**TWITTER SENTIMENT ANALYSIS.**

**Introduction**

This project aims to extract tweets using different hashtags of leaders from twitter and categorise each tweet on the basis of sentiment scores (Positive , Negative ,Neutral) to find out the most influential leader amongst Narendra Modi, Donald Trump, Jacinda and predict its accuracy using Naïve Bayes modelling approach.

R Libraries that would be needed to achieve this task:

1. twitteR

2. Ggplot2

3. Lattice

4.caret

5.tm

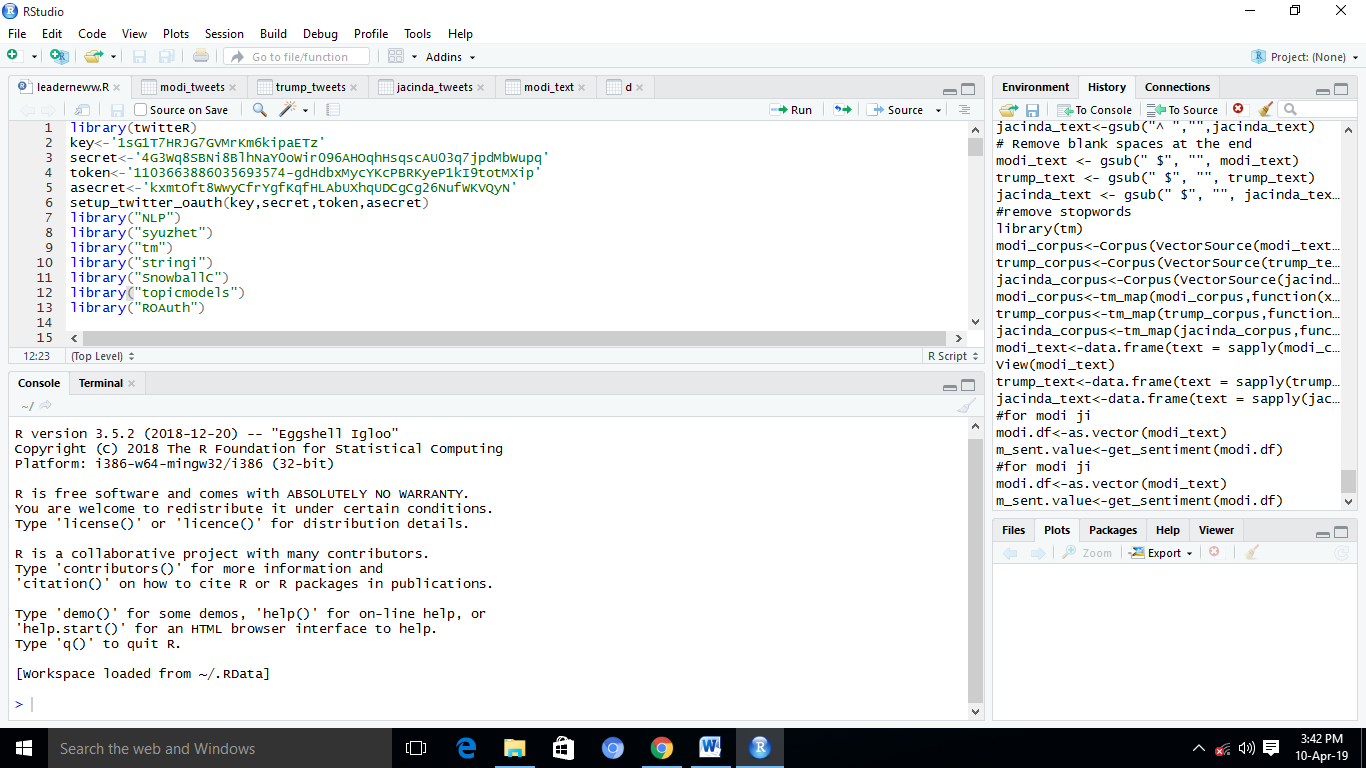
6.syuzhet

**Step 1: Gathering the data**

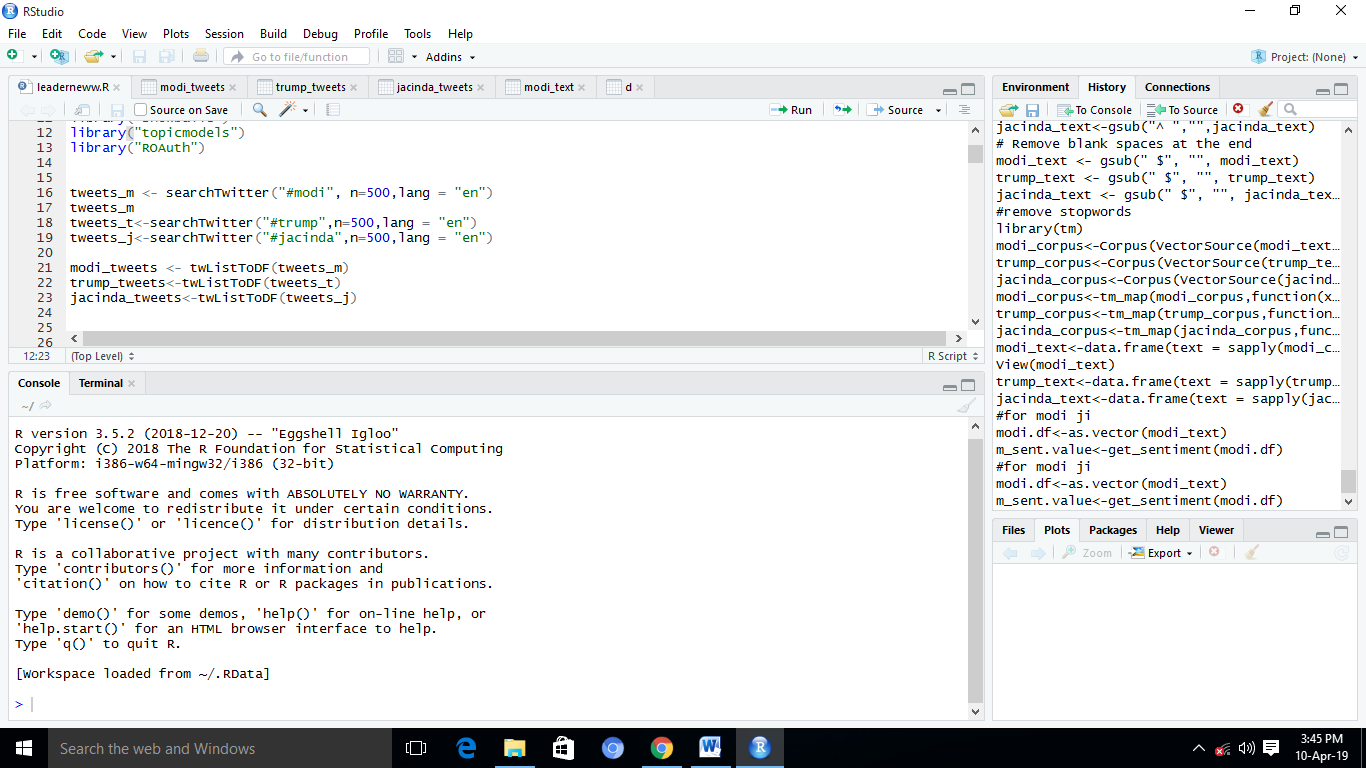
1. Crawl Tweets Against Hash Tags

To have access to the Twitter API, one need to login the Twitter Developer website and create an application. Enter your desired Application Name, Description and your website address making sure to enter the full address including the http://. You can leave the callback URL empty.

2. After registering, create an access token and grab your application’s Consumer Key, Access token and Access token secret from Keys and Access Tokens tab.



3.Extracting tweets using different hashtags.



**Step 2: Pre-processing of data**

we first need to remove noise and preprocess tweets by using the following steps:

1.Lower Case - Convert the tweets to lower case.

2.URLs - Eliminate all of these URLs via regular expression matching

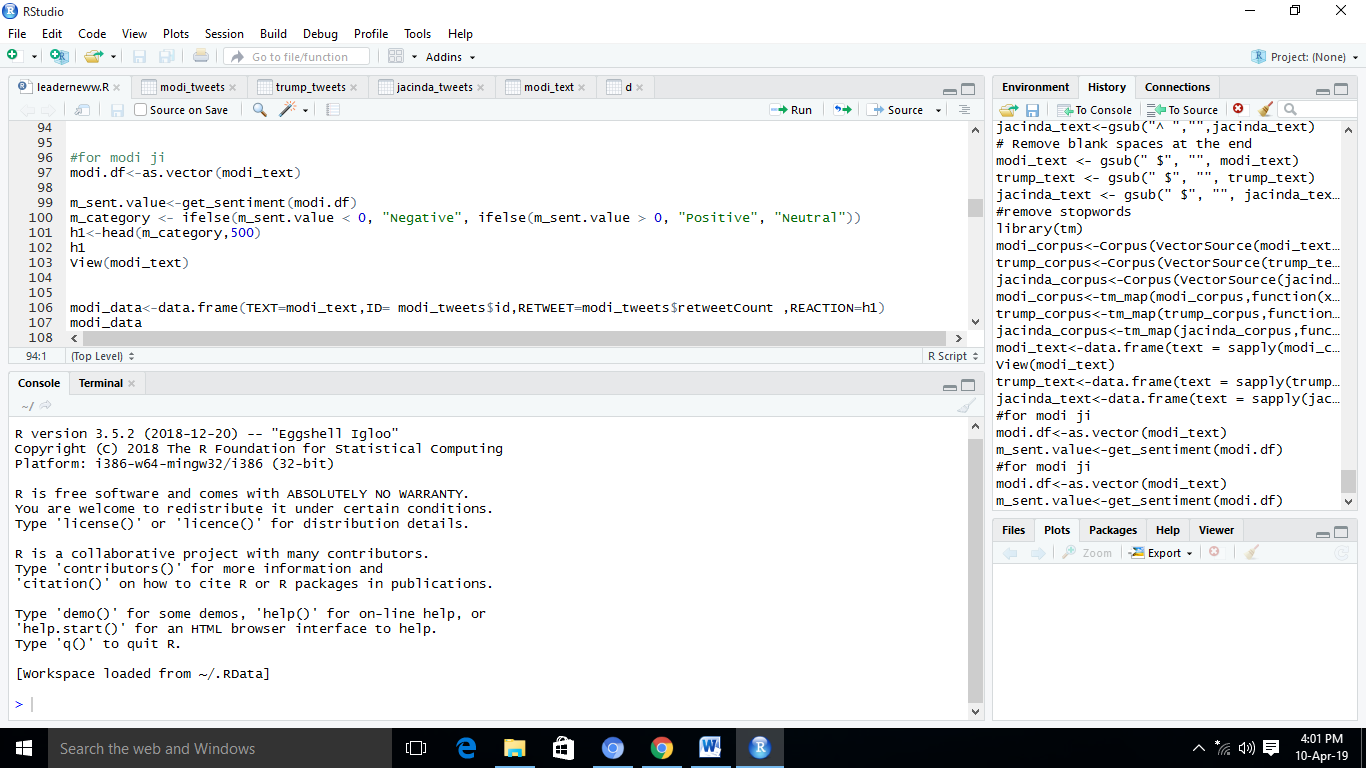
3.@username - Remove "@username" via regex matching.

4.Punctuations and additional white spaces - remove punctuation at the start and ending of the tweets, e.g: ' the day is beautiful! ' replaced with 'the day is beautiful'. We also replace multiple whitespaces with a single whitespace.

5.stop words-Remove the stop words like but ,then ,there (common English words) using corpus function defined in “tm” package.

**Step 3: Sentiment Analysis**

In this step we examine each tweets’s sentiment score using get\_sentiment() method from syuzhet package and categorise score less than 0 as Negative ,greater than 0 as positive and equal to zero as neutral.



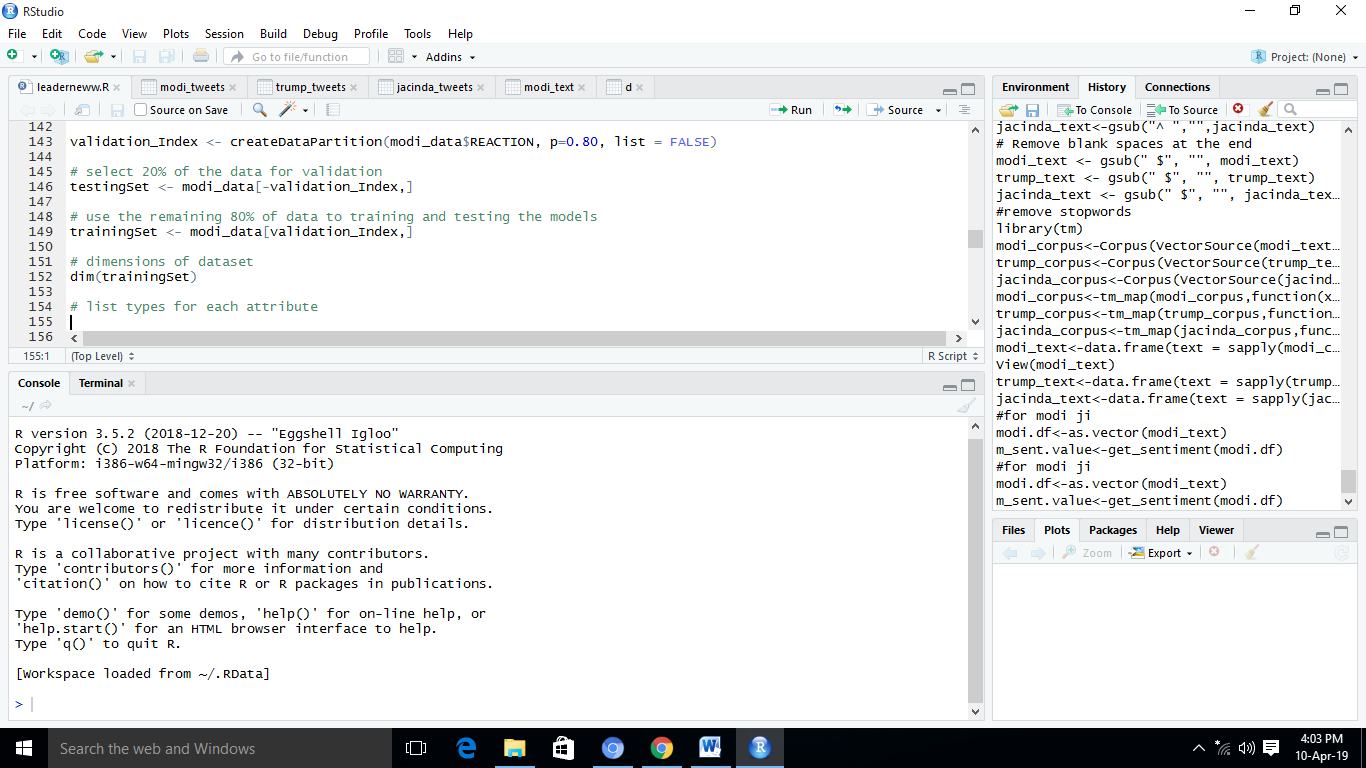
**Step 4: Training and Testing the model on data**

Here we are applying Naïve Bayes Model.

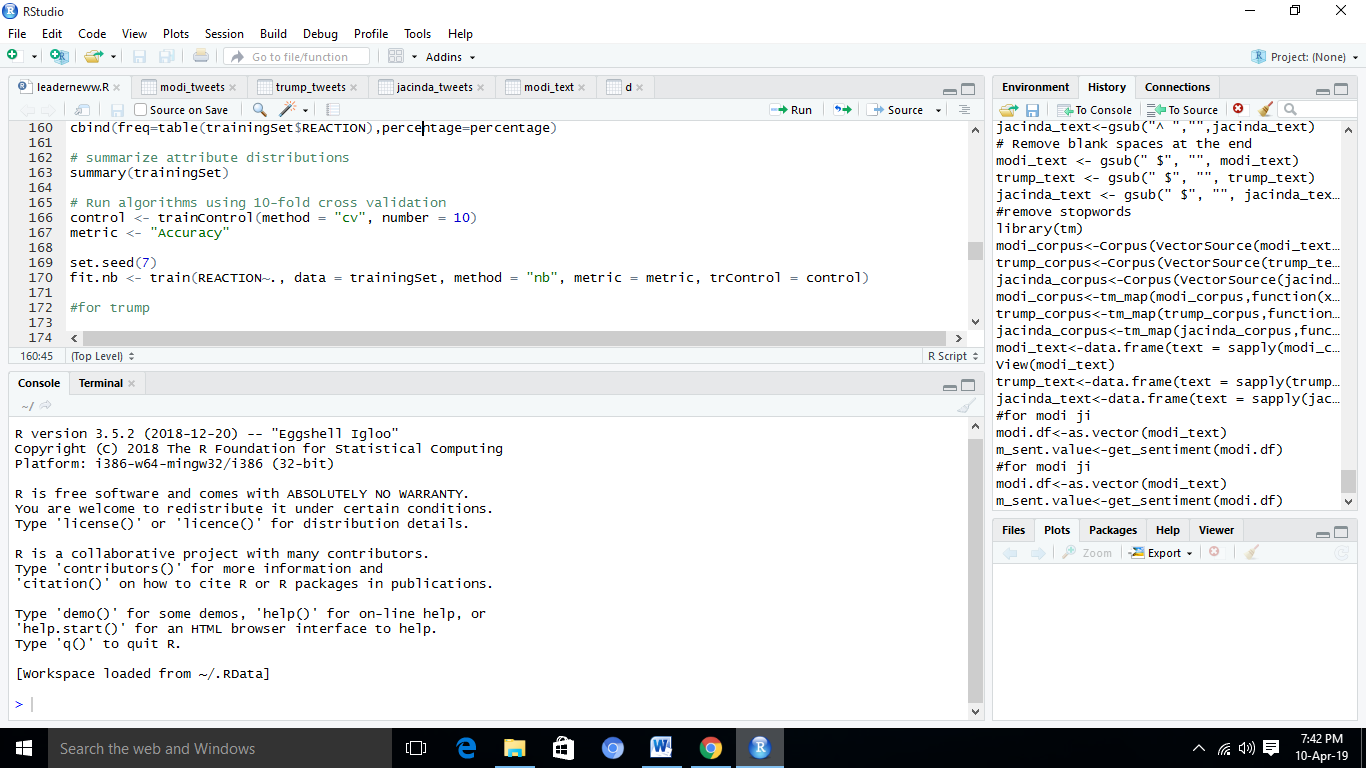
For training a model we initially split the model into 2 sections which are ‘**Training data**’ and ‘**Testing data**’.

First, we train the classifier using ‘**training data set**’, tune the parameters using ‘**validation index**’ and then test the performance of our classifier on unseen ‘**test data set**’.

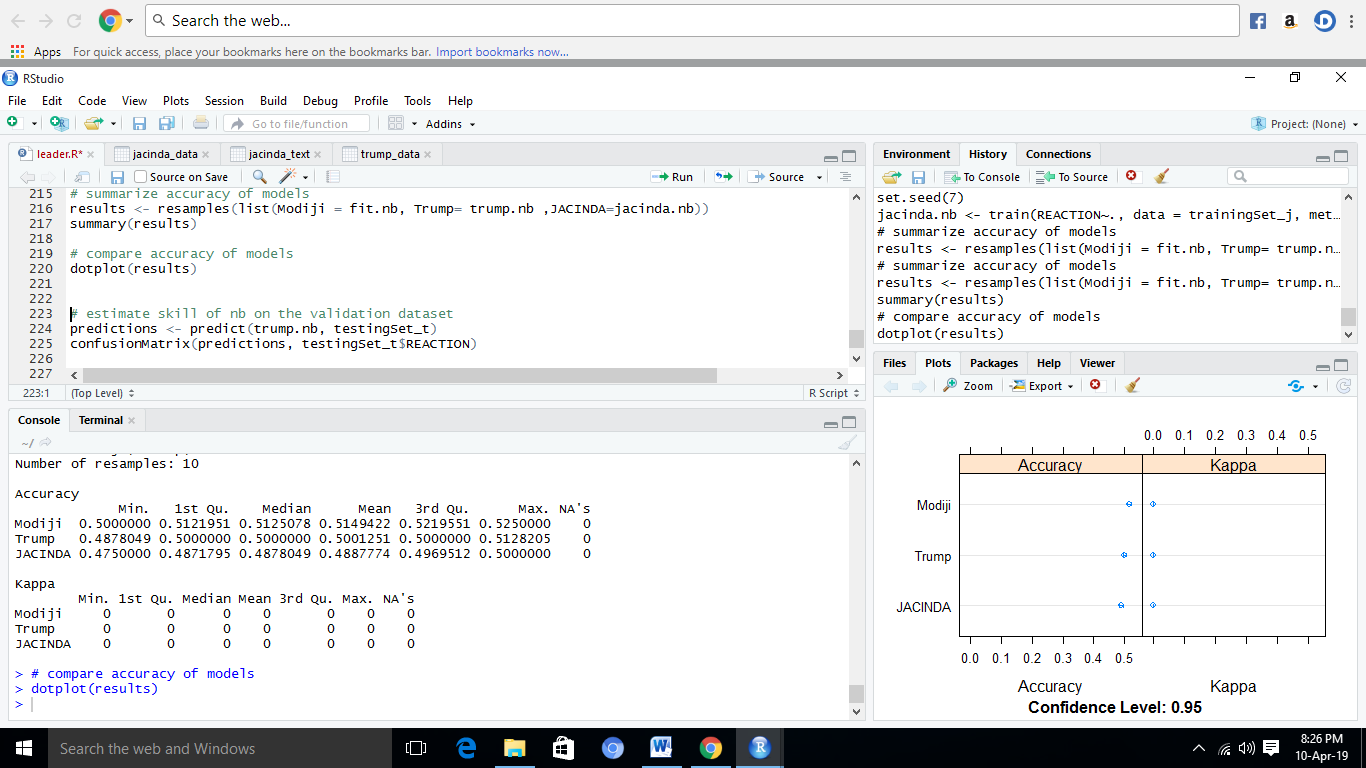
For this step, we are going to split our dataset into 4:1 ratio i.e. 80% training dataset and 20% testing dataset.



Applying Naïve Bayes model for every leader’s training data set, to compare the accuracy.



Finally, resample all the model result and plot the accuracy



By this we can conclude that Narendra Modi is concluding to be the most influential amongst Trump and Jacinda.