Final Project Report – US Election 2020 Sentiment Analysis

**Introduction:**

Forecasting the Presidential Elections has become a trend in Academia. Recently, US 2020 Elections were held, and the main contestants were Joe Biden from Democratic party and Donald Trump from Republican Party. I wanted to do Sentiment Analysis on Twitter data to get results of who will be the possible winner of US 2020 Elections.

**Problem Statement:**

1. Is Trump liked more or Biden?
2. Trump vs Biden Tweet count in India?
3. Trump or Biden, who has more positive tweets?

**Related Work:**

In last decade, there was a steady increase in data mining and finding out the opinions in text data very much. This particular opinion data from Twitter is used for stock market prediction and also for analyzing Social movements across United states and many other countries. Tumasjan (2010) found tweet volume about the political parties to be a good predictor for the outcome of the 2009 German election, while Choy et al. (2011) failed to predict with Twitter sentiment the ranking of the four candidates in Singapore’s 2011 presidential election. Now, I am using data from Kaggle, which is very apt for this project and would like to get good results from this data for deep understanding of the cultural and political practices at work through the use of Twitter data.

**Methods:**

Tools used -

I worked with Python Programming Language in IDE Google Colab. I used Pandas for Reading the Data set and also for Data Transformation, NumPy also for data Transformation. Matplotlib for Data Visualizations.

Load Data –

First, I loaded the Two data sets “Trump Data”, “Biden Data” into two different data frames.



Data Transformation –

I drop () function to drop few columns in the data set as they were not useful for my analysis.

I used loc [] method to get only the data of Trump in Trump Dataset and Biden in Biden Data set.

I used Concat () method to mix both the data frames into one data frame.

I used Sort\_values () method to sort the data frame on the column “created\_at” in ascending order.

A picture containing graphical user interface

Description automatically generated

Data Visualization –

Xtick () function is used in bar plot to keep labels in x axis. Here I used Xticks for labeling Donald Trump and Biden on x-axis.

Used Bar plot from Matplotlib for this.

Here, I used default blue color as we don’t have many variables to compare to.

Graphical user interface, text, application

Description automatically generated

Chart, bar chart

Description automatically generated

Fig:

Tweets count of each candidate is visualized through pie chart. Here I defined a function to calculate the pie chart with input of country name.



Chart, pie chart

Description automatically generated

Orange and Blue colors differentiate each candidate in pie chart.

Data Cleaning –

In order to do Sentiment Analysis, we have to remove the noise in the tweets like punctuation, numbers, stop words, spaces and convert them to lower case.

Graphical user interface, text, application, email

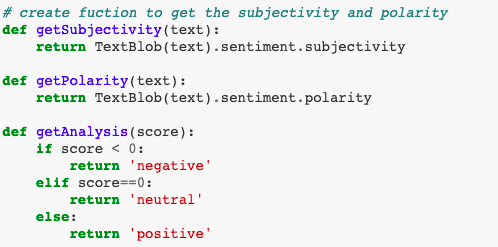
Description automatically generated

Dropped the Null Values using function dropna ()

Sentiment Analysis:

I have created functions to calculate the subjectivity score, polarity score and analysis based on score to find the sentiment of the tweets.

Here I had used NLTK, TextBlob Libraries for finding the scores.

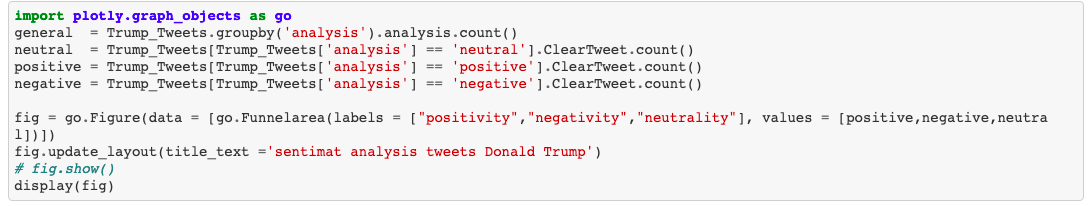


Imported Counter from Collections library to count the values of each sentiment positive, negative and neutral.

Text

Description automatically generated

I imported graph\_objects from plotly library to make a funnel area chart.



Funnel Area is used to show data in different stages.

Chart, funnel chart

Description automatically generated

Fig:

Chart, funnel chart

Description automatically generated

Fig:

We can clearly see from the chart that Trump has 39.8%, Joe Biden has 45.5% positive tweets.

**Results:**

From the sentiment analysis, we found out that Joe Biden has more chance of winning the US 2020 Election even though the data had more likes for Donald Trump.

**Discussion:**

From the Visualizations people can infer that Donald Trump has more likes than his counterpart from the bar chart. The Pie chart tells that In India, there was more discussion about Joe Biden than Donald Trump. From Sentiment Analysis, through funnel area graph chart we found that Donald Trump has fewer positive sentiments and more negative sentiments in the tweets than Joe Biden.

**Future Work:**

We should have ideally same number of tweets from all the states, well that is quiet impossible to have that but we can try by Distributing the data Uniformly. We can also use Supervised and Un Supervised Machine Learning Algorithms predict the sentiments with better accuracies.

**References:**

Wang, Hao, et al. *A System for Real-Time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle*.

KA, A. (2020, November 12). Data Visualization: US Election 2020. Retrieved from Medium website:

<https://medium.com/swlh/data-visualization-us-election-2020-a3d3531b611c>

‌