**Sentiment Analysis On Twitter Data**

**Name:-** Sundaram Saurabh

**University Roll Number:-** 2013539

**Problem Statement:-** Sentiment Analysis On Twitter Data.

**Motivation:-** For every student curiosity and the urge of being good at something is always the motivation. Social media websites have emerged as one of the platforms to raise users opinions and influence the way any business is commercialized. Opinion of people matters a lot to analyze how the propagation of information impacts the lives in a large-scale network. These days, the application of such analysis can be easily observed during public elections, movie promotions and many other fields. I have used Python in this project.

**Methodology Used:-**

* **Programming Language:-** Python
* **Tool:-** Jupiter Notebook

Sentiment analysis is a process of determining whether a text is positive, negative or neutral. Sentiment analysis is mainly used in E-commerce industry to understand whether a customer is happy about their product or not satisfied about their product.

To do twitter sentiment analysis, I have used twitter developer account with the help of which I have created a twitter API (app). For each and every app, four set of codes are generated that are as follows:

* API Key
* API Secret Key
* Access Token
* Access Token Secret

These are the four code that I have used to access the real time tweets from twitter. In this project, I have used two main python libraries. The first one is **Tweepy** which is used to access the twitter API and the second one is **TextBlob** which is used to process the text and also analyze the sentiment out of it. **TextBlob** can also be used to retrieve polarity and subjectivity. Polarity and subjectivity are the most common terminology used in sentiment analysis.

Polarity refers to emotions expressed in a particular statement. For example, if a person tweeting is feeling happy or sad or angry, and this can be expressed in a textual information. Polarity number ranges from **-1 to +1. -1** indicates that the statement is negative whereas **0** represents that the statement is neutral and **+1** represents that the statement is positive.

Subjectivity is nothing but expressing one’s own feelings, belief or opinion. Similar to polarity, subjectivity is number also ranges from**-1 to +1.** For example, if I am saying that I am very hardworking, so this has a subjectivity of **0.8** which means that this is my personal opinion and it is not a factual information.

For this project, I have used **Bill Gates** twitter account to see whether his tweets are mostly positive, negative or neutral.

**Libraries:**

I have imported several libraries for this project:

1. **tweepy**: To access twitter api
2. **textblob**: To process text and analyze the sentiments
3. **pandas**: To work with dataset
4. **numpy**: To work with arrays
5. **re**: To remove stopwords
6. **matplotlib**: To create charts, define parameters and color them

In a csv file, I have stored all the key codes and upload in the data frame using read\_csv() function of pandas.

Next, I have stored the API credentials in different variables. Then, I have created an authentication object to authenticate the credentials. After that the access token and access token secret are set and at last the API was created.

Now, I have extracted 100 tweets from the twitter account of Bill Gates and showed his 5 recent tweets. I have created a data frame with a column called **Tweets** that will contain the posts from his account and showed first ten rows of the data frame.

The data is a little bit dirty because it contains @ symbol, hyperlinks, retweets, and may contain hash tags, so I have cleaned it up by creating a function to remove these symbols from the tweets. Then, I have applied that function to the tweets and shown the result.

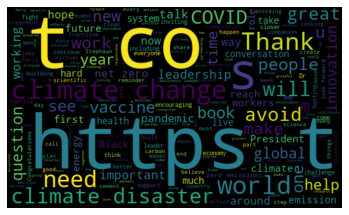
I want to add the tweets’ subjectivity and polarity to the data frame. In order to do this, I have created two functions: one to get the Subjectivity (how subjective or opinionated the text is —> a score of 0 is fact, and a score of +1 is very much an opinion) and the other to get the tweets called Polarity(how positive or negative the text is, —> score of -1 is the highest negative score, and a score of +1 is the highest positive score).

Next, I have stored the results into two columns, one called Subjectivity and the other called Polarity and shown the results.

To see how well the sentiments are distributed, the good way is to understand the common words by plotting a word cloud.

A word cloud (also known as text clouds or tag clouds) is a visualization, the more a specific word appears in the text, the bigger and bolder it appears in the word cloud.

From the word cloud I have got, it looks like the words such as **http, climate change, climate disaster, need, vaccine, covid, etc.** appears a lot in Bill Gates past 100 tweets.



Now, I have created a function to analyse the tweets i.e., whether a tweet is positive, negative or neutral and added the result in a new column named **Analysis** in the data frame and the modified data frame is shown.

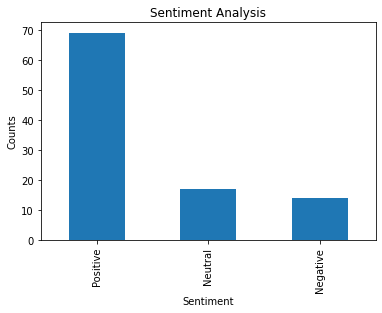
Now, I have printed all the positive tweets, negative tweets and neutral tweets separately and calculated their percentage respectively

The percentage of tweets came as follows:

* Positive Tweets 🡪 69%
* Negative Tweets 🡪 14%
* Neutral Tweets 🡪 17%

At last, I have plotted a bar graph in which in x-axis the sentiments are present and in y-axis the count of the sentiments is present.

The figure is shown below.



That’s all for this mini project.