AN INDUSTRIAL TRAINING REPORT

on

<Twitter Sentimental Analysis >

Submitted by

MOHD FIROJ
Roll No:161300052

Department of Computer Engineering & Applications

Institute of Engineering & Technology

GLA University
Mathura- 281406, INDIA
21 November, 2018

Declaration:-

Dated: 25/07/2018

Ref No. DU/PL18/A52/13495

TO WHOM IT MAY CONCERN

This is to certify that **MOHD. FIROJ, Bachelor of Technology** Student from **GLA University, Mathura (UP)** has successfully completed the project at our organization from **11th June 2018 to 25th July 2018.** Under training period he has worked on the project entitled **"Twitter Analysis"** using **Python & Data Science Technology** under the guidance of our Technical Head **Mr. Tanay.** During this project he was an active member of the team working on the project. The work carried by him was satisfactory and wish him all the best for his future assignments.

Best Regards,

Auth. Signature

Dept. of CEA, GLAU, Mathura

STUDENT ID NO.- A52/13495

DUCAT

An ISO 9001 : 2000 Company

The council of research and training,
Ducat Noida certifies that

Mohd. Firoj

has been awarded a certificate of proficiency in

Python & Data Science 6Weeks

for having successfully completed the program with

A *Grade.*

The program was conducted during the period

11/06/2018   25/07/2018

given under the seal of Ducat Noida.

Program Co-ordinator

Director

Dept. of CEA, GLAU, Mathura

Abstract-

Social networks are the main resources to gather information about people's opinion and sentiments towards different topics as they spend hours daily on social medias and share their opinion. Twitter is a microblogging platform which provides a tremendous amount of data which can be used for various applications of Sentiment Analysis like predictions, reviews, elections, marketing, etc. Sentiment Analysis is a process of extracting information from large amount of data, and classifies them into different classes called sentiment In this technical report, we show the application of sentimental analysis and how to connect to Twitter and run sentimental analysis queries.

Dept. of CEA, GLAU, Mathura

1.Introduction-

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. Sentimental Analysis is done by using various machine learning techniques, statistical models and Natural Language Processing (NLP) for the extraction of feature from a large data.

Sentiment Analysis can be done at document, phrase and sentence level. In document level, summary of the entire document is taken first and then it is analyze whether the sentiment is positive, negative or neutral. In phrase level, analysis of phrases in a sentence is taken in account to check the polarity. In Sentence level, each sentence is classified in a particular class to provide the sentiment.

Sentimental Analysis has various applications. It is used to generate opinions for people of social media by analyzing their feelings or thoughts which they provide in form of text. Sentiment Analysis is domain centered, i.e. results of one domain cannot be applied to other domain. Sentimental Analysis is used in many real life scenarios, to get reviews about any product or movies, to get the financial report of any company, for predictions or marketing.

Twitter is a micro blogging platform where anyone can read or write short form of message which is called tweets. The amount of data accumulated on twitter is very huge. This data is unstructured and written in natural language. Twitter Sentimental Analysis is the process of accessing tweets for a particular topic and predicts the sentiment of these tweets as positive, negative or neutral with the help of different machine learning algorithm.

Dept. of CEA, GLAU, Mathura

1.1 Overview and Motivation-

We have chosen to work with twitter since we feel it is a better approximation of public sentiment as opposed to conventional internet articles and web blogs. The reason is that the amount of relevant data is much larger for twitter, as compared to traditional blogging sites. Moreover the response on twitter is more prompt and also more general (since the number of users who tweet is substantially more than those who write web blogs on a daily basis). Sentiment analysis of public is highly critical in macroscale socioeconomic phenomena like predicting the stock market rate of a particular firm. This could be done by analysing overall public sentiment towards that firm with respect to time and using economics tools for finding the correlation between public sentiment and the firm's stock market value. Firms can also estimate how well their product is responding in the market, which areas of the market is it having a favourable response and in which a negative response (since twitter allows us to download stream of geo-tagged tweets for particular locations. If firms can get this information they can analyze reasons behind geographically differentiated response, and so they can market their product in a more optimized manner by looking for appropriate solutions like creating suitable market segments. Predicting the results of popular political elections and polls is also an emerging application to sentiment analysis. One such study was conducted by Tumasjan et al. in Germany for predicting the outcome of federal elections in which concluded that twitter is a good reflection of offline sentiment.

1.2 Objective:-

The main objective of this project is to perform the sentiment analysis on Indian Politicians like Narendra Modi,Rahul Ghandi, such that people opinions about these leaders which are extracted from Twitter.Thus to achieve this objective we build a classifier based on supervised learning and perform live sentiment analysis on twitter data.

Dept. of CEA, GLAU, Mathura

1.3 Summary of Similar Applications-

NCSU Tweet Sentiment Visualization App (Web App)- The NCSU Tweet Visualizer is particularly strong in the following-

Simply enter a keyword, and the Tweet Visualizer automatically pulls recent tweets (from the past week, though the time range is shorter for popular subjects).

By measuring pleasure, activation and dominance, the NCSU Tweet Visualizer offers far more dimensions than can be found in many other free sentiment analysis tools. Most of these tools only focus on the "pleasure" dimension and rate sentiment according to a three-value scale: positive, negative and neutral.

Enginuity (Web App)- It works differently from many of the free sentiment analytics tools out there. Instead of directly querying *tweets* related to a certain keyword, Enginuity allows you to search for *recent news stories* about the keyword.

The tool then queries both Twitter and Facebook to calculate how many times the story has been shared. It also analyzes whether the sentiment of social shares is positive or negative, and gives an aggregate sentiment rating for the news story.

Enginuity is thus a great tool for finding stories to share through your social channels, as well as getting a combined picture of sentiment about recent events trending on social media.

socialmention (Web App)- Socialmention is a basic, search engine-style web app for topic-level sentiment analysis on Twitter data. You can enter a keyword, and the tool will return aggregate sentiment scores for the keyword as well as related keywords.

One neat feature of socialmention is support for basic brand management use cases—the tool returns a "passion" score that measures how likely Twitter users are to discuss your brand, as well as the average reach of the Twitter users discussing your brand.

The caveat is that the tool still returns wonky results for lesser-known brands, but this is an issue with sentiment analysis in general.

Dept. of CEA, GLAU, Mathura

Steamcrab (Web App)- Steamcrab is a web application for sentiment analytics on Twitter data. It focuses on keyword searches and analyzes tweets according to a two-pole scale (positive and negative). Visualization options are limited to scatter plots and pie charts.

Revealed Context (API/Excel Add-in)- Revealed Context offers a free API for running sentiment analytics on up to 250 documents per day. There's an Excel add-in as well as a web interface for running analytics independently of the API

While Revealed Context doesn't offer an interface for directly scraping Twitter, it's simple enough to analyze a spreadsheet of tweets without using the API. With the API, you can build a pipeline that feeds recent tweets from the Twitter API into the Revealed Context API for processing.

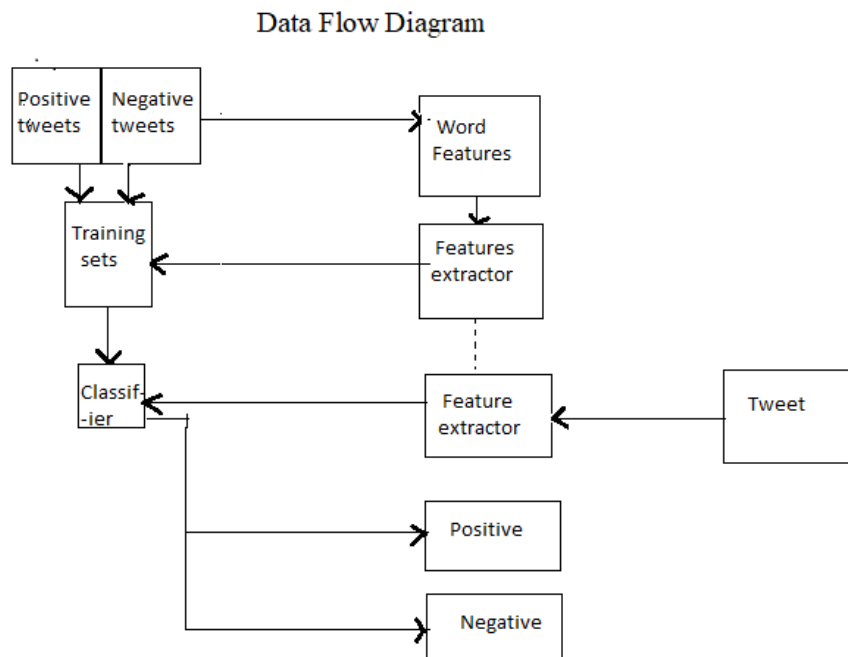Dept. of CEA, GLAU, Mathura

2. Company Profile-

 Ducat provides the best available programs which helps in enhancing the technical skills which   seems to be beneficial for all the applicants.
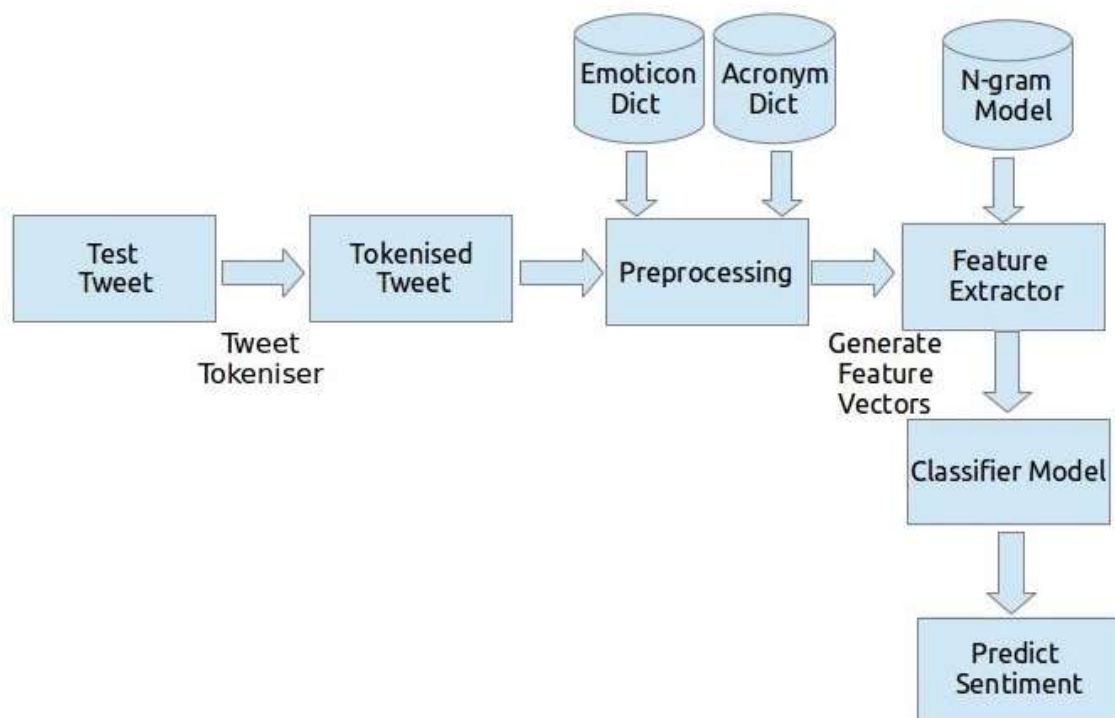
- **Software Development:** We provide the best and latest IT software training which helps all the fresher and the corporates to understand well and give them the knowledge to go hand in hand with the latest technologies. This does not only helps the companies but also increases the self-level to deal with all the necessary software.

- **Instructor led campus:** Ducat helps all the new instructors to get the best exposure to show their talent in right way.

- **Workshops and Placement Service:** At Ducat, workshops are held to increase the understanding level because theoretical values are always not enough and workshops helps in getting the practical knowledge which results in better understanding. As everything leads to the placement because if the institute does not provide placement services then it is ultimately bad for the applicants but we provide the best placement services and for that we give our best to give you the best

Dept. of CEA, GLAU, Mathura

3. Project Design-

3.1 Data Flow Diagram:-

Data Flow Diagram

Dept. of CEA, GLAU, Mathura

3.2 Class Diagram-

Dept. of CEA, GLAU, Mathura
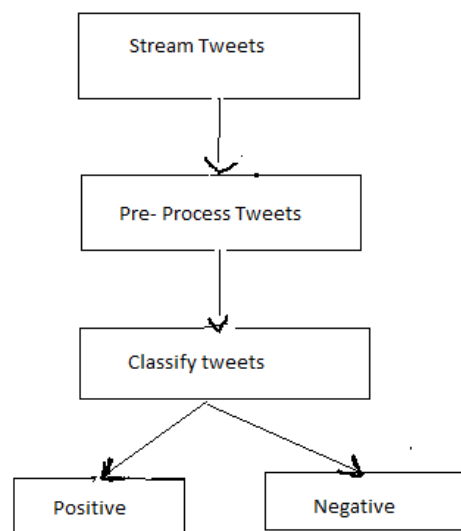
4. Implementation and User Interface-

As our goal is to achieve sentiment analysis for data provided from Twitter. We are going to build a classifier which consists of different machine learning classifiers.

Once our classifier is ready and trained we are going to follow the steps shown in figure

```
┌─────────────────────┐
│   Stream Tweets     │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│  Pre- Process Tweets│
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│   Classify tweets   │
└─────────────────────┘
        ╱        ╲
       ▼          ▼
┌──────────┐  ┌──────────┐
│ Positive │  │ Negative │
└──────────┘  └──────────┘
```

Step-1 First we are going to stream tweets in our build classifier with the help of Tweepy library in python

Step-2

Then we preprocess these tweets, so that they can be fit for mining andfeature extraction.

Step-3 After preprocessing we pass this data in our trained classifier, which then classify them into positive or negative class based on trained results.Since, Twitter is our source of data for analysis. We are going to stream the tweets from twitter in our database. For this we are going to use Twitter Application

Dept. of CEA, GLAU, Mathura

Twitter API (Application Programming Interface)

Twitter allows users to collect tweets with the help of Twitter API. Twitter provides two kinds of APIs: REST API and Streaming API. The differences between these are: REST APIs support connections for short time interval and only limited data can be collected at a time, whereas Streaming API provides tweets in real-time and connection for long time. We use Streaming API for our analysis. For collecting large amount of tweets we need long-lived connection and no limit data rate

ⓘ localhost:8888/notebooks/Twitter%20analysis_self.ipynb

ePirateBay - Down | ◇ ONLINE LABS | Day of the Program | Technical Report

Jupyter Twitter analysis_self Last Checkpoint: 07/23/2018 (unsaved changes) | Logout

File   Edit   View   Insert   Cell   Kernel   Widgets   Help | Trusted | Python 3 O

Code

Positive tweets percentage: 17.647058823529413 %

Positive tweets:
RT @vijayrupanibjp: Narendra Modi wants India among top 50 on ease of doing business index, double economy to $5trillion
https://t.co/Cz7Q3...
RT @MPMadanSaini: Hardworking honest and inspiring, the one name you can trust, PM Narendra Modi. Stand with him. Donate to the BJP. https:...
RT @VOANews: Ibrahim Mohamed Solih was sworn-in as the new president of Maldives on Saturday, Nov. 17 at a ceremony held in the capital Mal...
@TwitterIndia And Indian citizens voted Narendra Modi as PM from a backward class who has had great meetings with D... https://t.co/J4Fq3338Hh
RT @AbhinavDeekonda: @ThinkersPad @PMOIndia I don't see tweets from many prominent people I follow. PMO, Narendra Modi, etc. tweets don't a...
RT @thewire_in: "Pakistan frees a blasphemy accused, hells bells!
are they getting more liberal than my favourite BJP(?)" https://t.co/29bH...
RT @narendramodi: A special day and a special tribute.

Here is my article, on the landmark contribution of Sardar Patel to our country an...
RT @saket71: Dear @SwatiJaiHind the man speaking to the woman anchor on live TV isn't followed by Narendra Modi and therefore his abuse doe...
RT @SirrrJadeja: Sir Ravindra Jadeja And His Wife Riva Meets Indian PM Narendra Modi. It's A Great Honour 🙏in https://t.co/DrzFzNo77G
RT @pbhushan1: Read: CBI DIG's explosive allegations against Modi minister, NSA & top PMO official. He was investigating Modi//Shah's pet...

Negative tweets:
RT @aaptimist_: This is the guy who attacked @ArvindKejrival and not surprisingly he is a BJP member.

Dept. of CEA, GLAU, Mathura

Negative tweets:
RT @aaptimist_: This is the guy who attacked @ArvindKejriwal and not surprisingly he is a BJP member.

Here's the proof from his Facebook..
RT @ThePrintIndia: Angry Modi 2.0 in Madhya Pradesh has blown the bugle for 2019

Jyoti Malhotra @jomalhotra writes in #ModiMonitor: https:..
RT @investindia: Hon'ble Prime Minister, Narendra Modi, lauded the efforts of state govt.'s in improving India's rank in @Worl
dBank's Doin..
RT @KarnikaKohli: Watch | Why CBI official's petition is bad news for Narendra Modi, Ajit Doval https://t.co/064E4C3fIN | @rohi
ni_sgh and @..
RT @kapsology: Haribhai Parthibhai Chaudhary
Rakesh Asthana
CVC KV Chaudhary
Ajit Dhoval

What is the common thing between all these people..
RT @thewire_in: NSA Ajit Doval allegedly intervened in the CBI's investigation into Asthana and Kumar and ensured that their mo
biles - whic..
RT @ElvinBaruah: Suresh Chavhanke was arrested by UP Police for Communal Violence, He was Active executive of RSS. It's proved
 that Sudarsh..
RT @thewire_in: Watch | Why CBI official's petition is bad news for Narendra Modi, Ajit Doval https://t.co/kgKnnGGuwx | @rohini
_sgh and @sv..
RT @aurkuchbhi: Seeing so much hate against Brahmins should I conclude that no one is going to give their vote to Janeudhari br
ahmin Rahul..
RT @Manjula44001343: PM Narendra Modi aims to double Indian economy, eyes top 50 rank in 'ease of doing business' - For this to
 happen, Dr...

In [ ]:

Dept. of CEA, GLAU, Mathura

Bibliography-

https://www.wikipedia.com

Python for data analysis - Wes McKinney

https://www.udemy.com

Dept. of CEA, GLAU, Mathura

Appendices-
Code templates-

```python
Import re
Import tweepy
From tweepy import OAuthHandler
From textblob import TextBlob
class
TwitterClient(object):
'''
Generic Twitter Class for sentiment analysis.
'''
def
__init__(self):
'''
Class constructor or initialization method.
'''
# keys and tokens from the Twitter Dev Console
consumer_key =
'XXXXXXXXXXXXXXXXXXXXXXXX'
consumer_secret =
'XXXXXXXXXXXXXXXXXXXXXXXXXXX'
access_token =
'XXXXXXXXXXXXXXXXXXXXXXXXXX'
access_token_secret =
'XXXXXXXXXXXXXXXXXXXXXXXX'
# attempt authentication
try:
# create OAuthHandler object
self.auth =
OAuthHandler(consumer_key, consumer_secret)
# set access token and secret
self.auth.set_access_token(access_token,
access_token_secret)
# create tweepy API object to fetch tweets
self.api =
tweepy.API(self.auth)
except:
print("Error: Authentication Failed")
def
clean_tweet(self, tweet):
'''
Utility function to clean tweet text by removing links,
special characters
using simple regex statements.
'''
return
' '.join(re.sub("(@[A-Za-z0-9]+)|([^0-9A-Za-z \t])
|(\w+:\/\/\S+)", "
",
tweet).split())
```

Dept. of CEA, GLAU, Mathura

```python
def
get_tweet_sentiment(self, tweet):
'''
Utility function to classify sentiment of passed tweet
using textblob's sentiment method
'''
# create TextBlob object of passed tweet text
analysis =
TextBlob(self.clean_tweet(tweet))
# set sentiment
if
analysis.sentiment.polarity > 0:
return
'positive'
elif
analysis.sentiment.polarity ==
0:
return
'neutral'
else:
return
'negative'
def
get_tweets(self, query, count =
10):
'''
Main function to fetch tweets and parse them.
'''
# empty list to store parsed tweets
tweets =
[]
try:
# call twitter api to fetch tweets
fetched_tweets =
self.api.search(q =
query, count =
count)
# parsing tweets one by one
for
tweet in
fetched_tweets:
# empty dictionary to store required params of a
tweet
parsed_tweet =
{}
# saving text of tweet
parsed_tweet['text'] =
tweet.text
# saving sentiment of tweet
parsed_tweet['sentiment'] =
```

Dept. of CEA, GLAU, Mathura

```python
self.get_tweet_sentiment(tweet.text)
# appending parsed tweet to tweets list
if
tweet.retweet_count > 0:
# if tweet has retweets, ensure that it is
appended only once
if
parsed_tweet not
in
tweets:
tweets.append(parsed_tweet)
else:
tweets.append(parsed_tweet)
# return parsed tweets
return
tweets
except
tweepy.TweepError as e:
# print error (if any)
print("Error : "
+
str(e))
def
main():
# creating object of TwitterClient Class
api =
TwitterClient()
# calling function to get tweets
tweets =
api.get_tweets(query =
'Donald Trump', count =
200)
# picking positive tweets from tweets
ptweets =
[tweet for
tweet in
tweets if
tweet['sentiment'] ==
'positive']
# percentage of positive tweets
print("Positive tweets percentage: {}
%".format(100*len(ptweets)/len(tweets)))
# picking negative tweets from tweets
ntweets =
[tweet for
tweet in
tweets if
tweet['sentiment'] ==
'negative']
# percentage of negative tweets
```

Dept. of CEA, GLAU, Mathura

```python
print("Negative tweets percentage: {}
%".format(100*len(ntweets)/len(tweets)))
# percentage of neutral tweets
print("Neutral tweets percentage: {} %
\
".format(100*len(tweets -
ntweets -
ptweets)/len(tweets)))
# printing first 5 positive tweets
print("\n\nPositive tweets:")
for
tweet in
ptweets[:10]:
print(tweet['text'])
# printing first 5 negative tweets
print("\n\nNegative tweets:")
for
tweet in
ntweets[:10]:
print(tweet['text'])
if
__name__ ==
"__main__":
# calling main function
main()
```

Here is how a sample output looks like when above program is run:

Positive tweets percentage: 22 %

Negative tweets percentage: 15 %

Positive tweets:

RT @JohnGGalt: Amazing—after years of attacking Donald Trump the media managed to turn #InaugurationDay into all about themselves. #MakeAme...

RT @vooda1: CNN Declines to Air White House Press Conference Live YES! THANK YOU @CNN FOR NOT LEGITIMI...

RT @Muheeb_Shawwa: Donald J. Trump's speech sounded eerily familiar...

POTUS plans new deal for UK as Theresa May to be first foreign leader to meet new president since inauguration

.@realdonaldtrump #Syria #Mexico #Russia & now #Afghanistan.

Another #DearDonaldTrump Letter worth a read @AJEnglish

Negative tweets:

RT @Slate: Donald Trump's administration: "Government by the worst men."

RT @RVAwonk: Trump, Sean Spicer, et al. lie for a reason. Their lies are not just lies. Their lies are authoritarian propaganda.

RT @KomptonMusic: Me: I hate corn

Donald Trump: I hate corn too

Me: https://t.co/GPgy8R8HB5

It's ridiculous that people are more annoyed at this than Donald Trump's sexism.

RT @tony_broach: Chris Wallace on Fox news right now talking crap about Donald Trump news conference it seems he can't face the truth eithe...

RT @fravel: With False Claims, Donald Trump Attacks Media on Crowd Turnout

Dept. of CEA, GLAU, Mathura

Aziz Ansari Just Hit Donald Trump Hard In An Epic Saturday NIght Live Monologue

We follow these 3 major steps in our program:

♣

Authorize twitter API client.

♣

Make a GET request to Twitter API to fetch tweets for a particular query.

♣

Parse the tweets. Classify each tweet as positive, negative or neutral.

Now, let us try to understand the above piece of code:

♣

First of all, we create a Twitter Client class. This class contains all the methods to interact with Twitter API and parsing tweets. We use __init__ function to handle the authentication of API client.

♣

In get_tweets function, we use:

fetched_tweets = self.api.search(q = query, count = count)

to call the Twitter API to fetch tweets.

♣

In get_tweet_sentiment we use textblob module.

analysis = TextBlob(self.clean_tweet(tweet))

TextBlob is actually a high level library built over top of NLTKlibrary. First we call clean_tweet method to remove links, special characters, etc. from the tweet using some simple regex.Then, as we pass tweetto create a TextBlobobject, following processing is done over text by textblob library:

♣

Tokenize the tweet ,i.e split words from body of text.

♣

Remove stopwords from the tokens.(stopwords are the commonly used words which are irrelevant in text analysis like I, am, you, are, etc.)

♣

Do POS(part of speech) tagging of the tokens and select only significant features/tokens like adjectives, adverbs, etc.

♣

Pass the tokens to a sentiment classifier which classifies the tweet sentiment as positive, negative or neutral by assigning it a polarity between -1.0 to 1.0 .Here is how sentiment classifieris created:

♣

TextBlob uses a Movies Reviews dataset in which reviews have already been labelled as positive or negative.

♣

Positive and negative features are extracted from each positive and negative review respectively.

♣

Training data now consists of labelled positive and negative features. This data is trained on a Naive Bayes Classifierm

Then, we use sentiment.polarity method of TextBlobclass to get the polarity of tweet between -1 to 1.Then, we classify polarity as:

if analysis.sentiment.polarity > 0:

return 'positive'

```
elif analysis.sentiment.polarity == 0:
return 'neutral'
else:
return 'negative'
```
♣

Finally, parsed tweets are returned. Then, we can do various type of statistical analysis on the tweets. For example, in above program, we tried to find the percentage of positive, negative and neutral tweets about a query

Dept. of CEA, GLAU, Mathura