SENTIMENT ANALYSIS ON POLITICAL TWEETS BY PRESIDENTIAL CANDIDATES ON US ELECTION

SUBMITTED BY:

KRUNAL PATIL (3327)
SHAIK AFTAB (3345)
SURAJ KANYAL (3351)
SWAPNIL SINGH (3352)

Contents

Sentiment Analysis

- Sentiment analysis is contextual mining of text which identifies and extracts subjective information in source material, and helping a business to understand the social sentiment of their brand, product or service while monitoring online conversations.
- It is the most common text classification tool that analyses an incoming message and tells whether the underlying sentiment is positive, negative or neutral.

ABSTRACT

- Sentiment Analysis is a Natural Language Processing task that aims to obtain anyone's feelings expressed in positive or negative comments, questions and requests, by analysing a large series of documents.

- We want to hear others' opinions while making a decision.

- Second, it presents many challenging research problems, which had never been attempted. Reason was that there was little opinionated text in digital forms. It is thus no surprise that the inception and the rapid growth of the field coincide with those of the social media on the Web.

- Research has also spread outside of computer science to management sciences and social sciences due to its importance to business and society as a whole.

Motivation

- The Seeker’s Perspective
- The Giver’s Perspective

Applications of Sentiment Analysis

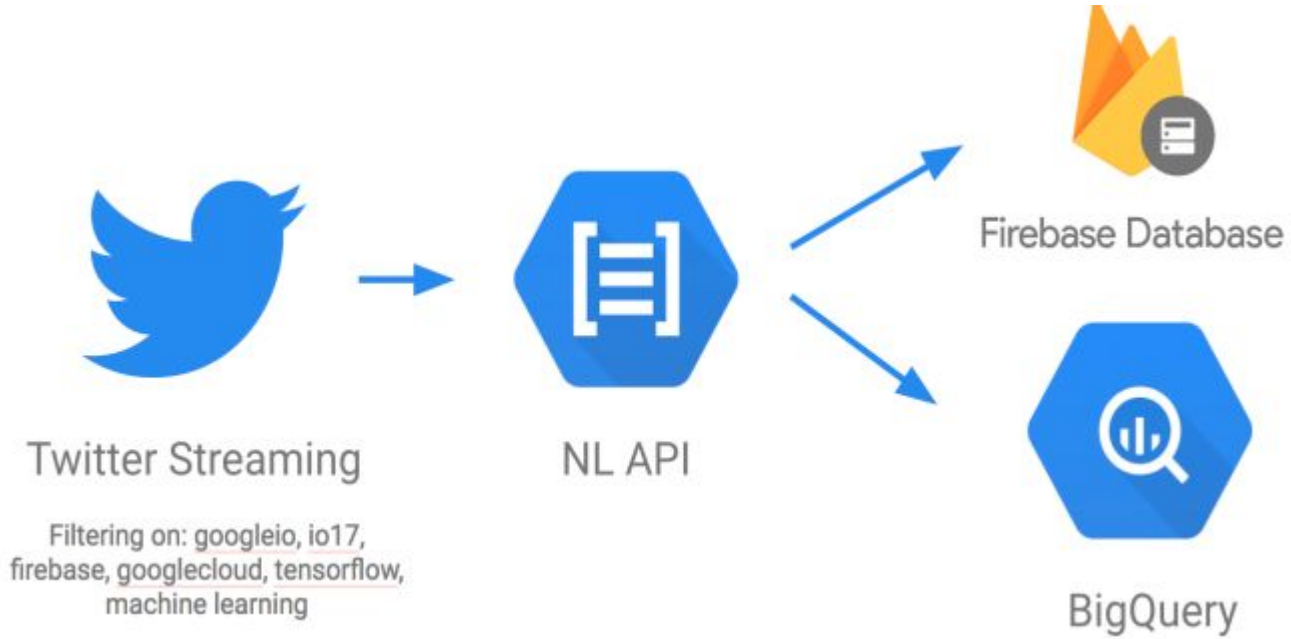
a. Review-related Websites
b. Automatic summarization of user reviews
c. Sentiment Analysis can also be used in trend prediction.
d. Applications in smart homes

Challenges

1. Identifying subjective portions of text. The same word can be treated as subjective
2. Associating sentiment with specific keywords
3. Domain dependence
4. Sarcasm Detection
5. Thwarted expressions
6. Indirect negation of sentiment
7. Order dependence
8. Entity Recognition
9. Identifying opinion holders

LITERATURE SURVEY

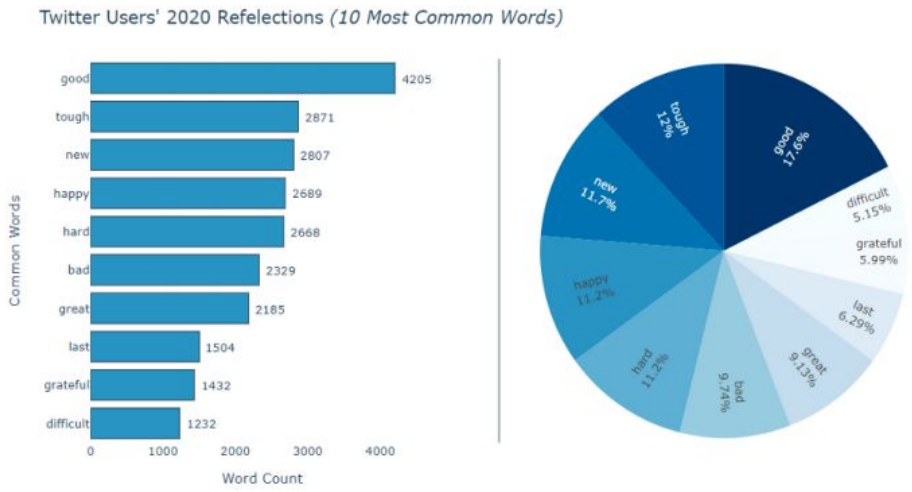
1. Project Strategy
2. Tweets Mining
3. Technology
4. Tweets Processing
5. Machine Learning

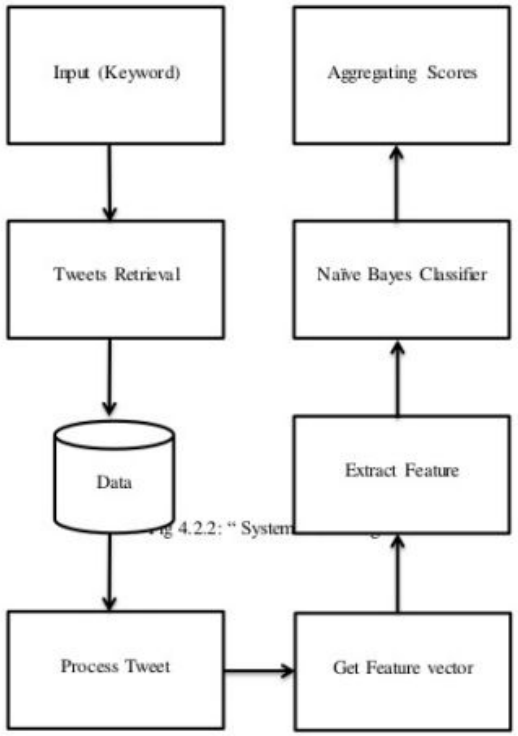


PROJECT DESIGN

Collect all the tweets and then run our algorithm to judge which person's thoughts are more effective in public. So, there are mainly three steps we will be working on,

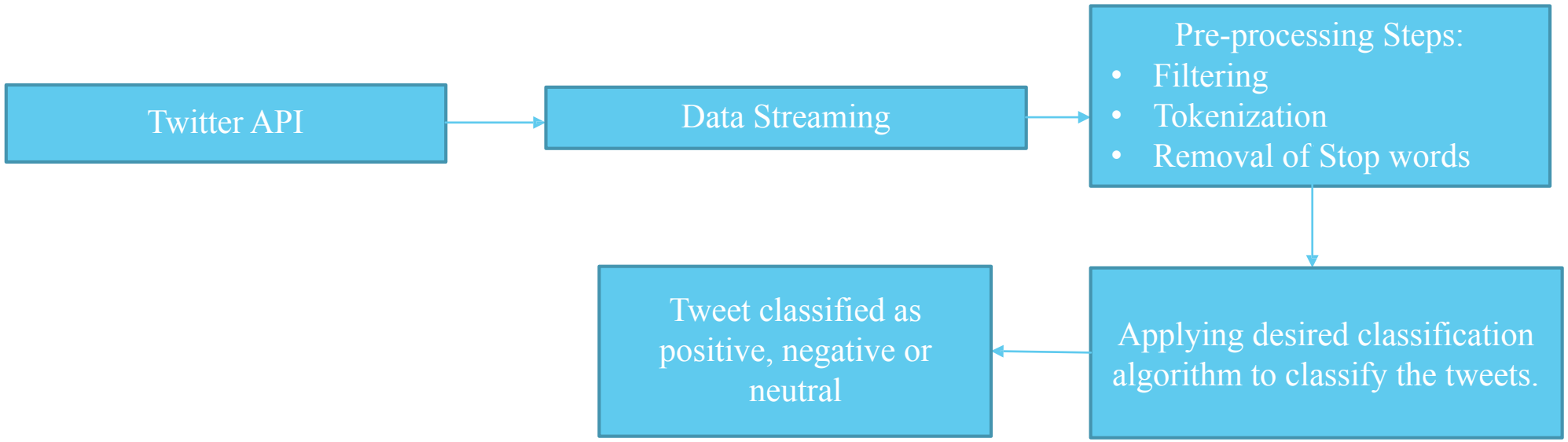
1. **TOOLS**
2. **DATASET CREATION**
3. **DATA ANALYSIS**
4. **DATA VISUALIZATION**





Flowchart

The following flowchart shall be followed while implementing the undertaken project.

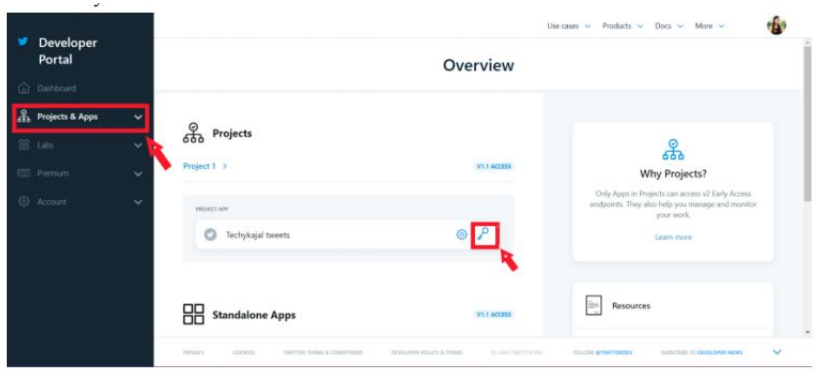


Twitter API

- An application program interface (**API**) is a set of protocols and tools for building software applications. Basically, an **API** specifies how software components should interact. Additionally, APIs **are** used when programming graphical user interface (GUI) components.
- The **Twitter API** is simply a set of URLs that take parameters. They URLs let you access many features of **Twitter**, such as posting a tweet or finding tweets that contain a word.
- Tweepy is open-sourced library, hosted on GitHub and enables Python to communicate with Twitter platform and use its API.

DATA CREATION

STEPS




```
USER1_TWITTER_USERNAME = "narendramodi"
USER2_TWITTER_USERNAME = "MamataOfficial"
```

```
USER1_POSTS = api.user_timeline(
                screen_name = USER1_TWITTER_USERNAME,
                count = 1000,
                lang = "en",
                tweet_mode = "extended"
            )

USER2_POSTS = api.user_timeline(
                screen_name = USER2_TWITTER_USERNAME,
                count = 1000,
                lang = "en",
                tweet_mode = "extended"
            )
```

	Tweets	subjectivity	polarity
0	वरिष्ठ पत्रकार शेष नारायण सिंह जी का निधन अत्य...	0.000000	0.000000
1	'Jal', 'Thal' and 'Nabh'...our armed forces ha...	0.000000	0.000000
2	: Fighting the Invisible Enemy: MoD's Response...	0.400000	0.200000
3	ગુજરાતના પૂર્વ નાણામંત્રી શ્રી પ્રતાપભાઈ શાહના...	0.000000	0.000000
4	Reviewed various aspects of the COVID-19 respo...	0.500000	0.000000
...	...	...	...
195	: Handing over of a Fast Patrol Vessel to Seyc...	0.500000	0.200000
196	: Joint e-inauguration of the new Magistrates'...	0.454545	0.136364
197	Speaking at the meeting with Chief Ministers.	0.000000	0.000000
198	Boosting friendship with Seychelles.	0.000000	0.000000
199	India will pay a fitting tribute to the great ...	0.625000	0.650000

Data Streaming

- Data streaming is the process of transferring a stream of data from one place to another, to a sender and recipient or through some network trajectory. Data streaming is applied in multiple ways with various protocols and tools that help provide security, efficient delivery and other data results.
- Twitter presents two kinds of APIs to extract the tweets:

(1) **Search API** is used for dumping old tweets . The training dataset is built for sentiment classification.

(2) **Streaming API** used for dumping live Sentiments . The current result will be displayed using this.

Pre-processing Steps

- In this step of the project , tweets are mined using Twitter Streaming API . Initially ,it cleans the unstructured textual data into structured textual data by removing punctuations and additional symbols.

1. Filtering : In this step , the special words , user names in twitter are removed.
2. Tokenization : is the act of breaking up a sequence of strings into pieces such as words, keywords, phrases, symbols and other elements called tokens.
3. Removal of Stop Words : Articles and other stop words are removed in this step.

TESTING

We performed some tests to check if our model was working fine or not. Now the case mentioned below, is the result of all the successful tests.

Title: Sentiment Analysis on US Election Campaign

Aim: To check is everything is working fine i.e we are getting the graph of polarity and subjectivity correctly or not.

Procedure: Run The Code , we can see which person tweets getting a positive response and which one is getting a negative response.

Expected Output: Joe Biden's Polarity graph is more accurate than Donald Trump's graph.

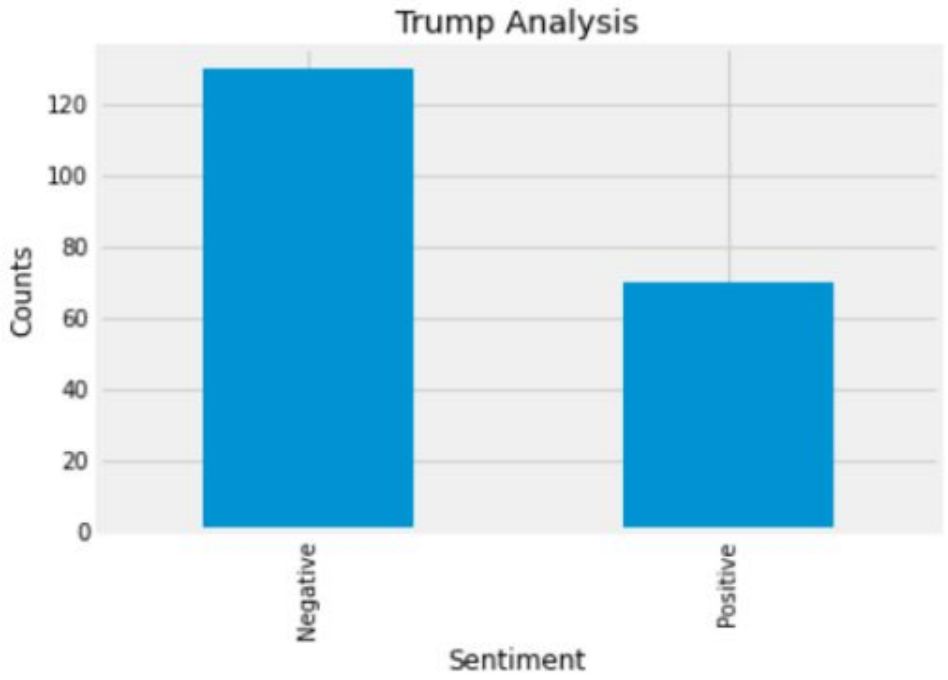
Observed Output: Joe Biden's Polarity graph is slightly more accurate than Donald Trump's graph.

Passed/Failed: Passed.

DATA VISUALIZATION

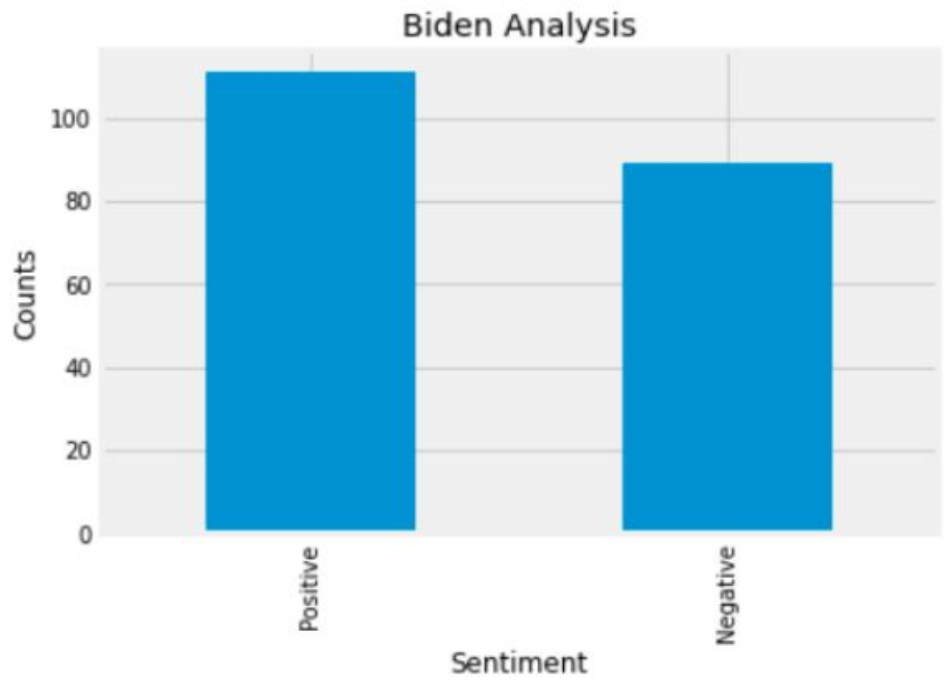
4.3.1. DONALD TRUMP

```
import matplotlib.pyplot

matplotlib.pyplot.title('Trump Analysis')
matplotlib.pyplot.xlabel('Sentiment')
matplotlib.pyplot.ylabel('Counts')
df['Analysis'].value_counts().plot(kind='bar')
matplotlib.pyplot.show()
```



4.3.2. JOE BIDEN

```
import matplotlib.pyplot

matplotlib.pyplot.title('Biden Analysis')
matplotlib.pyplot.xlabel('Sentiment')
matplotlib.pyplot.ylabel('Counts')
df1['Analysis'].value_counts().plot(kind='bar')
matplotlib.pyplot.show()
```

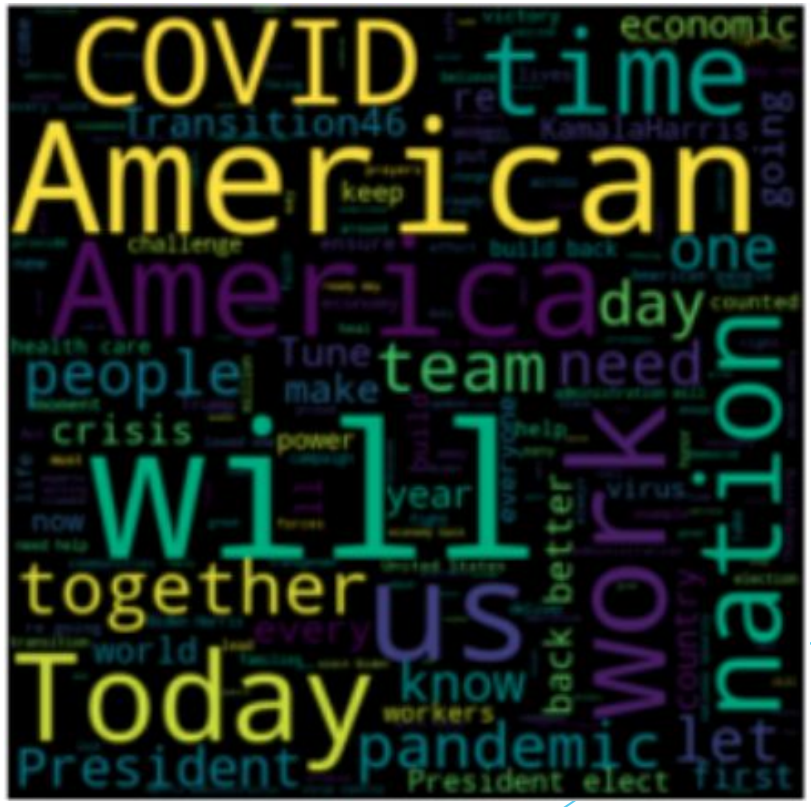


4.3.4.1. DONALD TRUMP WORD CLOUD

```
allwords=' '.join([twts for twts in df['Tweets']])
wordcloud=WordCloud(width=5000,height=5000,random_state=9).generate(allwords)

import matplotlib.pyplot
matplotlib.pyplot.imshow(wordcloud,interpolation='bilinear')
matplotlib.pyplot.axis('off')
matplotlib.pyplot.show()
```

4.3.4.2. JOE BIDEN WORD CLOUD

```
allwords1=' '.join([twts for twts in df1['Tweets']])
wordcloud1=WordCloud(width=500,height=500,random_state=21,max_font_size=119).generate(allwords1)

import matplotlib.pyplot
matplotlib.pyplot.imshow(wordcloud1,interpolation='bilinear')
matplotlib.pyplot.axis('off')
matplotlib.pyplot.show()
```



DEPLOYMENT

We have a GitHub repository and we will push all the code data into it. One important Thing, we also need to make a file called requirements.txt file and include all the Dependencies into that file.

Repository distribution should be based on HEROKU deployment steps

With respect to the project, our main.py file which we have made through FLASK app And pickle file in accordance with HEROKU deployment steps

FUTURE SCOPE

AREAS OF IMPROVEMENT:

1. The distribution of the data is poor
2.
3. The location identification part of the analysis needs
4. improvement:
5.
6. The prediction of the judgement for a state can be improved:
7.
8. Local language data processing

CONCLUSION

Sentiment analysis has become an important factor in the decision making process in a particular field.
In this paper we discussed techniques for preprocessing and information retrieval of tweets through twitter. Also we studied the Deep Learning technique: TensorFlow and TextBlob for text categorization which can be used to find out the polarity of textual tweet.
From study we can conclude that TextBlob acknowledges some properties of text like High Dimensional feature space, few irrelevant features, sparse instance vectors.
The performance of TensorFlow can be evaluated using precision and recall. Different results show that TextBlob gives good performance on text categorization as compared with Flair and Vader. With the ability to generalize high dimensional feature space, TextBlob eliminates the need for feature selection

THANK YOU