**A Model for prediction of consumer conduct using machine learning algorithm**

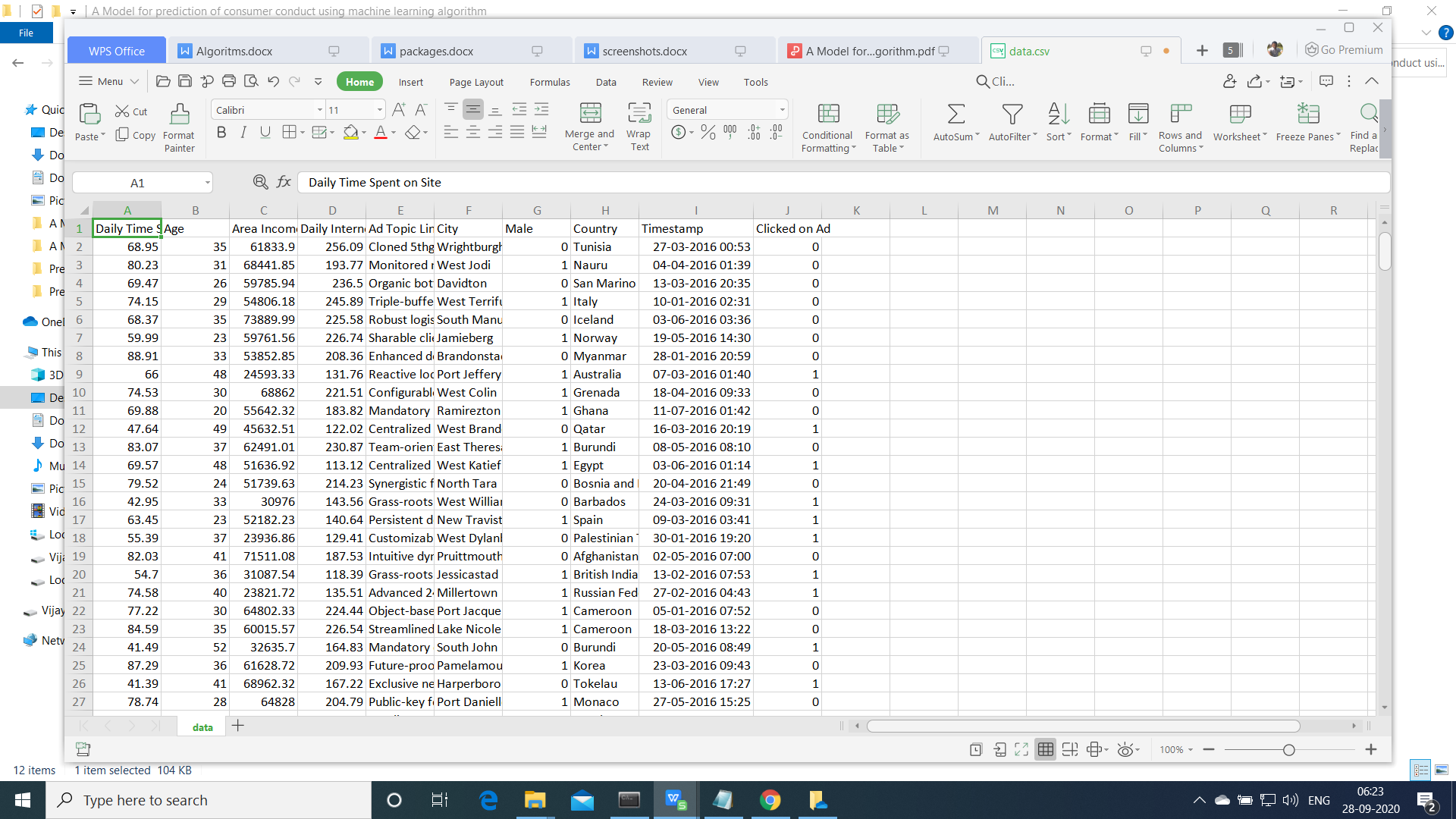
**Introduction:**

Predictive marketing empowers digital marketers with the ability to make intelligent, forward-looking decisions through **automatization**, **visualization** and **user experience innovations** coupled with powerful predictive models. Companies have always tried to predict the future of their business and today, enterprises have access to an infinite amount of data. It is all about being **proactive** and not reactive.

Customer behavior prediction identifies behaviors among groups of customers to predict how similar customers will behave under similar circumstances. Is about the creation of a **mathematic** **construct** to represent the common behaviors observed among groups of customers in order to **predict** how similar customers will **behave** under similar circumstances.

Customer behavior models are typically based on **data** **mining** of **customer** **data** and can be used to predict what a group of customers will do in response to a determine marketing action. If the model is right and the market follows, most customers in the group will act as predicted by the model.

Data Description:



Downloaded From kaggle

Location:https://www.kaggle.com/fayomi/advertising

**Existing Methods:Logistic, SVM, gaussian navie bayes, , Decision Tree**

**Proposed Method:randomForest**

**Support Vector Machine**

“Support Vector Machine” (SVM) is a supervised [machine learning algorithm](https://courses.analyticsvidhya.com/courses/introduction-to-data-science-2?utm_source=blog&utm_medium=understandingsupportvectormachinearticle) which can be used for both classification or regression challenges. However,  it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well (look at the below snapshot).

[](https://www.analyticsvidhya.com/wp-content/uploads/2015/10/SVM_1.png)

Support Vectors are simply the co-ordinates of individual observation. The SVM classifier is a frontier which best segregates the two classes (hyper-plane/ line).

**Logistic regression**

Logistic regression is named for the function used at the core of the method, the logistic function.

The [logistic function](https://en.wikipedia.org/wiki/Logistic_function), also called the sigmoid function was developed by statisticians to describe properties of population growth in ecology, rising quickly and maxing out at the carrying capacity of the environment. It’s an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits.

1 / (1 + e^-value)

Where e is the [base of the natural logarithms](https://en.wikipedia.org/wiki/E_(mathematical_constant)) (Euler’s number or the EXP() function in your spreadsheet) and value is the actual numerical value that you want to transform. Below is a plot of the numbers between -5 and 5 transformed into the range 0 and 1 using the logistic function.



Logistic Function

Now that we know what the logistic function is, let’s see how it is used in logistic regression.

# Naive Bayes Classifiers

Naive Bayes is a classification algorithm for binary (two-class) and multi-class classification problems. The technique is easiest to understand when described using binary or categorical input values.

It is called naive Bayes or idiot Bayes because the calculation of the probabilities for each hypothesis are simplified to make their calculation tractable. Rather than attempting to calculate the values of each attribute value P(d1, d2, d3|h), they are assumed to be conditionally independent given the target value and calculated as P(d1|h) \* P(d2|H) and so on.

This is a very strong assumption that is most unlikely in real data, i.e. that the attributes do not interact. Nevertheless, the approach performs surprisingly well on data where this assumption does not hold.

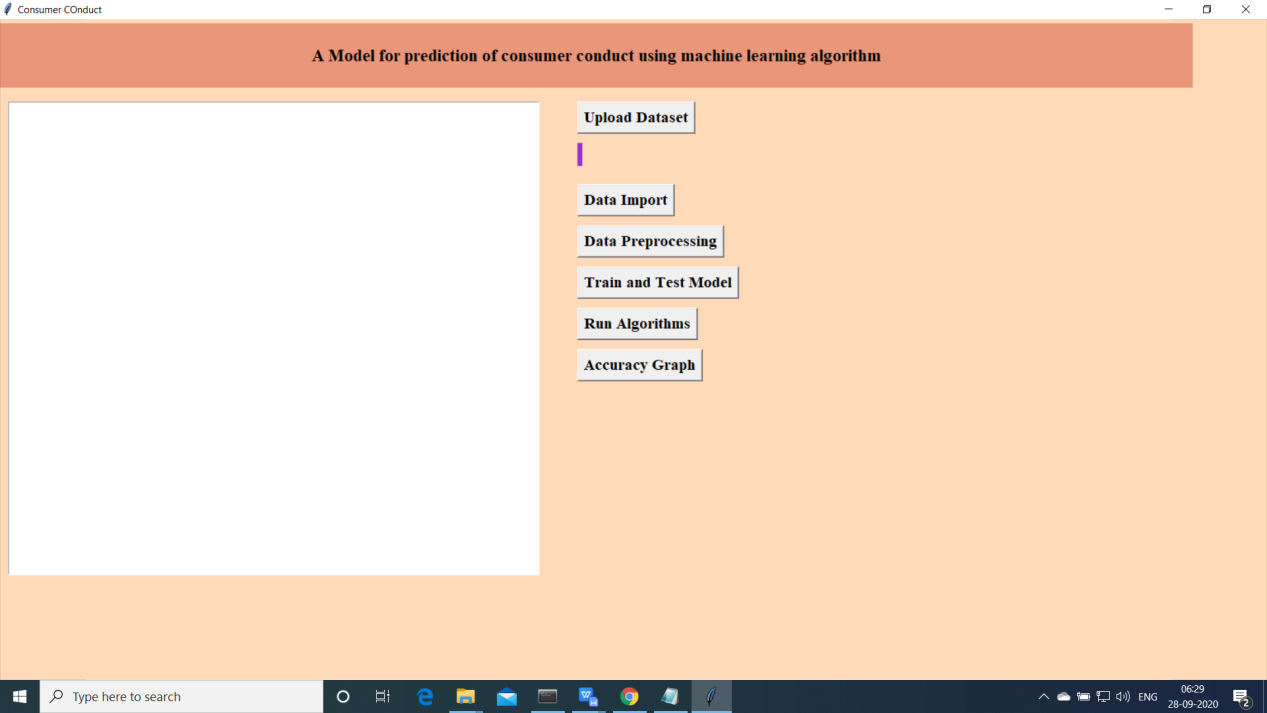
# Random Forest Algorithm

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of **ensemble learning,** which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

As the name suggests, **"Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset."** Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

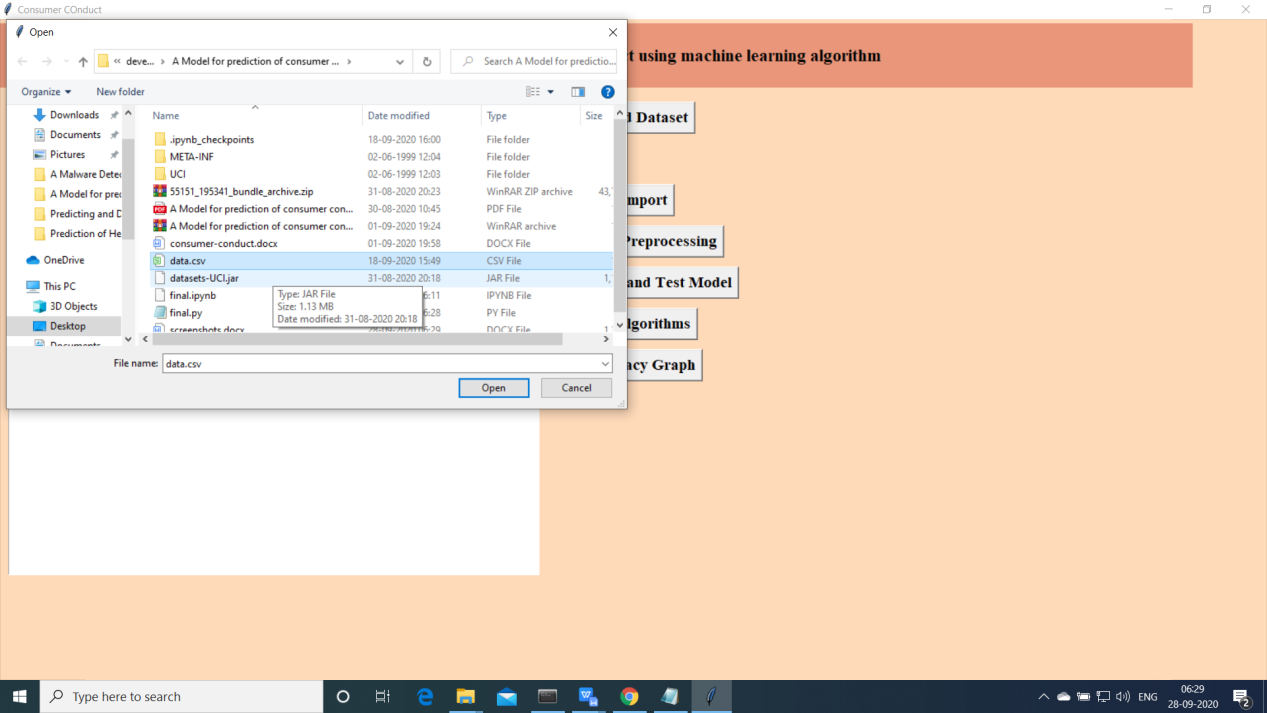
**Execution Steps:**

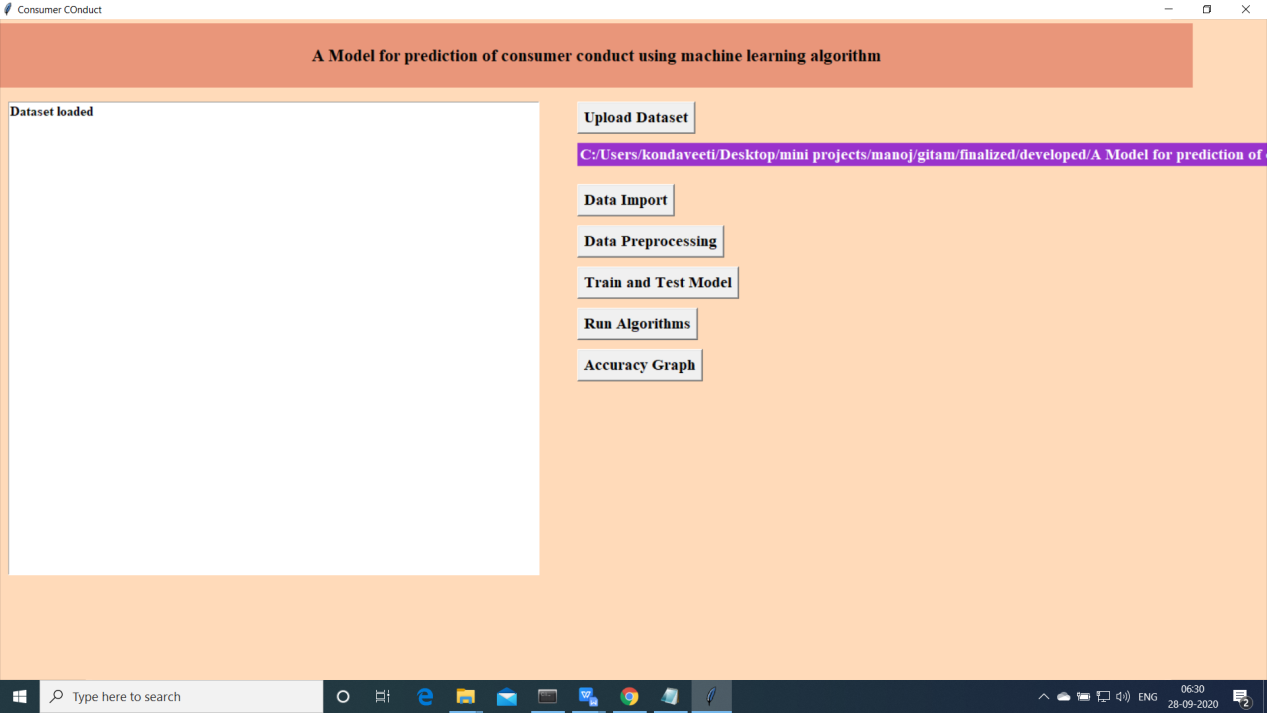
1. Open anaconda Prompt
2. Conda activate tf
3. Goto Project Directory
4. Python final.py



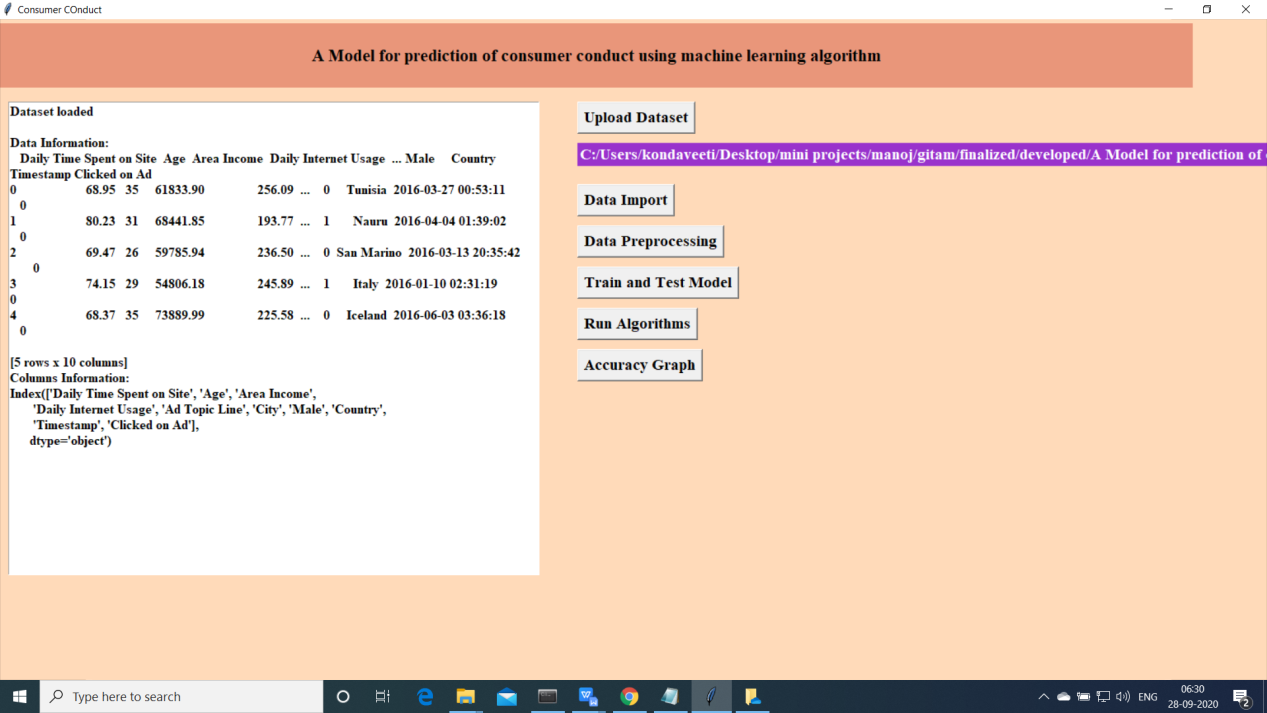
Above screen will be opened.

1. Now click on “Upload data ”



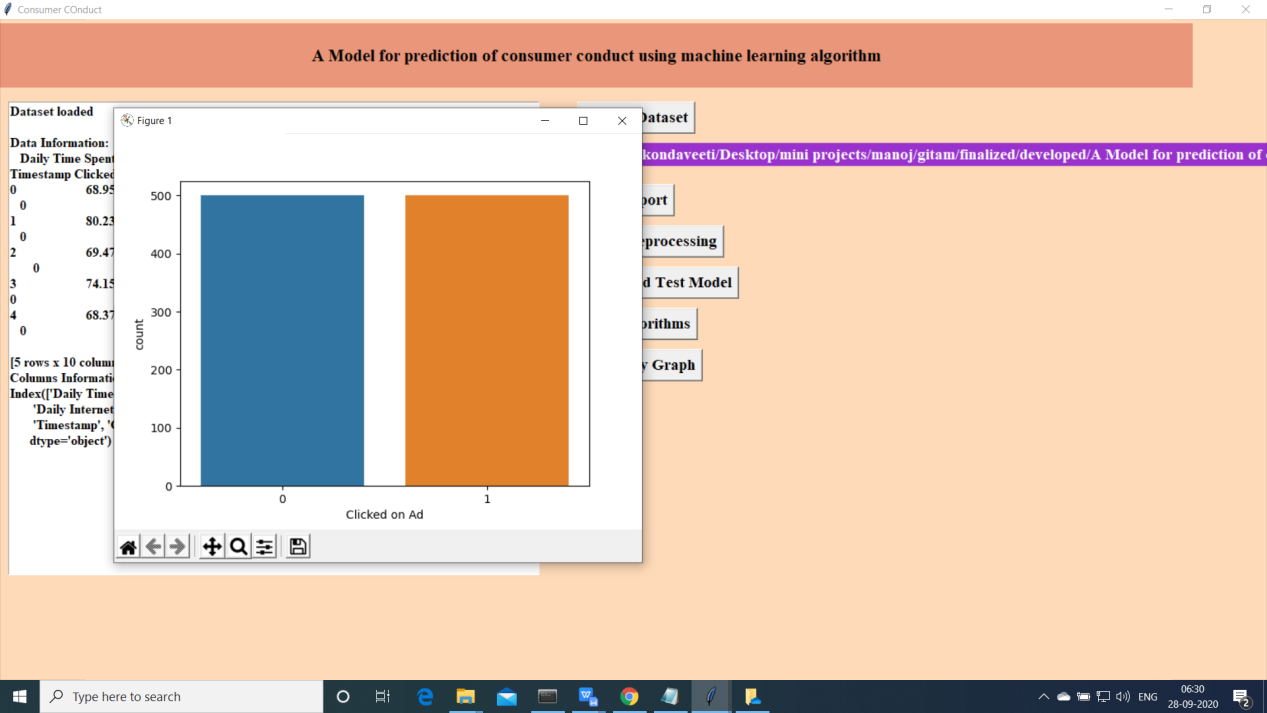


1. Import the data

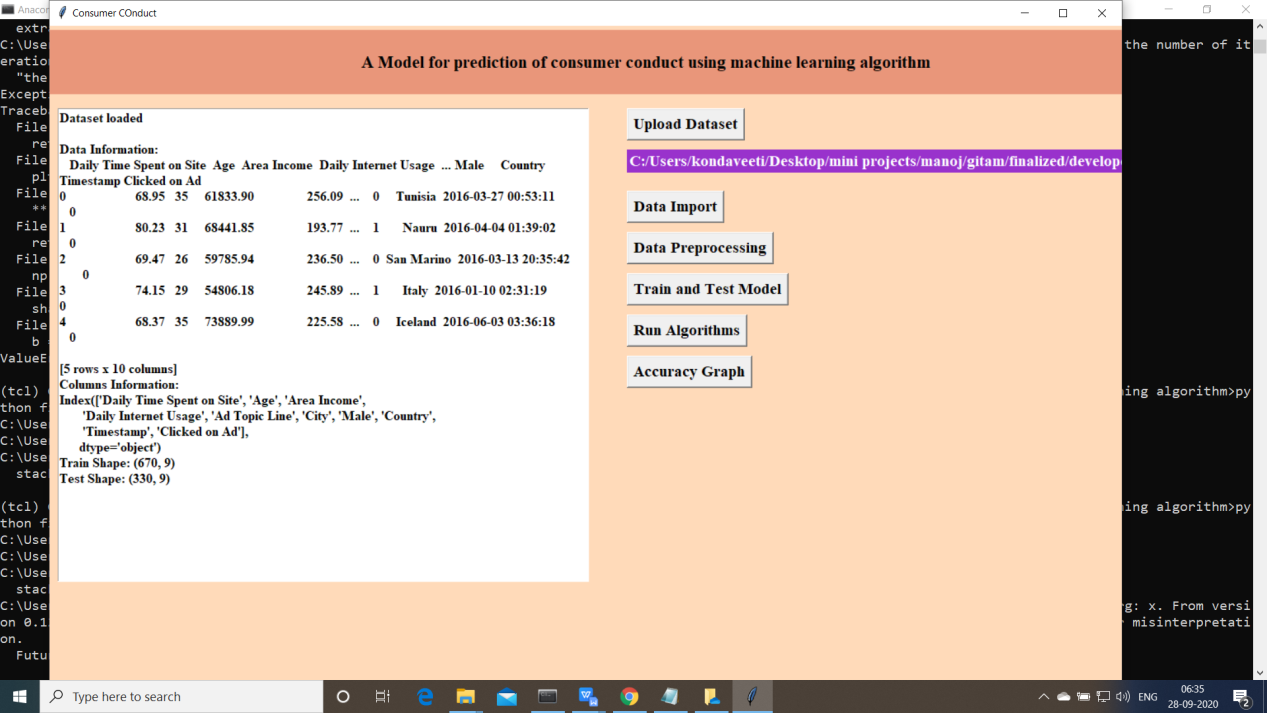


Upload the data and read the basic data information will be shown on the screen

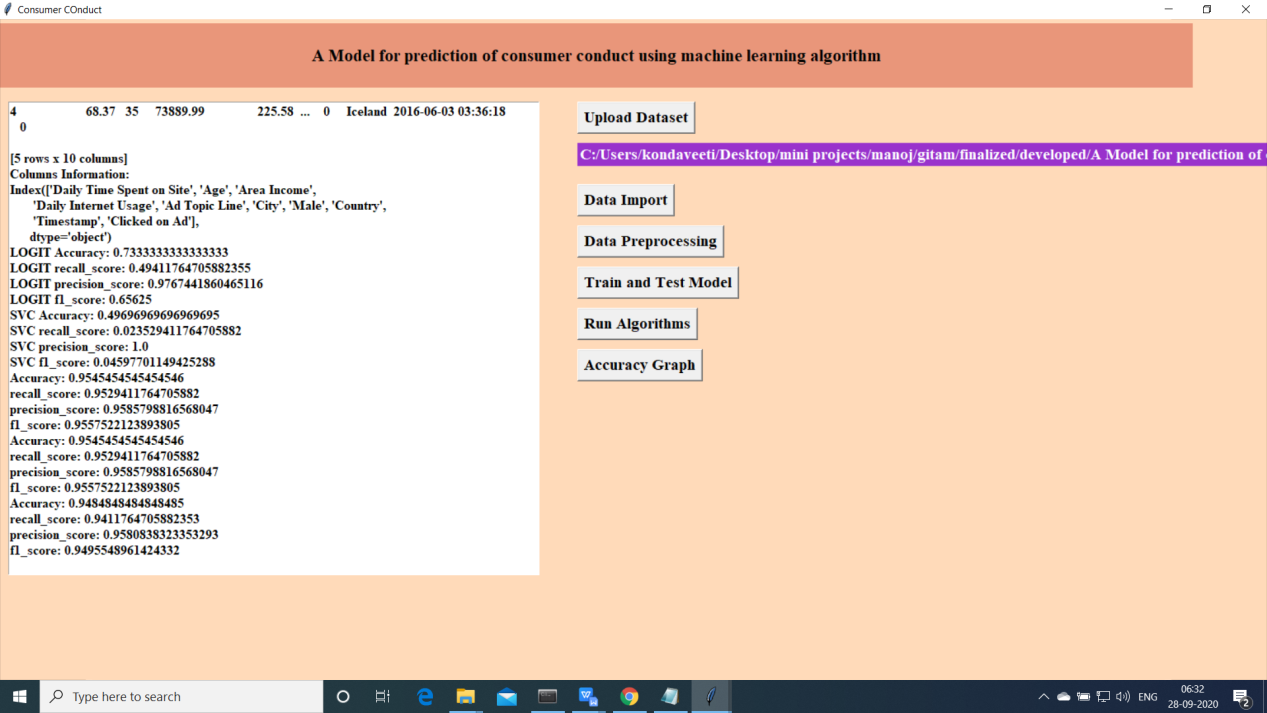
1. Data Preprocessing



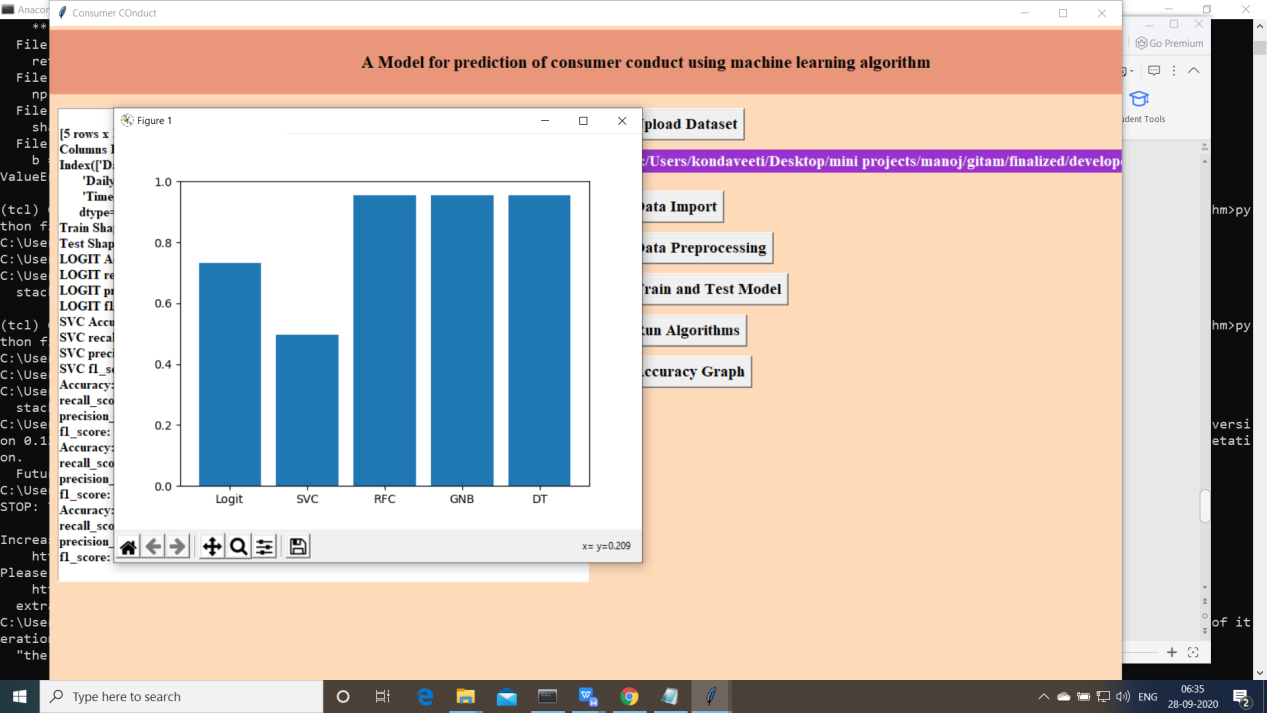
1. Now click on “Train and Test model”. split the data into train and test and traain will be used for training and to tets the performace we are using test data



1. Now click on “Run Algoruimns”. Mentioned algorithms will be run on the data



1. Accuracy Comparision for all the models



Navie bayes algorithm is performed better

Extension is Random Forest and perfromed well compare to other algorithms

Random Forest got an accuracy of 95.5%