

Sentiment Analysis of Twitter dataset

Tweets comparison of Joe Biden & Donald Trump during 2020 Presidential Elections.

1

Shubham Lokhande. Author, *Systems & Enterprises*, Stevens Institute of Technology, <mailto:slokhan1@stevens.edu>

ABSTRACT

Basic idea of this project is to get clear idea about the tweets by Joe Biden & Donald Trump during the presidential elections of 2020. The aim of the project is to analyze those tweets to find the sentiment and conclude that whose tweets have more positive impact.

I. INTRODUCTION

Sentiment analysis is the process of detecting positive or negative sentiment in text. It's often used by businesses to detect sentiment in social data, gauge brand reputation, and understand customers.

Sentiment analysis focuses on polarity (positive, negative, neutral) but also on feelings and emotions (angry, happy, sad, etc.), urgency (urgent, not urgent) and even intentions (interested v. not interested).

In this project same sentiment analysis was used to find the polarity of tweets. Python code was scripted accordingly to find the sentiment and the result of elections. In addition to the analysis code also helps to find most tweeted words by both.

II. BACKGROUND

Towards the end of course in Informatics for Engineering Management a mandatory project helps in better understanding of the concepts we learned in the subject. This project emphasizes on the teachings of course work performed in this subject taught by Dr. Carlo Lipizzi.

The idea of the project came after the Presidential elections of 2020. As elections happened recently and new government started working, I felt curious about the twitter war between the competing parties. This idea eventually generated a spark to find the positivity and negativity in tweets by them.

Many people around the world follow the electing party heads Joe Biden and Donald Trump. This analysis eventually helped to find polarity scores of positive, neutral, and negative tweets posted by them.

III. METHODOLOGY

Python scripts are the core of this project.

- Twitter dataset of the tweets by Biden and Trump in the form of csv files was obtained. [1]
- This data set was an opensource database used for different analysis.
- PyCharm was the tool used to run program.

```
import csv
from collections import Counter
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
import matplotlib.pyplot as plt
from wordcloud import WordCloud
```

Fig 1. Imported libraries

- After importing csv library, Csv files were opened in python.
- Two empty lists were made to store the data.
- Required row data was appended in the following list
- A thorough cleaning of the data was performed by defining a function to strip and split the data, to lowercase the data and to only have alphabets.
- Data was made to run through for loops.
- Stop words text file was used to remove unwanted and impact less words from the data.
- The cleaned data was stored in a new list.
- Counter library was imported to count most common words in a new data set.

```
a1 = sum(clean_biden_words,[])
Counter = Counter(a1)
most_occur1 = Counter.most_common(10)
print("The 10 most common tweeted words during election campaign by Joe Biden are:")
print(most_occur1)
print()
```

Fig 2. Common words using Counter

- Both files go through same processes.
- After importing Sentiment Intensity Analyzer library, an empty tuple was created to hold sentiment of the sentence.
- A function of sentiment of sentence was defined to analyze each sentence.
- This helped in printing the positive and negative sentiments of the database.
- Using the same library, polarity score was calculated.

```

SID_obj1 = SentimentIntensityAnalyzer()
polarity_dict1 = SID_obj1.polarity_scores(clean_biden_data)
print("Polarity of Tweets is as follows :")
print()
print("Polarity percentage of tweets is", polarity_dict1['neg'] * 100, "% Negative")
print("Polarity percentage of tweets is", polarity_dict1['neu'] * 100, "% Neutral")
print("Polarity percentage of tweets is", polarity_dict1['pos'] * 100, "% Positive")
print()
print("Overall polarity percentage of Joe Biden's tweets is", end=" : ")

```

Fig 3. Polarity of sentences.

- q) Same Analysis was performed for both the datasets.
- r) A Matplotlib and word cloud library was installed to plot a word cloud of the tweets by Biden and Trump.
- s) As a final result, pie chart was plotted between their election result. [2]

IV. RESULTS

- a) After running a script in python, following were the results obtained:

The 10 most common tweeted words during election campaign by Joe Biden are:

['vote', 82), ('president', 81), ('trump', 75), ('make', 54), ('donald', 52), ('election', 51), ('head', 46), ('day', 45), ('today', 36), ('tune', 35)]

The 10 most common tweeted words during election campaign by Donald Trump are:

['biden', 168), ('great', 147), ('thank', 133), ('joe', 102), ('vote', 89), ('election', 86), ('big', 85), ('people', 76), ('fake', 58), ('trump', 54)]

The top 3 positive sentiments on twitter by Joe Biden are:

['jill wish happy festival join hoping future filled peace prosperity hope festivities lead bountiful harvest year',

'receiving medal freedom friend great proud dedicated life serving country',

'middle president trump trying eliminate obamacare increase health care costs rushing supreme court justice help make health care']

The top 3 negative sentiments on twitter by Joe Biden are:

['domestic violence awareness recommit ensuring domestic violence treated crime ensuring survivors receive support spent entire career fighting domestic abuse continue fight',

'president trump plan address quit wants grow numb horrors death toll afford years failed',

'donald trump distract fact failed americans died million unemployment small businesses risk permanent closure let']

Polarity of Tweets is as follows:

Polarity of Tweets is as follows :

*Polarity percentage of tweets is 9.9 % Negative
Polarity percentage of tweets is 66.10000000000001 % Neutral
Polarity percentage of tweets is 24.0 % Positive*

Overall polarity percentage of Joe Biden's tweets is: Positive

The top 3 positive sentiments on twitter by Donald Trump are:

*['huge win today united states peace sudan agreed peace normalization',
'supreme court shows great wisdom american people win important case',
'days going win great state going win years']*

The top 3 negative sentiments on twitter by Donald Trump are:

*['prayers people vienna vile act terrorism evil attacks',
'killed fellow citizens illegal gangs democrats ruined',
'vote biden harris vote ban ban co*

Polarity of tweets is as follows:

Polarity of tweets is as follows :

*Polarity percentage of tweets is 15.0 % Negative
Polarity percentage of tweets is 65.4 % Neutral
Polarity percentage of tweets is 19.6 % Positive*

Overall polarity percentage of Donald Trump's tweets is: Positive

```
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
sentence_sentiment_dict1={}
def sentiment(sentence):
    sid_obj = SentimentIntensityAnalyzer()
    sentiment_dict = sid_obj.polarity_scores(sentence)
    sentence_sentiment_dict1[sentence]=sentiment_dict['compound']
for sentence in clean_biden_data:
    sentiment(sentence)
print('The top 3 positive sentiments on twitter by Joe Biden are:')
sorted_positive1 = sorted(sentence_sentiment_dict1.items(), key=lambda kv:
kv[1],reverse=True)
positive3 = sorted_positive1[0:3]
pos=[]
for z in positive3:
    pos.append(z[0])
print(pos)
print('The top 3 negative sentiments on twitter by Joe Biden are:')
sorted_negative1 = sorted(sentence_sentiment_dict1.items(), key=lambda kv:
kv[1])
negative3 = sorted_negative1[0:3]
neg=[]
for z in negative3:
    neg.append(z[0])
print(neg)
```