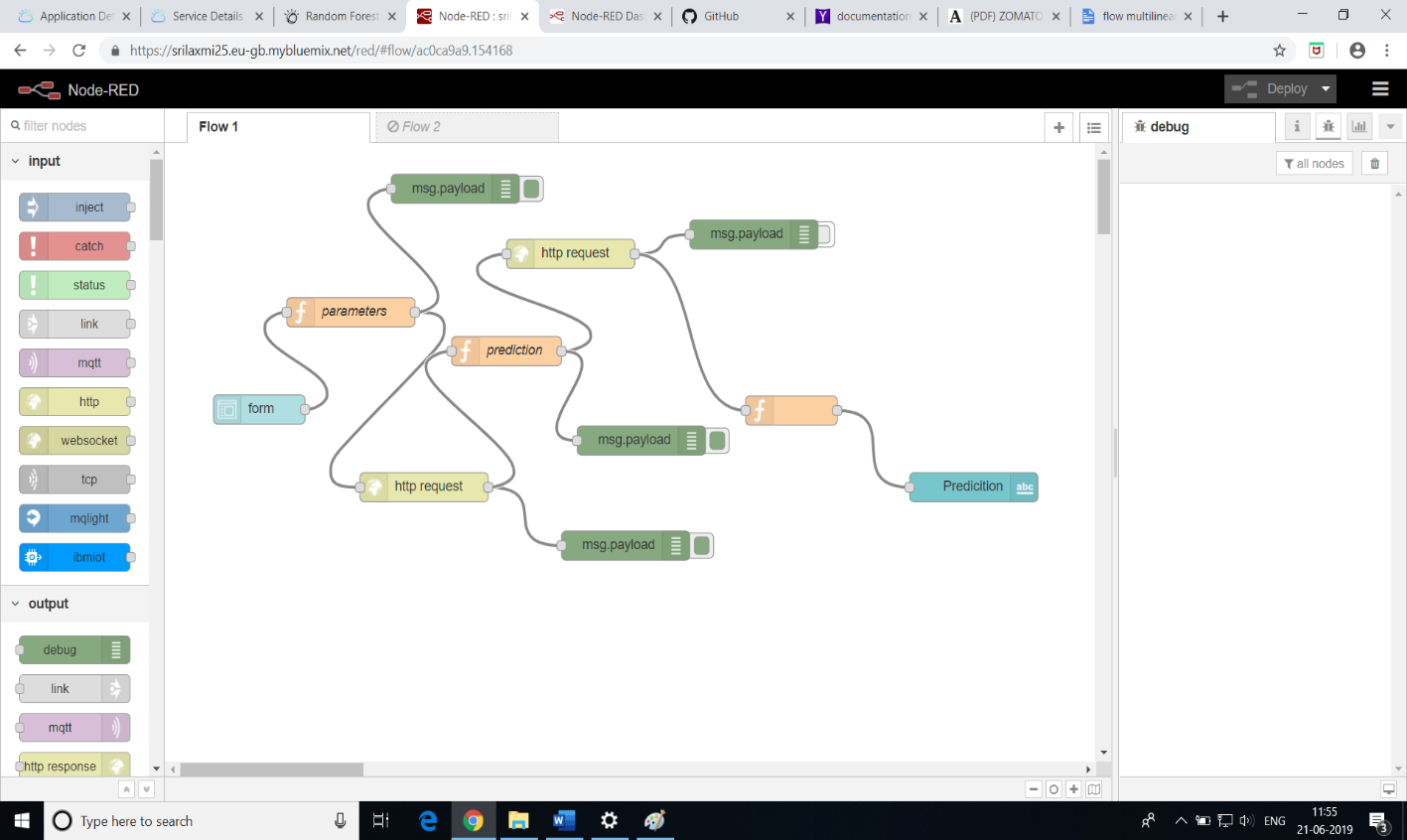
* **Random Forest :**

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees’ habit of over fitting to their training set.



**FIGURES AND TABLES:**

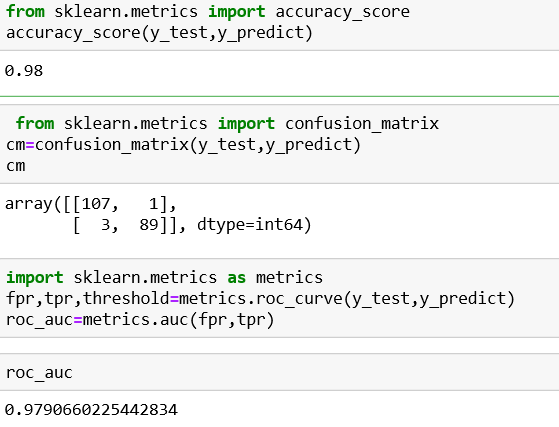
****



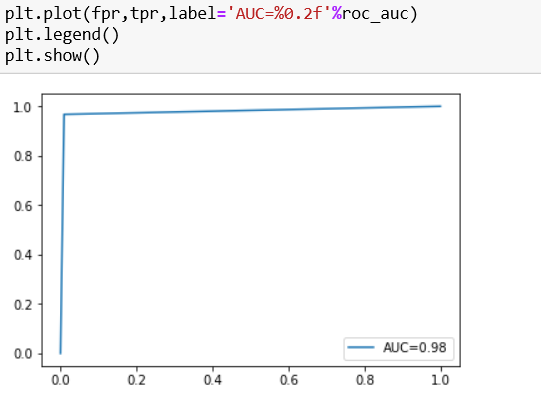
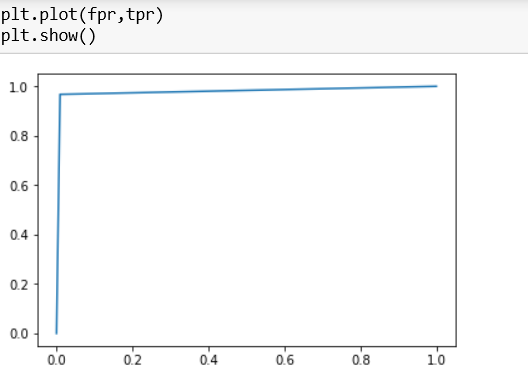
**STATISTICAL TECHNIQUES AND DATA VISUALIZATION:**

* **Statistical Technique:**

**In classification, accuracy\_score is used to calculate accuracy.**

****

* **Data visualization:**

****

**DATA MODELLING USING SUPERVISED ML TECHNIQUES:**

Supervised ml techniques in our model are

Model built using Decision Tree:

🡪importing DecisionTreeClassifier package

🡪Fitting model using x\_train and y\_train

🡪predicting values of y using x\_test

🡪calculating accuracy\_score

Model built using Random Forest:

🡪importing RandomForestClassifier package

🡪Fitting model using x\_train and y\_train

🡪predicting values of y using x\_test

🡪calculating accuracy\_score

Model built using K Nearest Neighbor:

🡪importing KNeighborsClassifier package

🡪Fitting model using x\_train and y \_train

🡪predicting values of y using x\_test ,calculate accuracy\_score

**FINDINGS AND SUGGESTIONS:**

**FINDINGS-**

**In this model, we have used**

* **Logistic Regression**
* **Random Forest**
* **K Nearest Neighbour**

**For prediction of data whether they will click ad or not. From all these classification algorithms, we have got 98% accuracy – Logistic Regression**

**94%accuracy- Decision Classification**

**96%accuracy- Random Forest**

**96%accuracy- K Nearest Neighbour**

**SUGGESTIONS-**

**If the data is inconsistent, preprocess the data and predict the value using classification algorithm.**

**Accuracy increases by giving more dependent variables/values.**

**CONCLUSION:**

* **In this project, we have used concepts of machine learning model are implemented by using classification algorithms to predict the number of users reviewing and clicking on advertisements based on their usage.**
* **Based on the predicted value, the administrator of the adviser knows for whom they have to send the advertisements so that the business rate will be increased**