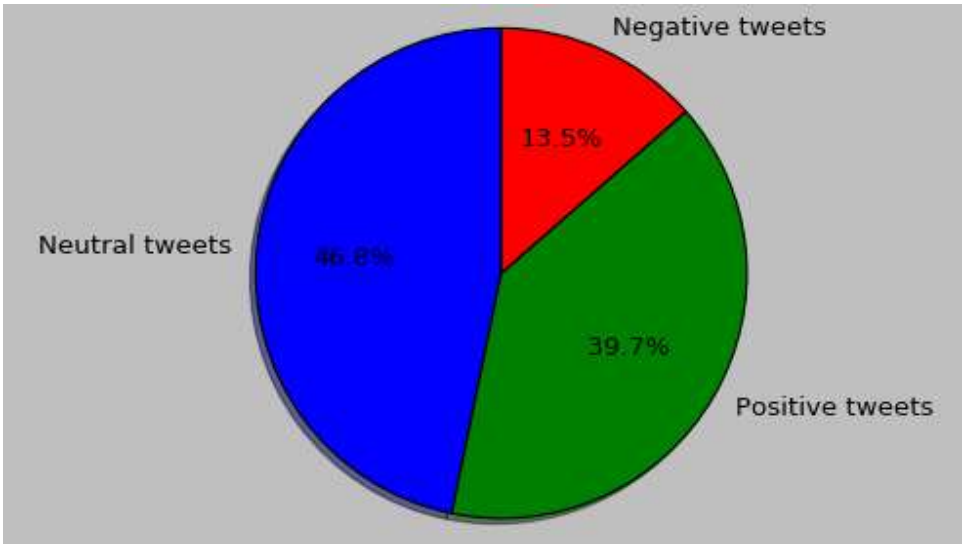
Twitter Sentiment Analysis using Python

Submitted by:-

Yash Agarwal(17dcs017)
Kartik Goyal(17uec061)
Sudhanshu Agrawal(17ucs161)

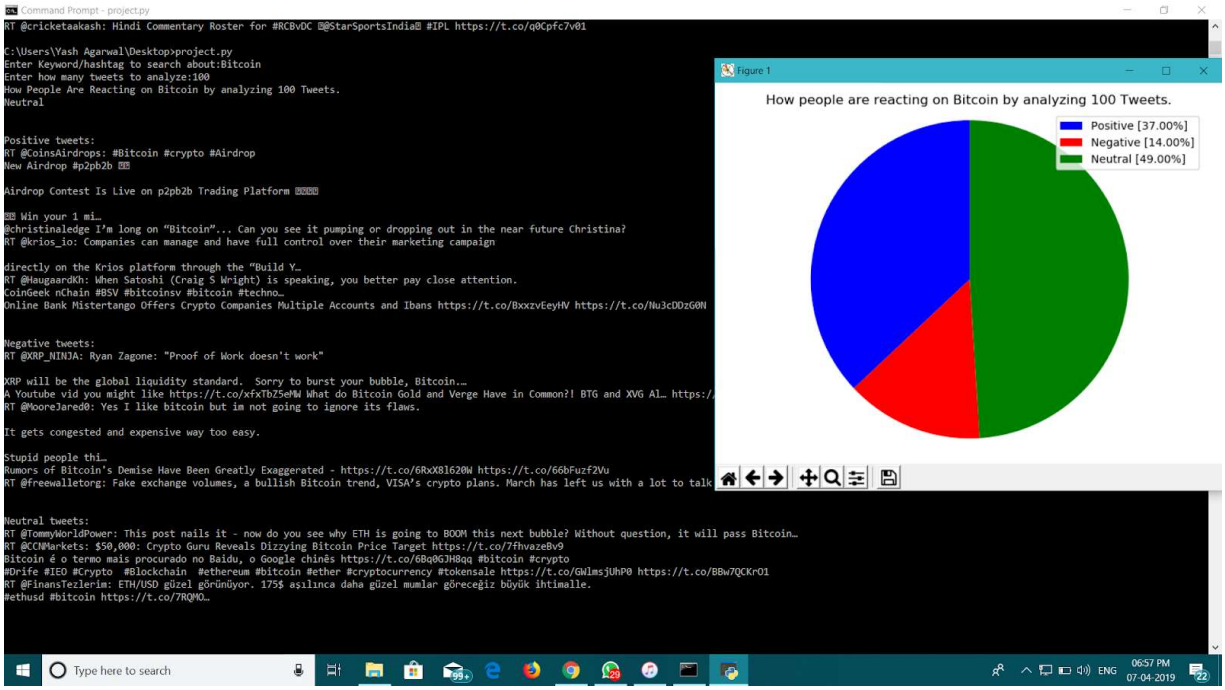
Project Description:

Our project is on Twitter Sentiment Analysis. Sentiment Analysis is the process of ‘computationally’ determining whether a piece of writing is positive, negative or neutral. It’s also known as **opinion mining**, deriving the opinion or attitude of a speaker.Our Project takes a keyword/hashtag and number of tweets ,to analyse from, as inputs and plots a pie chart for the same as shown below.



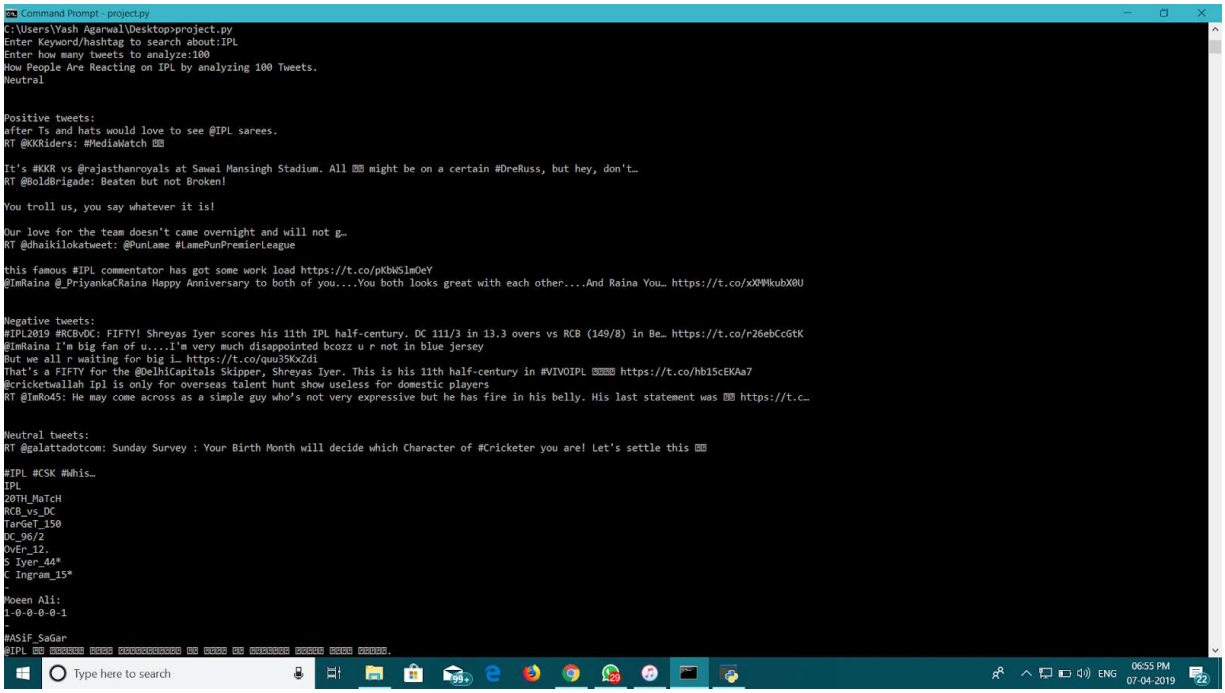
Screenshots:

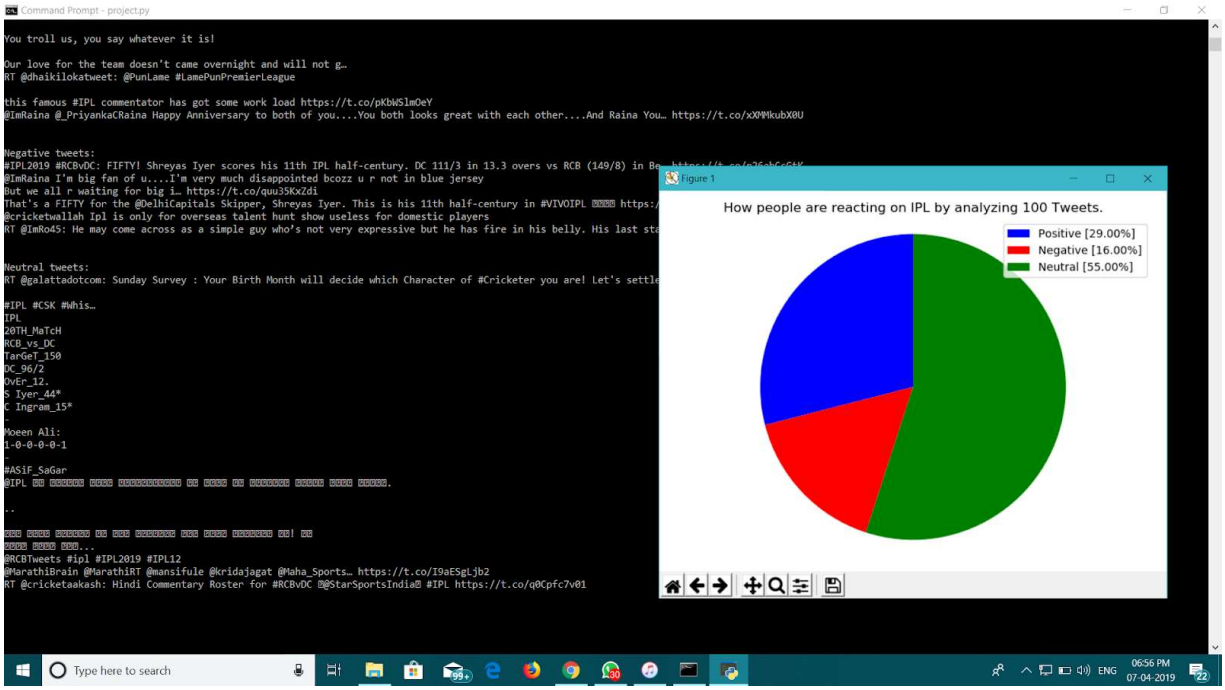
1.



This Screenshot shows the output of the code when the input string given is “Bitcoin” and 100 tweets are analysed.

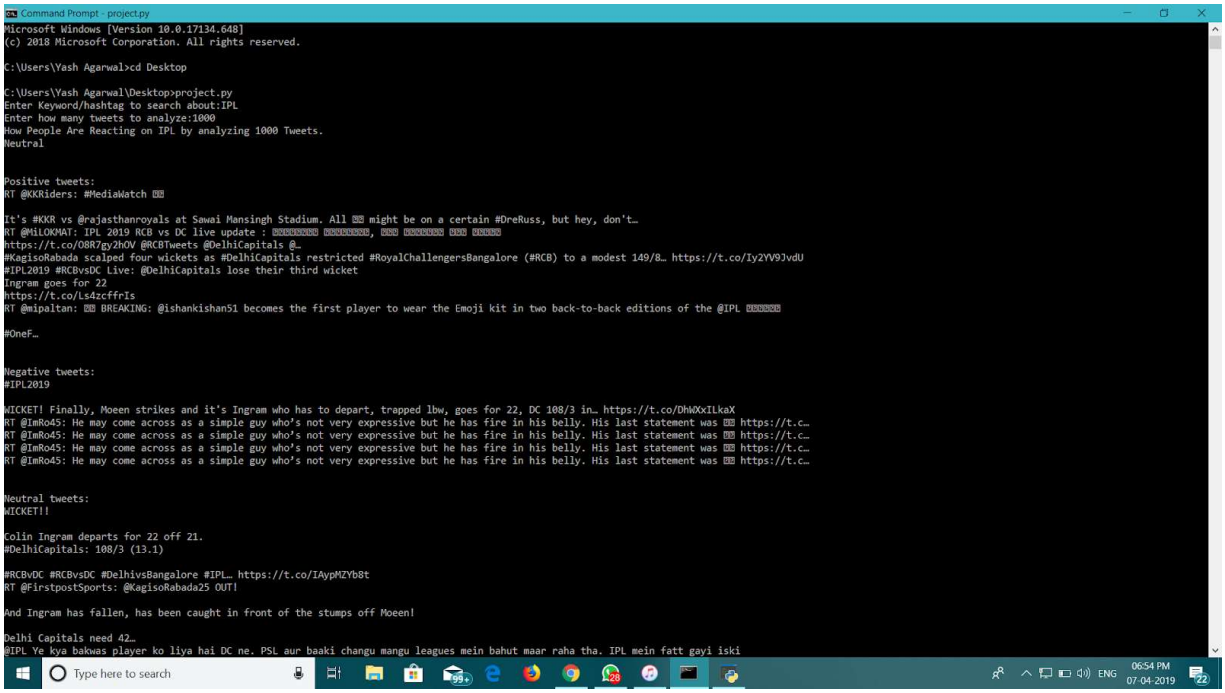
2.

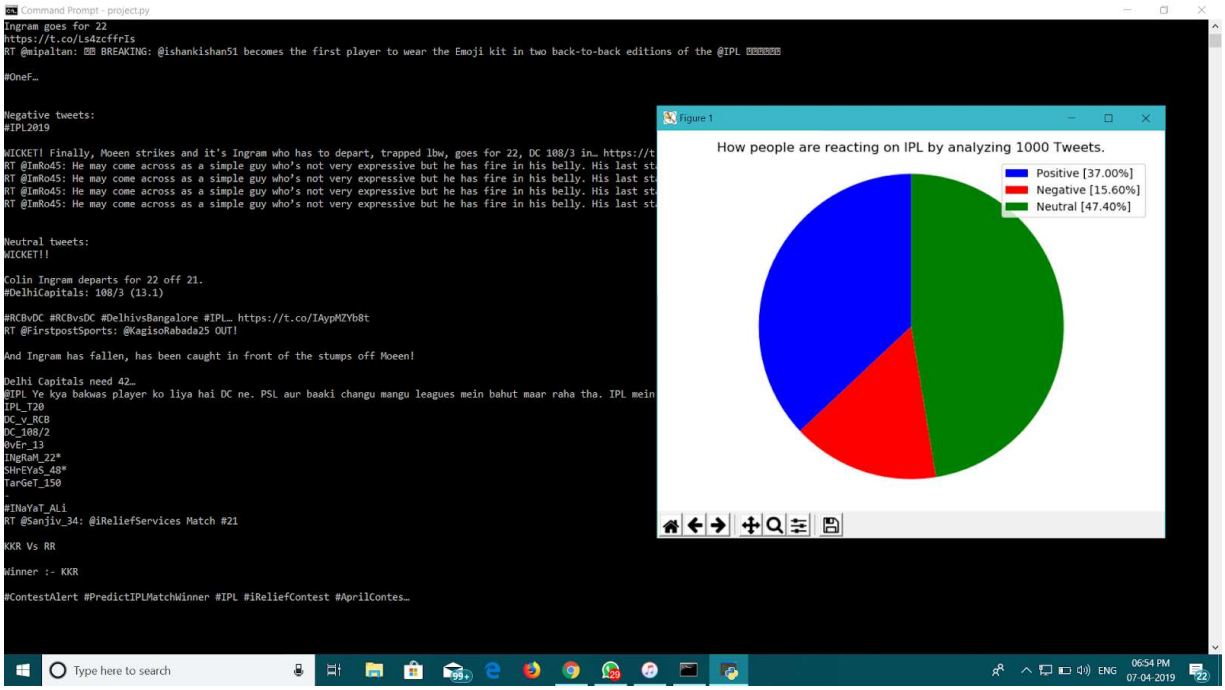




This Screenshot shows the output of the code when the input string given is “IPL” and 100 tweets are analysed.

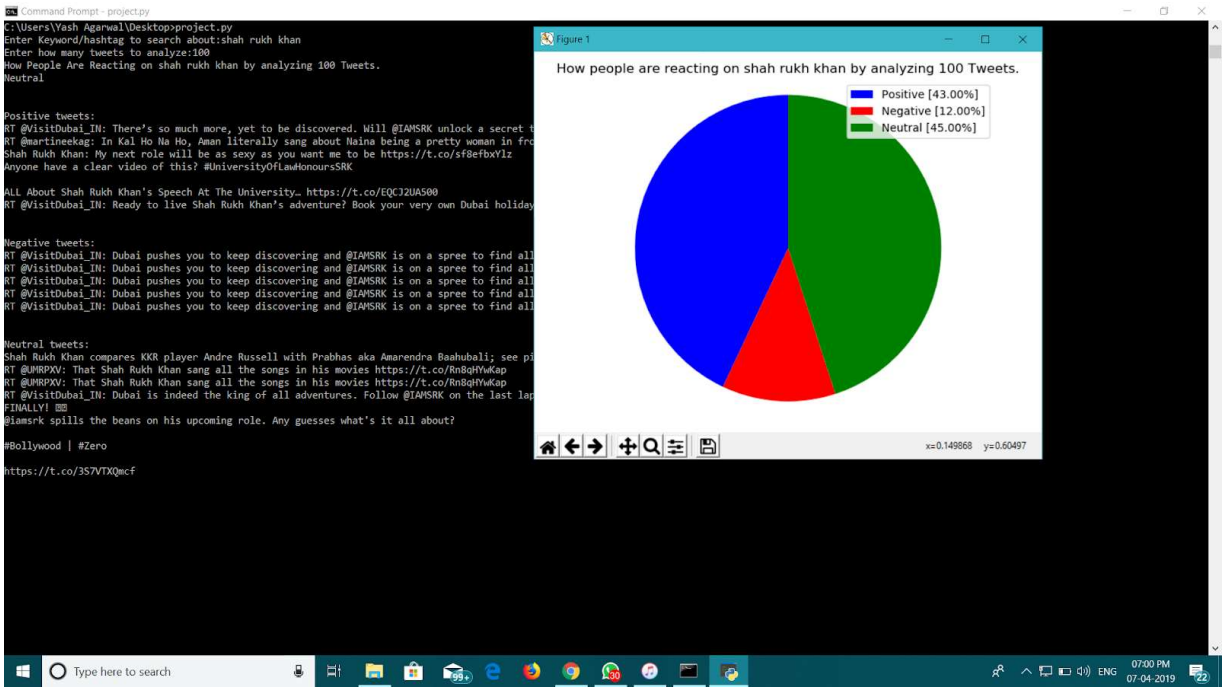
3.





This Screenshot shows the output of the code when the input string given is “IPL” and 1000 tweets are analysed.

4.



This Screenshot shows the output of the code when the input string given is “shah rukh khan” and 100 tweets are analysed.

To see the working of the code(video) *click here*

Installation:

Tweepy: tweepy is the python client for the official Twitter API.
Install it using following pip command:

pip install tweepy

TextBlob: textblob is the python library for processing textual data.
Install it using following pip command:

pip install textblob

Matplotlib:matplotlib is the python library for plotting data.
Install it using following pip command:

pip install matplotlib

Technology used:

Sentiment analysis combines natural language programming(NLP), text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information.
We have used Tweepy,TextBlob,Matplotlib,re,sys libraries.

Problem:

We believe, in today's time the major problem is still related to correct interpretation of context in which certain words are used.

It is still difficult for a vast majority of tools to precisely evaluate what truly is a negative, neutral, and a positive statement. we are not quite sure about the mechanics behind it, but at the moment it's not advanced enough to successfully deal with **sarcasm** or **context** of some of the discussions.

Think of a an example of someone being sarcastic in their Tweets, Facebook posts, whatever. Sometimes it's difficult to pick it up during a face-to-face conversation, let alone a tool analysing a context of a sarcastic social media update. Have a look at the example below:

Received an alert from BuzzSumo and used Google translate (guilty!) to decode it

While it's not the best example of someone being sarcastic, it's one of a few. This mention from Twitter was labelled as a negative one. The reason for it is the word "guilty" which is negative in its essence. The tool implies that it was used in a negative context in relation to our tool, however, in this particular case it was used jokingly and isn't a bad thing to our brand.

Still, it's a perfect example that in some instances, a pair of eyes of a person is essential to properly evaluate the sentiment of a piece of social media content.

Solution:

Definition

Sentiment analysis is a type of data mining that measures the inclination of people's opinions through natural language processing (NLP), computational linguistics and text analysis, which are used to extract and analyze subjective information from the Web - mostly social media and similar sources. The analyzed data quantifies the general public's sentiments or reactions toward certain products, people or ideas and reveal the contextual polarity of the information.

Sentiment analysis is also known as **opinion mining**.

What are we doing:

In our code we are first using Tweepy library to extract tweets from Twitter Database then we are using TextBlob library to calculate sentiment polarity of each tweet and finally Matplotlib library to demonstrate our analysis on a graph.Our code show the overall response for the input,whether it is positive or negative or neutral.Also prints 5 tweets of each sentiment (5 of positive sentiment, 5 of negative sentiment, 5 of neutral).

```
from textblob import TextBlob        #library to analyse sentiment
import re
import sys,tweepy                #library to extract twitter data
import matplotlib.pyplot as plt      #library to plot graph
```

Textblob:

TextBlob is a Python (2 and 3) library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.

Here we are using TextBlob to analyse each tweet based on it's sentiment.

```
for tweet in tweets:
    # cleaning tweet and analysis the polarity of tweet
    analysis = TextBlob(clean_tweets(tweet.text))
    polarity+=analysis.sentiment.polarity

    # count positive tweet and storing in list
    if(analysis.sentiment.polarity>0.00):
        positive=positive+1
        if(pcount<5):
            pcount+=1;
            ptweets.append(tweet)

    # count negative tweet and storing in list
    elif(analysis.sentiment.polarity<0.00):
        negative=negative+1
        if(ncount<5):
            ncount+=1;
            ntweets.append(tweet)

    # count neutral tweet and storing in list
    elif(analysis.sentiment.polarity==0):
        neutral=neutral+1
        if(nucount<5):
            nucount+=1;
            nutweets.append(tweet)
```

Tweepy:

Tweepy is the python client for the official Twitter API.The API class provides access to the entire twitter RESTful API methods. Each method can accept various parameters and return responses. When we invoke an API method most of the time returned back to us will be a Tweepy model class instance. This will contain the data returned from Twitter which we can then use inside our application.

Given below is the process to get Consumer Key,Consumer Secret,Access token,Access token secret.

Authentication:

In order to fetch tweets through Twitter API, one needs to register an App through their twitter account. Follow these steps for the same:

- Open this link and click the button: 'Create New App'
- Fill the application details. You can leave the callback url field empty.
- Once the app is created, you will be redirected to the app page.
- Open the 'Keys and Access Tokens' tab.
- Copy 'Consumer Key', 'Consumer Secret', 'Access token' and 'Access Token Secret'.

Here we are using tweepy to extract tweets from the twitter API.

```
# keys and tokens from the Twitter Dev Console

consumerKey='mgbOcqvdyJD3uh5pPQfPKtsWM'

consumerSecret='3DSpBeQB2IegVzFICK6a08w3TTCoNKFQ6H7q6WafBQgUirtJpC'

accessToken='923927317470093312-DhzXggAVIzISaCbtVEadhWtViYXaooN'

accessTokenSecret='l5SnMQuJuWP1oCEbIqAdWkLydR2qr7VWTKj2pGoUIhTVN'


# attempt authentication

try:

    # create OAuthHandler object

    auth=tweepy.OAuthHandler(consumer_key=consumerKey,consumer_secret=consumerSecret)

    # set access token and secret

    auth.set_access_token(accessToken,accessTokenSecret)

    # create tweepy API object to fetch tweets

    api=tweepy.API(auth)

except:

    print("Error: Authentication Failed")
```

Matplotlib:

It is a Python 2D plotting library. It has a module named pyplot which makes things easy for plotting by providing feature to control line styles, font properties, formatting axes etc.

Here matplotlib is used for plotting the graph(pie chart).

```
# ploting the graph and labeling of the sentiments

labels=['Positive ['+str(positive)+"%]",'Negative ['+str(negative)+"%]","Neutral ["+str(neutral)+"%]"]

sizes=[positive,negative,neutral]

colors=['blue','red','green']

patches,texts = plt.pie(sizes,colors=colors,startangle=90)

plt.legend(patches,labels,loc = "best")

plt.title("How people are reacting on "+searchTerm+" by analyzing " +str(noOfSearchTerms)+" Tweets.")

plt.axis("equal")

plt.tight_layout()

plt.show()
```

Advantages:

- Companies use Twitter Sentiment Analysis to develop their business strategies, to assess customers' feelings towards products or brand, how people respond to their campaigns or product launches and also why consumers are not buying certain products.
- In politics Sentiment Analysis Dataset Twitter is used to keep track of political views, to detect consistency and inconsistency between statements and actions at the government level. Sentiment Analysis Dataset Twitter is also used for analyzing election results.
- Twitter Sentiment Analysis also is used for monitoring and analyzing social phenomena, for predicting potentially dangerous situations and determining the general mood of the blogosphere.

Disadvantages:

- The time taken to analyse the data and produce output is directly proportional to the number of tweets to be analysed. As the no. of tweets increases,time taken to respond also increases.
- The results are not permanent.With every new tweet the results may change and one must do the process again to get the results right.
- Lesser the number of tweets to be analysed,lesser the accuracy.