TWITTER SENTIMENTAL ANALYSIS

Project Report

Submitted in fulfillment of the requirement for
the degree of Bachelor of Technology
In
Computer Science and Engineering and Information Technology

By
Abhinav Kirti Gupta (151274)
Abhimanyu Chauhan (151475)

Under the supervision of
Dr. Hari Singh

To

Department of Computer Science & Engineering and Information Technology

Jaypee University of Information Technology Waknaghat, Solan-173234, Himachal Pradesh

Certificate

Candidate's Declaration

I hereby declare that the work presented in this report entitled "Twitter Sentimental Analysis" in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering/Information Technology submitted in the department of Computer Science & Engineering and Information Technology, Jaypee University of Information Technology Waknaghat is an authentic record of my own work carried out over a period from July 2018 to May 2019 under the supervision of Dr. Hari Singh (Assistant Professor, Senior Grade, Computer Science &Engineering Department).

The matter embodied in the report has not been submitted for the award of any other degree or diploma.

(Student Signature)

Abhinav Kirti Gupta(151274) Abhimanyu Chauhan(151475)

This is to certify that the above statement made by the candidate is true to the best of my knowledge.

(Supervisor Signature)
Supervisor Name: Dr. Hari Singh
Designation: Assistant Professor (Senior Grade)
Department name: Computer Science and Engineering and Information Technology

Dated:

ACKNOWLEDGEMENT

I want to take this chance to thank almighty for blessing me with his grace and taking my job to a successful culmination. We owe our profound gratitude to our project supervisor, Dr. Hari Singh who took keen interest and guided us all along in my project work titled - “Twitter Sentimental Analysis” till the completion of our project by providing all the necessary information for developing the project. The project development helped us in research and we got to know a lot of new things in our domain. We are really thankful to him.

TABLE OF CONTENT

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

This project addresses the problem of sentiment analysis in twitter; that is classifying tweets according to the sentiment expressed in them: positive, negative or neutral. Twitter is an online micro-blogging and social-networking platform which allows users to write short status updates of maximum length 140 characters.

It is a rapidly expanding service with over 200 million registered users - out of which 100 million are active users and half of them log on twitter on a daily basis - generating nearly 250 million tweets per day. Due to this large amount of usage we hope to achieve a reflection of public sentiment by analyzing the sentiments expressed in the tweets. Analyzing the public sentiment is important for many applications such as firms trying to find out the response of their products in the market, predicting political elections and predicting socioeconomic phenomena like stock exchange. The aim of this project is to develop a functional classifier for accurate and automatic sentiment classification of an unknown tweet stream.

CHAPTER 1: INTRODUCTION

1.1) Introduction :

Sentimental Analysis refers to the way of characterizing word into the 3 categories namely: positive, negative or neutral. This can be collected from social network sites because people upload their emotions over them on regular basis.

Let's take a look over the concept of social media and it's most popular platform i.e. Twitter :

1.1.1) Social Media

Social Media is a group of Internet-based applications that create on the ideological and technological foundations of Web2.0 which is allowed to build and exchange of user-generated contents. In a discussion of Internet, World Start identified that a trend of internet users is increasing and continuing to spend more time with social media by the total time spent on mobile devices and social media in the U.S across PC increased by 37 percent to121 billion minutes in 2012, compared to 88 billion minutes in 2011[20].

On the other hand, businesses use social networking sites to find and communicate with clients, a business can be demonstrated damage to productivity caused by social networking. As social media can be posted so easily to the public, it can harm private information to spread out in the social world. In addition, mentioned that social media is also being used for advertisement by companies for promotions, professionals for searching, recruiting, social learning online and electronic commerce. Electronic commerce or E-commerce refers to the purchase and sale of goods or services online which can via social media, such has Twitter which is convenient due to its 24-hours availability, ease of customer service and global reach. Among the reasons why business tends to use more social media is for getting insight into consumer behavioral tendencies, market intelligence and presents an opportunity to learn about customer review and perceptions. Twitter Sentiment Analysis The sentiment can be found in the comments or tweet to provide useful indicators for many different purposes[21]. Also, and stated that sentiment can be categorized into two groups, which is negative and positive words. Sentiment analysis is a natural language processing technique to quantify an expressed opinion or sentiment within a selection of tweets.

1.1.2) Twitter

Twitter is a prevalent ongoing micro blogging administration that enables clients to share short data known as tweets which are constrained to 140 characters (now limit is extended). Clients compose tweets to express their feeling about different subjects identifying with their day by day lives. Twitter is a perfect stage for the extraction of overall population conclusion on explicit issues.

A gathering of tweets is utilized as the essential corpus for assessment examination, which alludes to the utilization of supposition mining or normal language handling. Twitter, with 500 million clients and million messages for each day, has rapidly turned into an important resource for associations to invigilate their notoriety and brands by separating and examining the slant of the tweets by the general population about their items, administrations showcase and even about contenders featured that, from the online life produced suppositions with the mammoth development of the internet, super volumes of feeling writings as tweets, surveys, web journals or any exchange gatherings and discussions are accessible for investigation, accordingly making the internet the quickest, most containing and effectively available vehicle for assumption examination.as an e-commerce marketing tool .

Though, the micro blogging platform has been developed a few years' time for promoting foreign trade website by using a foreign micro blogging platform as Twitter marketing. The instant of sharing, interactive, community-oriented features are opening e-commerce, launched a new bright spot which it can be shown that micro blogging platform has enabled companies do brand image, product important sales channel, improve product sales, talk to the consumer for a good interaction and other business activities involved. said, in fact, the company’s manufacturing such products have started to poll theses micro blogs to get a sense of general sentiment for a product. Many times these companies study user reactions and reply to users on micro blogs.

Furthermore, it is stated that machine learning methods can generate a fixed number of the most regularly happening popular words which assigned an integer value on behalf of the frequency of the word in Twitter[20].

Opinion Mining alludes to the expansive region of common language handling, content mining, computational semantics, which includes the computational investigation of assessments, feelings and feelings communicated in content In spite of the fact that, view or frame of mind dependent on feeling rather than reason is frequently informally alluded to as a slant .

Consequently, loaning to an equal for opinion mining or estimation investigation expressed that sentiment analysis areas including law, governmental issues, innovation and promoting. In prior days numerous online networking have given web clients road for opening up to express and impart their musings and insights.

1.2) Problem Statement :

As mentioned earlier, huge amount of data is available over the web especially social media like twitter. Analyzing this could be beneficial in predicting the next response on given extracted data. This could be beneficial over the following areas:

1.2.1) Science and Technology: The popularity of any scientific product/automobiles/home appliances can be known before its launch.

1.2.2) Cricket: In the IPL-auction the likelihood of a player getting into a team could be found also by checking the enthusiasm of the cricketer amongst audience.

1.2.3) Politics: Since the people are responsible in forming a government thus it could be easily predicted beforehand the most suitable candidate for the election purpose based upon theirs popularity as reflected over twitter.

1.3) Objectives:

Since the above stated terms are mostly related to public's general routine and highly populated everyday on social media sites. Therefore, the fields of science, politics, sports etc. are used in automatic sentiment classification of an unknown tweet stream.

1.4) Methodology :

(1) Twitter data is extracted through twitter API “tweepy” which is the official API for extracting tweets.

(2) The extracted data is analyzed through python at runtime, by means of mathematical functions for the average as well as standard deviation the variation of particular #Hashtag or “Search Term” is calculated and the results are stored in a “<search-word>.csv” file (that too at run time, if the file does not exist previously the o.s. dynamically creates on at run time).

(3) For the purpose of visualization of data the matplotlib library is used which helps in showing the 2-D plot for given number of days.

1.5) Organization :

The chapters of the work are divided into five chapters:

Chapter 1:Introduction

This chapter highlights of social media in modern life and the impact of it in finance, governance as well as marketing. It also describes about one of it’s biggest platform i.e. Twitter about some existing methods used for sentimental analysis.

Chapter 2 : Literature Review

Here we discuss from the research papers in journals and platforms such as those from ACM, IEEE, Springer, JIIT , Google Scholar etc. The diversity in these papers give idea about techniques involving ANN, convolution, etc.

Chapter 3: System Development

The hardware and software specifications are described here, moreover various libraries used for the task are discussed over here.

Chapter 4 : Results and Performance Analysis

The alignments of test set is covered here covering the day to day domains such as Technology, Cricket and politics

Chapter 5 : Conclusion

The summary covered is in this chapter which depicts the use of the platform for the mentioned tasks.

CHAPTER2 : LITERATURE SURVEY

Following papers were taken into consideration:

i) A Sentimental Education Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts[1]:

It starts by classifying a paragraph into two groups, first one being subjective sentences and the second being objective sentences. Only the subjective part is taken into analysis and the objective part is removed from the data set.

Separate group of classes are created of the data words, then using particular formula of individual and associated scores the sentiment-class scores are calculated for each word. The class where the partitioning cost is optimal is chosen as the base class for word.

ii) Domain-specific Sentiment Analysis using Contextual Feature Generation: [2]

Clue set contains most likely feature words for the word that is to be checked. This method has a four step algorithm for generating the new clue set of sentiment words. Initially this is consisting of sentence as well as polarity(training ex.) then generates corresponding clues. Second step involves the identification of sentiment topics from training (sample) example. Then the sentiment clues which gets identified to sentiment topics are putted to it's current clue set. This updated classifier is then used for identifying other domains under sentiment clues from the sample data set .

iii) Exploiting New Sentiment-Based Meta-level Features for Effective Sentiment Analysis: [3]

This method contains large set of pre-classified words/sentence. For a new word that has to be checked the features(which discuss issues for the work) are checked for similarities with the pre classified data set. The one with the most resembling contents matched is considered as the true sentiment word for the given content. Thus the given content's sentiment is taken as the sentiment of the pre classified sentence with whom the maximum feature similarities are got.

iv) Facebook Impact and Sentiment Analysis on Political Campaigns:[4]

The measure of sentiment of user towards each political party is checked using sentiment index which takes the log of ratio of total words raised to the power positive and negative added to unity.

$$Sentiment\ Index = ln\left[\frac{1 + TOTAL^{POSITIVE}}{1 + TOTAL^{NEGATIVE}}\right]$$

The study indicated that the emotions of voters can be found but not their intention to vote. It was found that the political party that won had a bad perception on social media while the one having good was unable to win the Mexican election.

v) SentiML: functional annotation for multilingual sentiment analysis:[5]

This approach allowed a multi level representation of three categories namely

i) Target

ii) Modifier

iii) Appraisal group.

Target is the expression that the sentiment refers to. Modifier is the expression conveying the sentiment and appraisal group includes the set of targets and modifiers.The meaning of a sentence gets reversed when certain words were present which are k/a modifiers. For classifying the identified attribute is classifies into it's categories.

vi) Evaluation of Features on Sentimental Analysis:[7]

This involves 4 step procedure. Firstly data is preprocessed, in it initially all text are converted into lowercase words for simplicity of feature extraction. Then the words ending with apostrophizes are converted back to original form like don't -> do not, any non ASCII character is removed. This is followed by removal of stop words (eg. a,an the) as the do not convey any feature so removal of

stop words is preferred. Second setup involves partitioning data into training and test data. For training data after performing stemming(removing suffixes like ing
eg.: computing->compute) feature selection is to be performed in which various statistical methods are applied to check whether sentiment of review can be extracted from the count of words in each sample.

vii) Review of Sentimental Analysis Methods using Lexicon Based Approach :[8]

This approach includes the calculation for the inclination of sentiments in words by checking their semantic alignment of words in the document .Phrases can also be used. Lexicon-based approach dictionaries can be created automatically as well as manually , with the help of seed words we can expand the list of words. This research was focused mainly on using adjectives as predictors of the semantic alignment of text. It starts with compilation of adjectives and their (SO) Sentiment Orientation into a dictionary and from there patterns are matched for sentimental calculation.

viii) Sentimental Analysis of Twitter Data using Text Mining and Hybrid Classification Approach: [9]

This begins with the extraction of tweets which is then pre-processed followed by application of a classifier algorithm. It involves 5 step mechanism. Firstly a structured data is formed, Then common grammar words like verbs ,preposition articles (c/a stop words) are removed. This is followed by the step where words ending with "ing","ize","ed" are reduced to their root word. This is known as steaming. The term frequency – inverse term frequency score is calculated for each term and a 2D matrix is created in which rows represent documents and columns represent the word extracted from document after the above preprocessing which is filled with the TF-IDE score.

ix) Sentimental Analysis of Flipkart reviews using Naïve Bayes and Decision Tree algorithm:[10]

Naïve Bayes approach for classifying is based on probabilistic classifying where Bayes Theorem is used and theory of total probability is used. The paper claims it's suitability for large datasets. Just like text mining it starts with tokenization then removal of stop words followed by text transformation. Then features are selected for the parts of document that contribute for positive and negative words. These parts are joined in such a way such that the probability is maximum of the resulting sentence existing in either of the two +ve, -ve terms.

x) Sentiment Analysis using Neuro-Fuzzy and Hidden Markov Models of Text:[11]

The data were taken from different files c/a data set which was combined into a one source file called "corpus". Once combined the text were converted into array of words and further step was to sort this array of collection. Like other works stemming was skipped so that the words like remind and reminded make different sense to the text in consideration. Now a final part of calculation of membership degree of each term is calculated by an analytical formula before being processed by the neural network.

xi) Rating Prediction Based on Social Sentiment From Textual Reviews:[12]

This paper has two type of list:

i) Word List: Contains both positive polarity and negative polarity words.

ii) Topic list: Contains list of topics that will be depicted as the root sentiment.

The model involves the following steps: Firstly pre-processing tasks such as stop word removal, noise word removal are performed and then remaining product features are extracted using Latent Dirichlet Allocation (which computes the relationship of reviews and topics with words). The generative process is followed that matches user words to a set of it's most probable topic list. Secondly sentiment degree is matched by the SDD(sentiment degree dictionary).From the combined word list(sentiment dictionary) from where a review is matched before the product feature and assigned a score +1.0,-1.0for positive and negative respectively. Finally the words with prefix words as negative from ND(negation dictionary) are checked and a default value coefficient of +1.0 c/a negation check coefficient is added the sentiment polarity is reversed , and the coefficient is set to −1.0 if the sentiment word is preceded by an odd number of negative prefix words and hence therefore the normalized score is generated.

xii) Part- Of- Speech tagging:[13]

It uses the HMM (Hidden Marker Model) that assumes data to be in unobserved states. In it the results are a probabilistic function of the classes to which the classification is to be performed. The data set are pre classified as objective or subjective using Naive Bayes Model. The Sequence recognition feature includes probabilistic distribution of subjectivity degree. The HMM parameters were calculated from them using specific equations and results stored.

xiii) SA using Binary to Multi-Class Classification :[14]

Binary Classification refers to classifying the work in either of two polarity i.e positive or negative. This paper's model works on a pattern based approach and had accuracy of 87.5% in binary classification .The author defined seven classes for which tweets were matched namely happy, sad, anger, love, hate, sarcasm and neutral.

$$\rho(t) = \frac{PW(t) - NW(t)}{PW(t) + NW(t)}$$

Emotional scores for words are calculated using Senti-Strength that assigns a score ranging from -1 to -5 and +1 to +5 based on the severity of the sentences. This is followed by getting a total of 4 parameters

i) Net score for PW (+ve words)

ii) Net score for NW(-ve words)

Using it ratio of emotional words is calculated as:

xiv) Opinion Mining: [15]

In this paper the most frequent words are to be generated as output in the form of a word cloud and it's size will depends upon the extent of the frequency of occurrence of words collected. This implements the sentimental analysis algorithm for searching the most used words in smartphone industry. The data set is taken from CSV (common separated value) file. The category to which these words used for is also checked. Each frequency is noted for the term.

Fig 2.1: Word Cloud

These words are displayed in a clustered manner and it's size depends upon the frequency meaning large frequency words are having size that is large in number.

xvi) Sentimental Analysis of Product Based Reviews Using Machine Learning Approaches: [18]

The method involves the collection of data from several e-commerce websites which are firstly preprocessed to remove words like 'or', 'that' and suffixes like 'ing' ,' ness' are reduced. This is passed to feature extractor that searches for a character like domain, product, feature and adjective. These features are cross checked with earlier supplied dataset whose correct polarity are known and for correct result the result are shown else the user has to again redraw the data set.

xvii) Sentimental feature selection for sentiment analysis of Chinese online reviews:[19]

This method includes dividing the text into segments, from where a sequence of sentimental features is selected. This selection is based upon the resemblance with training sample set. A weight for the feature is calculated and each set of feature is associated with its weight to generate a group containing text along with their weight. Thus reviews are structured together in this way and the polarity is checked by knowing the extent of the magnitude of the weight.

CHAPTER 3: SYSTEM DEVELOPMENT

3.1) System Requirement:

Hardware:

Data Bus: 64 bit

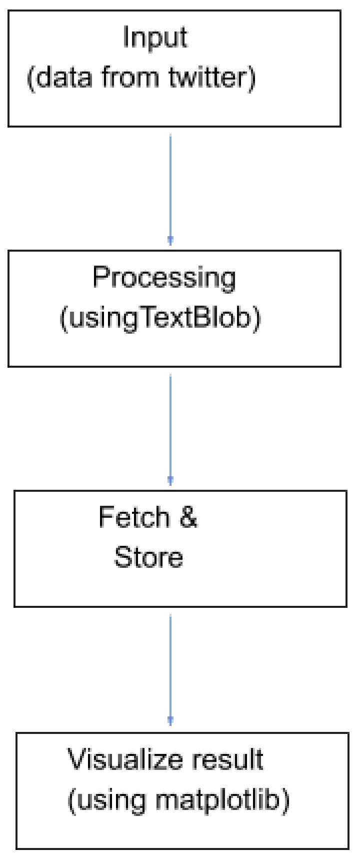
Processor: Intel Core i5 (or above)

Software:

Operating System: Windows 10

Python version : 3.7

3.2) Model Development :



Fig(3.1): System development

3.2.1) Input: The tweets are take these can include the live streams also or tweets generated in a particular date.

3.2.2) Processing: The textblob library is an open source library that is maintained in github and includes functions specifically for natural language processing tasks. Through it the characterization is either positive, negative or neutral is performed

3.2.3) Fetch & Store: On run time when the results are calculated then the need to be stored in file to be viewed at a later stage. Thus those results are stored in csv format and the file name is the search-term itself.

3.2.4) Data Visualization: The matplotlib is used to extract data from the saved files to show the variation in positive and negative w.r.t their creation dates.

3.3) Libraries Used:

1) Tweepy
2) Textblob
3) Matplotlib

3.3.1) tweepy: Tweepy is the official Programming interface for dealing with twitter related works.It was utilized for getting tweets. This Programming interface has techniques for doing retweets, refresh status, seek terms, get client data, getting devotees and so forth.

- **Classes Utilized:**

i) API : This class has methods to confirm an application to interact with twitter.

ii) OAuthHandler: It takes customer key and buyer mystery as parameters that are novel for every application.

iii) cursor : This class can play out all the page association related works. This was utilized in getting tweets from a few courses of events.

- **Methods utilized:**

i) set_access_token: It takes get to token and access token mystery as parameters and is utilized in verifying that the 4 parameters relate to a substantial application.

ii) Cursor : The programming interface protest was passed as first contention and the term to be sought as second .

iii) items() : This was utilized to process result per page for the tweets.

Arguments : noOfSearchTerms: The estimation of terms to coordinate.

3.3.2.) textblob: It is that library that has vast methods for the purpose ofdata processing.

- **Classes Utilized :**

i) Sentiment: The examination class gives techniques that can be utilized in getting conclusion from words or some other dialect handling.

- **Method Utilized:**

ii) Polarity: It distinguishes extremity from a content that is inputted in it's bracket.

3.3.3.) Matplotlib :

Matplotlib is accounts for the graphical visualization of data for 2D plotting .

- **Classes utilized:**

i) pyplot: This incorporates capacities to give a pictorial perception of the inputted information.

It was utilized in producing the pie outline.

- **Methods Utilized:**

i) pie : Produces pie outline for first contention with hues in second contention.

ii) legend() : Creates legend that contains depiction identified with shading.

iii) title() : Assign title to outline for better comprehension of pie chart.

iv) axis() : sets hub information limit for the sort of diagram being utilized.

v) light_layout() : Naturally change subplot parameters to give cushioning.

vi) appear() : Presentations a figure , here it shows the pie outline with title, legend as provided.

CHAPTER 4:
RESULTS AND PERFORMANCE ANALYSIS

4.1) Test Plan:

4.1.1)New Set:

Here both the keyword variable and the number of tweet variable are reset and manually stored.

4.1.2)Same Key Words but Different number of tweets:

Keeping the keyword/hashtag same user is asked to input a new number of tweets to be matched.

4.1.3)Different Key Words but same number of tweets :

Keeping the number of matching words/ hashtags same the value of them is to be re entered that will be used in getting the sentiments reflected over.

4.2) Classification:

Results are classified over 2 categories:

4.2.1) Variable search set.

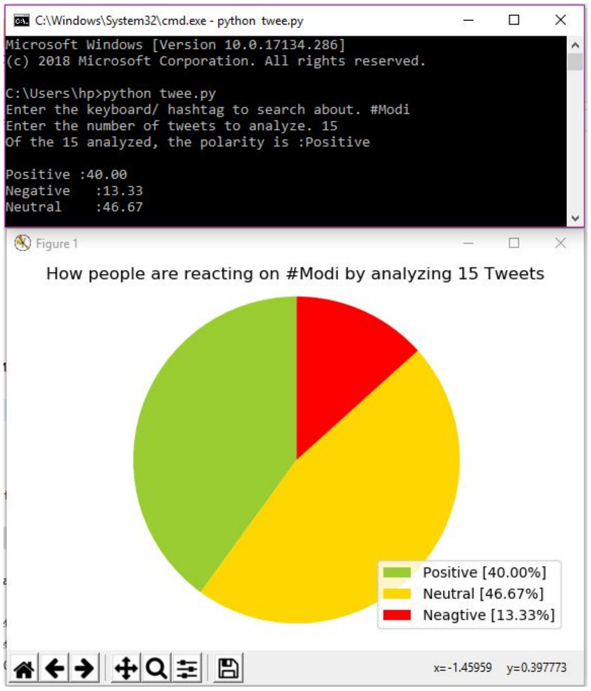
4.2.2) Continuous search set

4.2.1) Variable search set:

Using the test plan as discussed in chapter: 5 the value of trending words are matched also the number of tweets for matching is varied and the results are plotted. Since the ratio of positive, negative and neutral are combined to form the whole set so pie chart is chosen to depict their corresponding shares in respective categories.

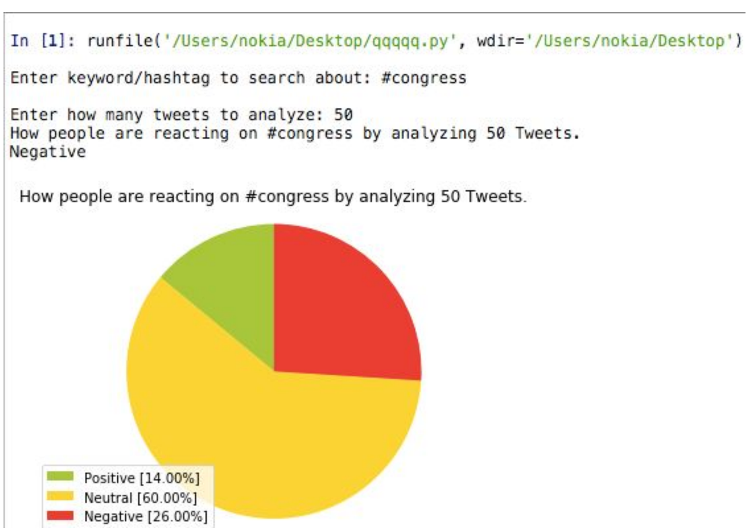
Results :

4.2.1.1) Test Case#1:



Fig(4.1): Test Case #1

4.2.1.2) Test Case#2:



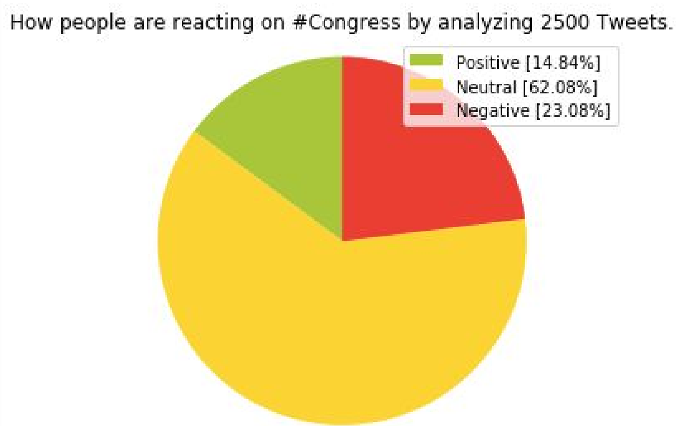
Fig(4.2): Test Case #2

4.2.1.3) Test Case #3:

```
In [2]: runfile('/Users/nokia/Desktop/qqqqq.py', wdir='/Users/nokia/Desktop')

Enter keyword/hashtag to search about: #Congress

Enter how many tweets to analyze: 2500
How people are reacting on #Congress by analyzing 2500 Tweets.
Negative
```



Fig(4.3) Test Case #3

4.2.2) Continuous search set:

The following domains are collected from twitter by setting the tweet limit to hundred and the results are plotted for the starting week of May for the domains underlined.

4.2.2.1) Cricket

1.1 #ChampionsLeagueFinal

1.2 Kings XI Punjab

1.3 Ipl 2019

4.2.2.2) Technology

2.1 Fog Computing

2.2 5G

2.3 #OnePlus7Pro

2.4 Cryptocurrency

4.2.2.3) Politics

3.1 #2019election

3.2 Rahul Gandhi

3.3 Narendra Modi

3.4 Arvind Kejriwal

1) Cricket:

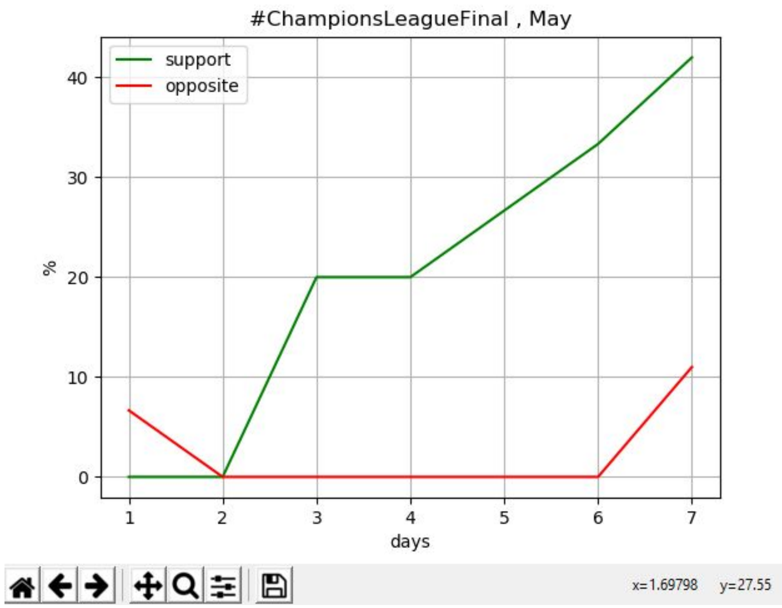
1.1) #ChampionsLeagueFinal

Day	Positiv e	Negativ e
1	0	4.67
2	0	0
3	20	0
4	20	0
6	33.33	0
7	42	11

	Positive	Negativ e
Mean	19.2216 7	2.945
St. Dev	14.5887 5	4.34838 6
Maximu m	42	11
Minnimu n	0	0

Table(4.1): Variation of tweets on : #ChampionsLeagueFinal for the first week of May.

Table(4.2): Mathematical analysis of #ChampionsLeagueFinal graph.



Fig(4.4): Variation of tweets on : #ChampionsLeagueFinal for the first week of May

Observation:

From the graph we can conclude that the maximum positive tweets recorded were on May 7 with an estimate of 42 %. The curve for positive is in increasing type of graph with average recorded value approximately 20% and maximum opposition tweets were equal to 11% on the same day as for supporting event.

1.2) Kings XI Punjab:

Day	Positive	Negative
1	28.64865	14.13514
2	30.0885	11.50442
3	32.84615	4.076923
4	39.00709	10.28369

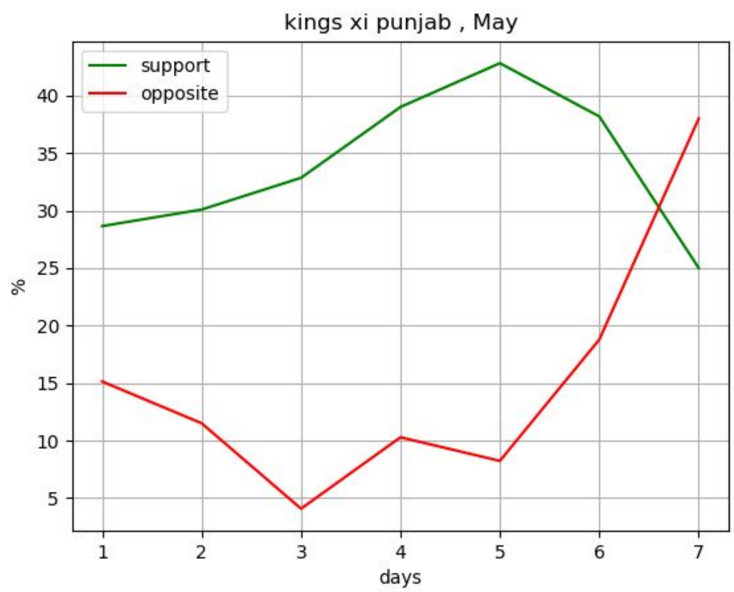
	Positive	Negative
Average	33.7983	14.14559
St.Dev	4.931295	10.30127
Max	42.81588	38

5	42.81588	8.231047
6	38.18182	18.78788
7	25	38

Table(4.3): Variation of tweets on : Kings XI Punjab for the first week of May.

Min	25	4.076923

Table(4.4): Mathematical analysis of the Kings XI Punjab graph.



Fig(4.5): Variation of tweets on : Kings XI Punjab for the first week of May.

Observation:

There has been an increase in tweets for opposition for this team which reached a maximum of 38% from 8.2 % in just an interval of two days only, moreover the support reduced from 42% to just 25% that was a decline in around 20 % of popularity while parallel the increase in opposition had around 30 %.

1.3) IPL 2019

Day	**Positive**	**Negative**

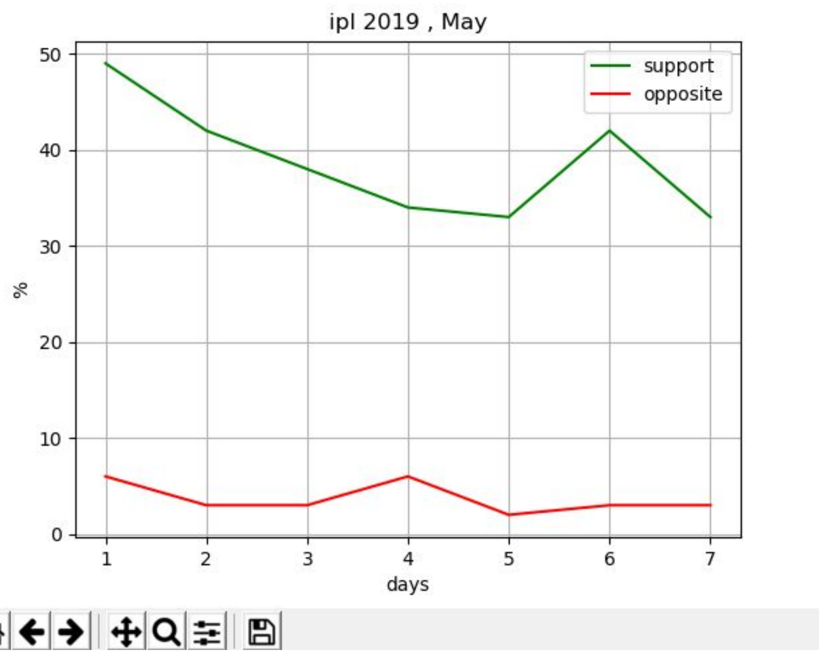
	Positive	**Negative**

1	49	6
2	42	3
3	38	3
4	34	6
5	33	2
6	42	3
7	33	3

Table(4.5): Variation of tweets on : IPL 2019 for the first week of May.

Average	38.71429	3.714286
St.Dev	4.547568	1.484615
Max	49	6
Min	33	2

Table(4.6): Mathematical analysis of the IPL 2019 graph.



Fig(4.6): Variation of tweets on : ipl2019 for the first week of May.

Observation:

The overall craze for IPL as refleted from tweets had been in it'sfavour , the tweets in opposition could hardly had an average of less than 4%. On the other hand the maximum supporting tweets were close to 50 % on the start of the month while the minimum being at a fraction of 33% in support and 2% for non favourable tweets

Inference:

As of May7, we conclude that the sentiments(average) for the search term: “IPL 2019” were the highest amongst the 3 terms takes.

IPL 2019	38.0
Kings XI Punjab	33.7
#ChampionsLeagueFinal	19.2

Table(4.7) : Average comparison of Cricket related area.

On the other hand the maximum positive sentiments recorded were for IPL 2019.

IPL 2019	49
#ChampionsLeagueFinal	42.8
Kings XI Punjab	42

Table(4.8) : Maximum recorded positive sentiments for Cricket.

The order of opposition tweets is as follows:

Kings XI Punjab	38
#ChampionsLeagueFinal	33.7
IPL 2019	19.2

Table(4.9) : Minimum recorded conflicting sentiments for Cricket related area.

2) Technology:

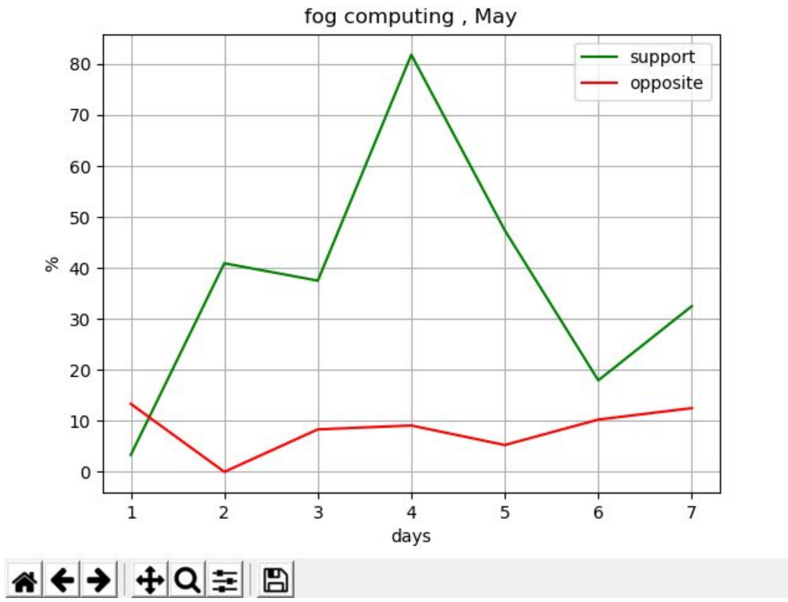
2.1) Fog Computing:

Day	Positive	Negative
1	3.333333	13.33333
2	40.90909	0
3	37.5	8.333333
4	81.81818	9.090909
5	47.36842	4.263158
6	17.94872	10.25641
7	32.5	12.5

Table(4.10): Variation of tweets on : Fog Computing for the first week of May.

	Positive	Negative
Average	37.33968	8.396735
St.dev	22.82712	4.233802
Max	81.81818	12.5
Min	3.333333	0

Table(4.11): Mathematical analysis of the Fog Computing graph.



Fig(4.7): Variation of tweets on : Fog Computingfor the first week of May.

Observation:

Fog computing is the emerging art of technology that has the potential in replacing cloud computing sooner or later. From, the data collected it was revealed that the maximum popularity reflected was 83 % a standard deviation of 22 % is sufficient in revealing that there had been an high variation for it's support since the minima was 3.3% recorded i.e a very high difference of 80% .

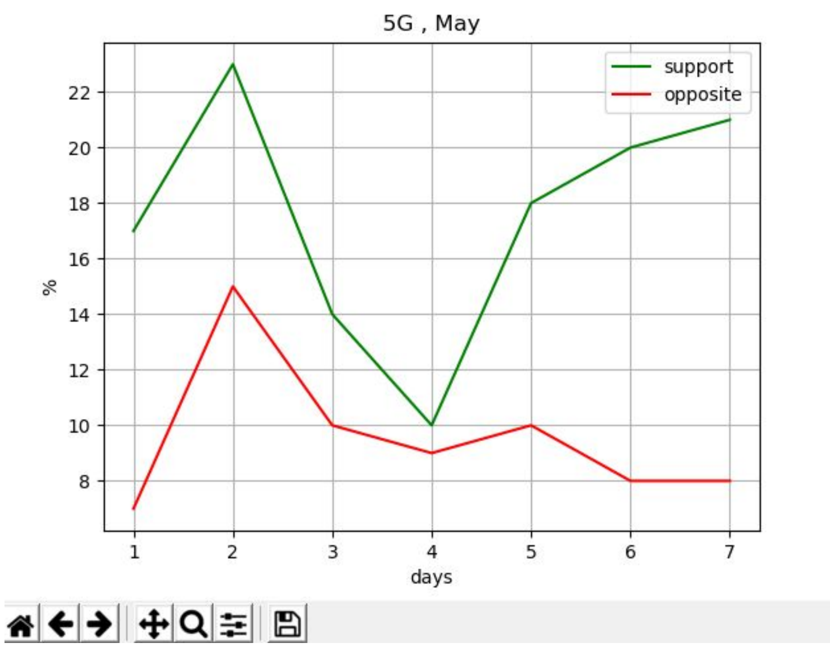
2.2) 5G:

Day	Positive	Negative
1	17	7
2	23	15
3	14	10
4	10	9
5	18	10
6	20	8
7	21	8

Table(4.12): Variation of tweets on : 5G for the first week of May.

	Positive	Negative
Average	17.57143	9.571429
St.dev	4.100771	2.441144
Max	23	15
Min	10	7

Table(4.13): Mathematical analysis of the 5G graph.



Fig(4.8): Variation of tweets on : 5G for the first week of May.

Observation:

On each interval taken, the tweets counted in opposition were not more than the tweets considered in it'ssupport . Maximum recorded tweets covered an altitude of 15% in opposition while the one in support were 23% as recorded. Taking the minimum value in account the number was 7% and 10% in opposition and support respectively.

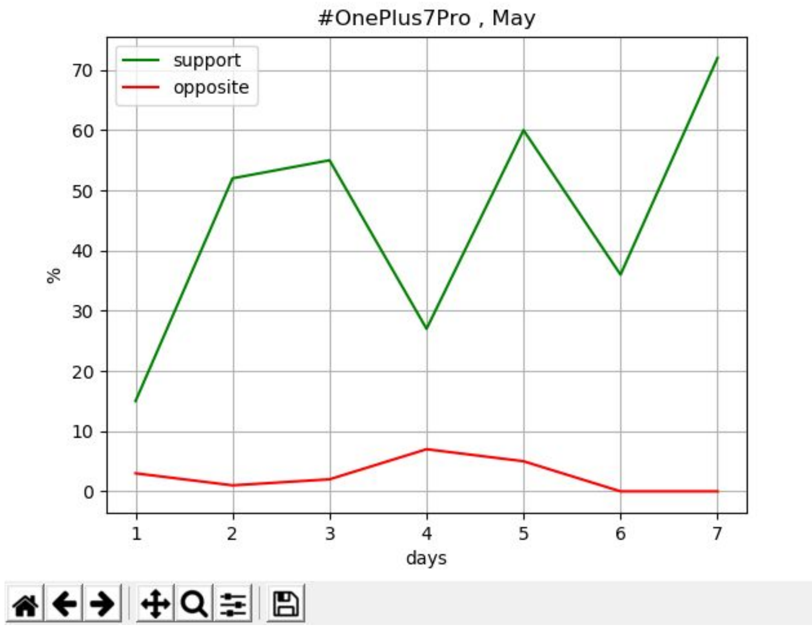
2.3) #OnePlus7Pro

Day	Positive	Negative
1	15	3
2	52	1
3	55	2
4	27	7
5	60	5
6	36	0
7	72	0

Table(4.14): Variation of tweets on : #OnePlus7Pro for the first week of May.

	Positive	Negative
Average	44.28571	2.571429
St.Dev	18.54504	2.441144
Max	72	7
Min	15	0

Table(4.15): Mathematical analysis of the #OnePlus7Pro graph.



Fig(4.9): Variation of tweets on : #OnePlus7Pro for the first week of May.

Observation:

It appears that this handheld device will be one of the next most success in device technology. Making place in top 10 trending hashtags with good support of an average 45% tweetsin support against 2.5% in opposition and the maximum support of 72% when no tweets were recorded in opposition reveals the fact of it'sbecocmming a grand success for technological goods.

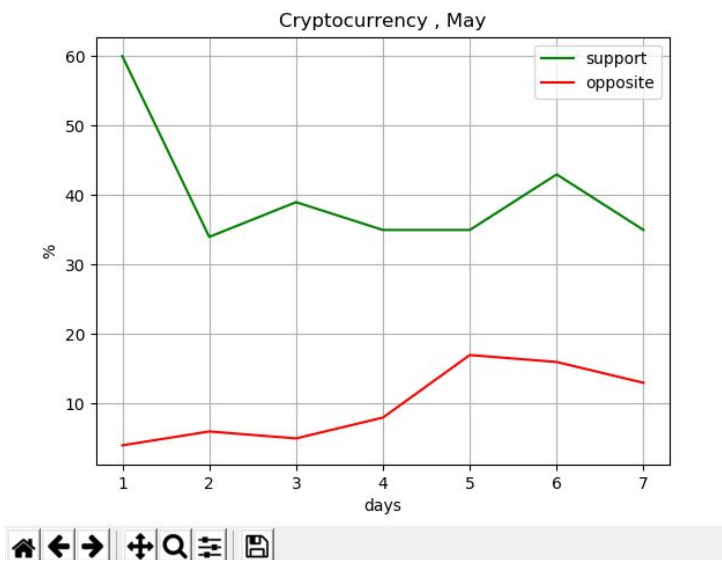
2.4) Cryptocurrency:

Day	Positive	Negative
1	60	4
2	34	6
3	39	5
4	35	8
5	35	17
6	43	16
7	35	13

Table(4.16): Variation of tweets on : Cryptocuency for the first week of May.

	Positive	Negative
Average	40.14286	9.857143
St.Dev	8.626017	4.997959
Max	60	17
Min	34	4

Table(4.17): Mathematical analysis of the Cryptocurrency graph.



Fig(4.10): Variation of tweets on : Cryptocurrency for the first week of May.

Observation:

Cryptocurrency also showed on each sample interval a net positive support towards it also the maximum opposition[17%] is even half of minimum favor[34%] as recorded.

Cryptocurrency could help inn contributing in saving the global warming scenario also since using this technology could replace the current use of paper-based currency for which to manufacture a lot of damage had been caused to the ecosystem. Thus the support reflected form the platform of colleting tweets is true.

Inference:

From the weekly analysis we conclude that fog computing got the maximum positive tweets on Saturday i.e May 4 while the average analysis showed that

#OnePlus7Pro	44.28 %
Cryptocurrency	40.14%
Fog computing	37.3%
5G	17.57%

Table(4.18) : Minimum recorded conflicting sentiments for Cricket related area.

Fog computing	81.8 %
#OnePlus7Pro	72.0%
Cryptocurrency	60.0%
5G	23.0%

Table(4.19) : Maximum recorded positive sentiments for Cricket.

#OnePlus&Prohad the most highlighted positive with maximum reaching to 72% on May 7.

In terms of variation

Cryptocurrency	17.0 %
5G	14.0%
Fog Computing	12.5 %
#OnePlus7Pro	7.0 %

Table(4.20) : Average comparison of Cricket related area.

We see that 5G had the minimum standard deviation and the maximum of opposite were close to the average in positive.

3) Election:

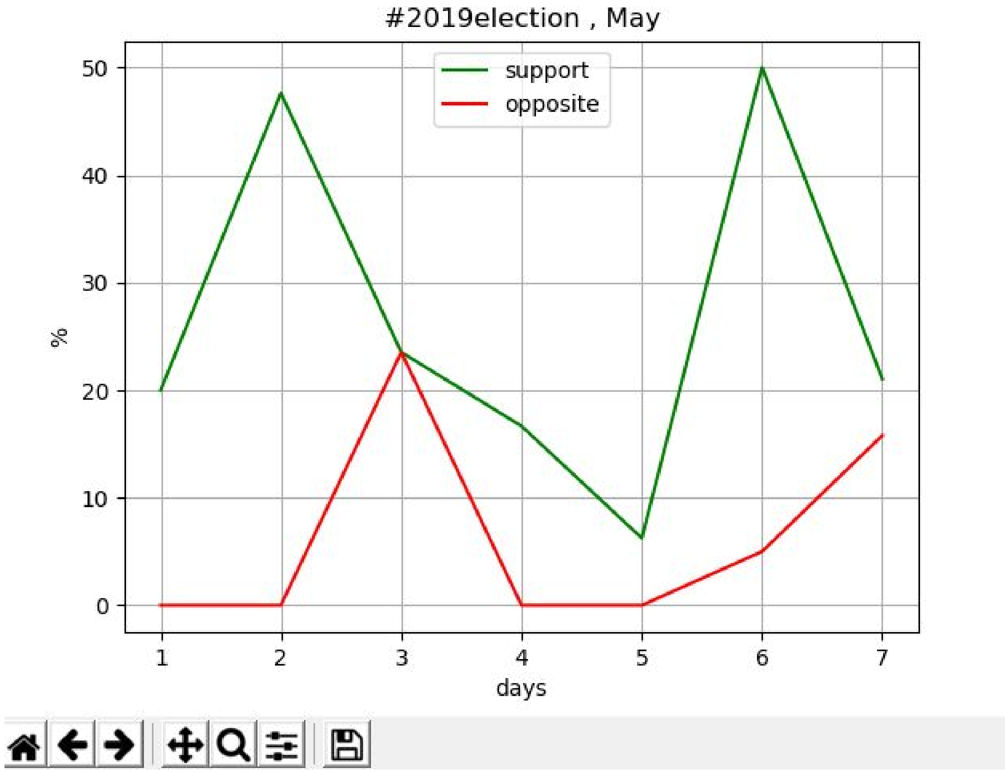
3.1) #2019Election:

Day	Positive	Negative
1	20	0
2	47.61905	0
3	23.52941	23.52941
4	14.66667	0
5	4.25	0
6	50	5
7	21.05263	14.78947

Table(4.21): Variation of tweets on : #2019Election for the first week of May.

	Positive	Negative
Average	24.44539	4.331269
St.Dev	14.0511	8.842639
Max	50	23.52941
Min	4.25	0

Table(4.22): Mathematical analysis of the #2019Election graph.



Fig(4.11): Variation of tweets on : #2019Election for the first week of May.

Observation: This topic comes under the most prestigious and debatable current issues for the month as the elections are undergoing too in the month of May, the publibs sentiments collected as and aomputed showed that the maximum favourable support was 50 % while that for opposition was 23.5%. There had been 4 intervals when no tweets were recorder in the later category that reflects a chance for a peaceful public reaction in the election areas.

3.2) Rahul Gandhi:

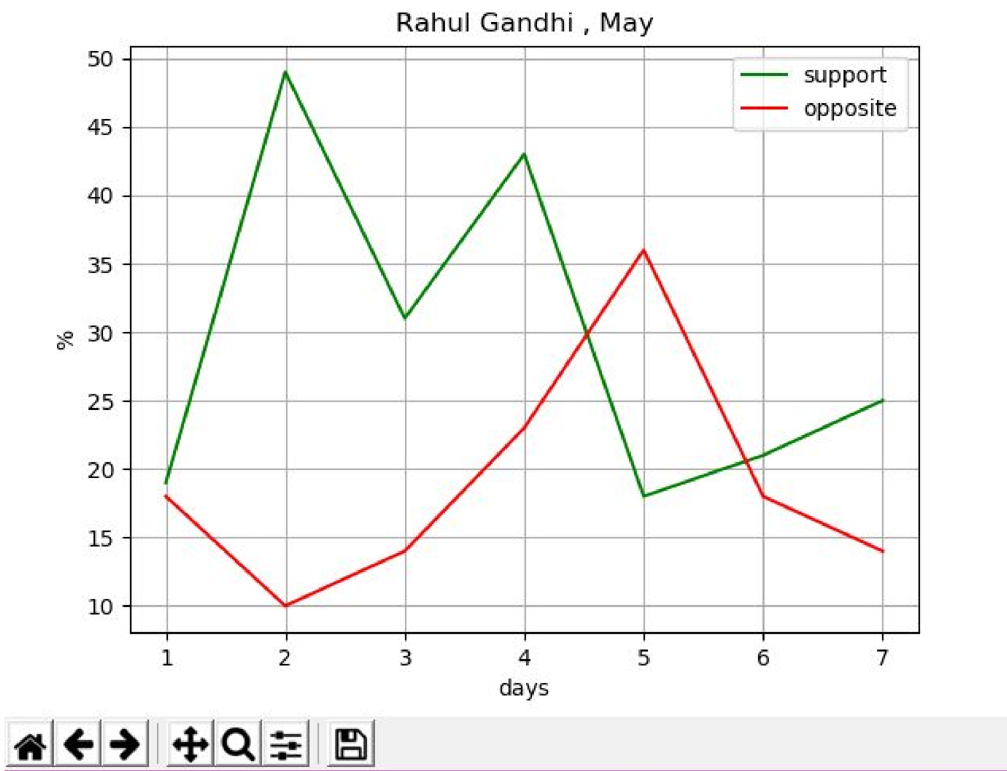
Day	Positive	Negative
1	19	18
2	49	10
3	31	14
4	43	23
5	18	36
6	21	18
7	25	14

Table(4.23): Variation of tweets on :

	Positive	Negative
Average	29.42857	19
St.dev	11.33713	7.91021
Max	49	36
Min	18	10

Table(4.24): Mathematical analysis of

Rahul Gandhi for the first week of May. thetweets on Rahul Gandhi graph.



Fig(4.12): Variation of tweets on : Rahul Gandhi for the first week of May.

Observation:

For the starting half the tweets in favor were much higher as compared to those in opposition.

The transition occurred on 5th when the favourable [18%] were merely half of those tweets collected in opposition [36%] and that sample accounted for the most tweets in opposition.

For the favourable the maxima was 49% and an average of 30%(approx.) was recorded.

3.3) Narendra Modi:

Day	Positive	Negative
1	12.5	12.5
2	7.7	0.0
3	11.1	0.0
4	24.08	8.7

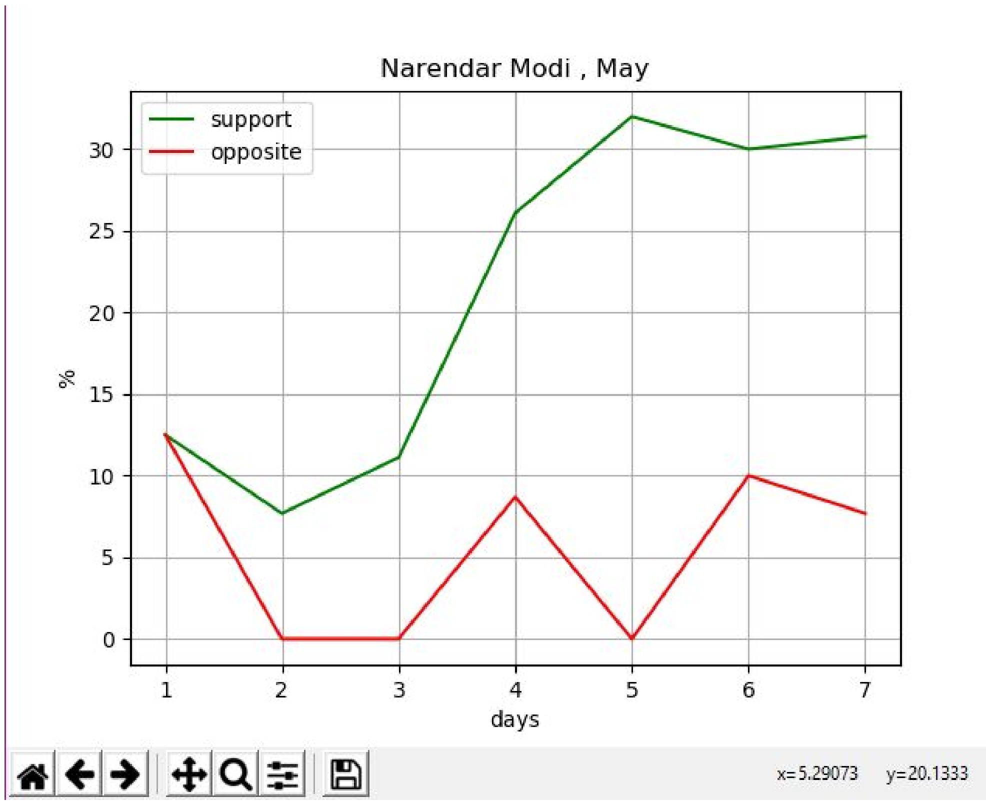
	Positive	Negative
Average	21.44	4.55
St.dev	9.6	4.0
Max	32.0	12.5

5	32.0	0.0
6	30.0	10.0
7	30.7	7.7

Table(4.25): Variation of tweets on : Narendra Modi for the first week of May.

Min	7.7	0.0

Table(4.26): Mathematical analysis of the tweets on Narendra Modi graph.



Fig(4.13): Variation of tweets on : Narendra Modi for the first week of May.

Observation:

Namo being the PM too had on maximum interval tweets in support for as compared to those posted not in favour. In this curve the maximum reflected as posted were 32 % while in the other re were merely 12.5% .

3.4) Arvind Kejriwal

Day	Positive	Negative

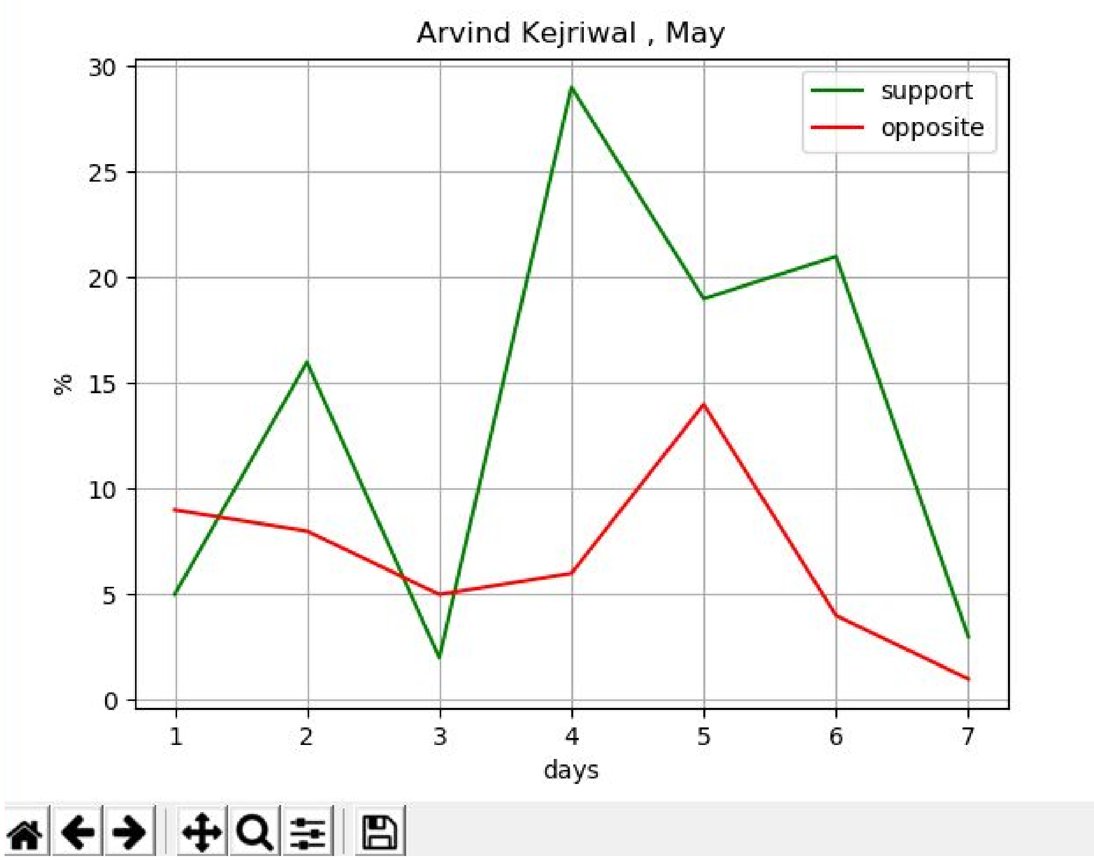
	Positive	Negative

1	5	9
2	16	8
3	2	5
4	29	6
5	19	14
6	21	4
7	3	1

Table(4.27): Variation of tweets on : Arvind Kejriwal for the first week of May.

Averag e	13.5714 3	4.71428 6
St.Dev	9.61928 3	3.84389 3
Max	29	14
Min	2	1

Table(4.28): Mathematical analysis of the tweets on Arvind Kejriwal graph.



Fig(4.14): Variation of tweets on : Arvind Kejriwal for the first week of May.

Observation: The CM of Delhi had a mix reaction on the twitter on some days' the positive were more and increasing while other it was the opposite situation reflected. With a maximum support received of 29% showed that this election his popularity had been not that high as it were in the previous elections.

Inference:

From the tweets over politicians collected we see the order of average collected for seven consecutive days over politicians as:

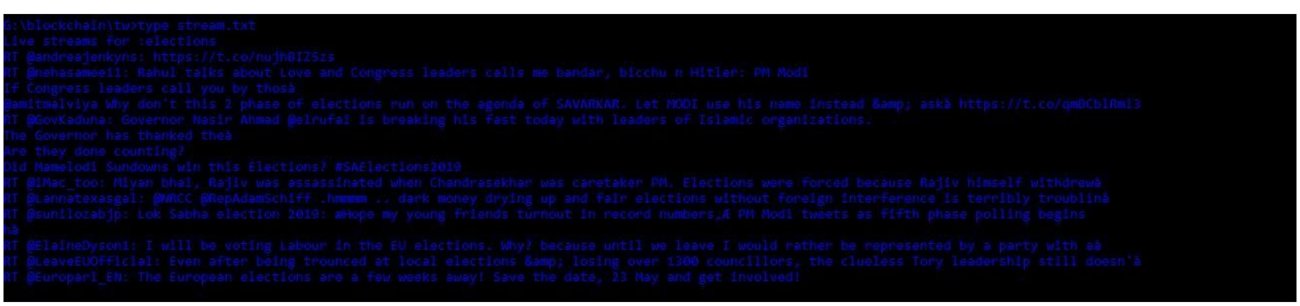
Rahul Gandhi	29.4 %
Narendra Modi	27.14 %
Arvind Kejriwal	13.5 %

Table(4.29) : Average comparison of Politicians.

This showed that in the starting of May month the popularity of Rahul Gandhi was higher then other some competing political leaders.

On May 5 the tweets against each candidate were one amongst the highest in “opposition” and for Narendra Modi(27%) were highest among collected and there was a steep decline on the positive ones(10%)

Along with Example of Live Tweets Collected over term: “election”



```
G:\blockchain\tw>type stream.txt
Live streams for :elections
RT @andreajenkyns: https://t.co/nujh8IZ5zs
RT @nehasameel1: Rahul talks about Love and Congress leaders calls me bandar, bicchu n Hitler: PM Modi
If Congress leaders call you by thosâ
@amitmalviya Why don't this 2 phase of elections run on the agenda of SAVARKAR. Let MODI use his name instead &amp; askâ https://t.co/qmBCblRml3
RT @GovKaduna: Governor Nasir Ahmad @elrufai is breaking his fast today with leaders of Islamic organizations.
The Governor has thanked theâ
Are they done counting?
Did Mamelodi Sundowns win this Elections? #SAElections2019
RT @iMac_too: Miyan bhai, Rajiv was assassinated when Chandrasekhar was caretaker PM. Elections were forced because Rajiv himself withdrewâ
RT @Lannatexasgal: @NRCC @RepAdamSchiff .hmmmm .. dark money drying up and fair elections without foreign interference is terribly troublinâ
RT @sunilozabjp: Lok Sabha election 2019: #Hope my young friends turnout in record numbers,Â PM Modi tweets as fifth phase polling begins
hâ
RT @ElaineDyson1: I will be voting Labour in the EU elections. Why? because until we leave I would rather be represented by a party with aâ
RT @LeaveEUOfficial: Even after being trounced at local elections &amp; losing over 1300 councillors, the clueless Tory leadership still doesn'â
RT @Europarl_EN: The European elections are a few weeks away! Save the date, 23 May and get involved!
```

Fig(4.15): Live streaming for “election”.

CHAPTER 5: CONCLUSION

4.3.1) Conclusions:

Through this paper we present a different way of analysis for prediction for topics like product launch success, political campaigns etc.

In general the opinions were collected through polling each individuals from regions to know the status of their taste This method consumed a lot of human work and was tedious too.

In contrast to that if the required work is carried from the largest social network then it could be efficiently done also, we can analyze the trend in the same as collection of tweets from previous posts is still feasible.

As we have seen their occurs a vast variation of polarity and sentiments collected over time that helps in governing the mood of global people across different geographical locations. Moreover some words like study showed zero negative ratios.

4.3.2) Future Scope:

The project can be combined with Operating system kernel so as to provide a better interface with the users and pages that contain data mining tools can be better implemented to govern the variation of keywords/hash tags over time.

4.3.3) Applications:

- Political Campaign strategy management
- Product launch
- Trend analysis
- Most popular sentiments

REFERENCES:

[1]:B. Pang and L. Lee(2008), "Opinion mining and sentiment analysis" . Foundations and Trends in Information Retrieval 2(1-2), pages 1-135.

[2]: Y. Kim and S. Myaeng. "Opinion analysis based on lexical clues and their expansion" . In Proceedings of 6th NTCIR Evaluation Workshop, 2007.

[3]: M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi. "Measuring User Influence in Twitter: The Million Follower Fallacy". Int'l AAAI Conference on Weblogs and Social Media (ICWSM), 2010.

[4]: Rodrigo Sandoval-Almazan, David Valle-Cruz , "Facebook impact and sentiment analysis on political campaigns on Political rallies",19th Annual International Conference -May 2018.

[5]: Marilena Di Bar,S Sharoff, Martin Thomas , "SentiML: functional annotation for multilingual sentiment analysis", ACM, Florence, Italy, September 2013.

[6]: Justin Martineau, and Tim Finin, "Delta TFIDF An Enhanced Element Space for Notion Examination",In Third International ICWSM Conference, Pg 258-261, November 2009.

[7]: Shahana P.H,Bini Ommanb , "Evaluation of Features on Sentimental Analysis", International Conference on Information and Communication Technologies, Pg 1585-1592, 2014.

[8]: Rahul Rajput, Arun Kumar Solanki, "Review of Sentimental Analysis Methods using Lexicon Based Approach", International Journal of Computer Science and Mobile Computing, IJCSMC, Vol. 5, Issue. 2, February 2016, pg.159 – 166.

[9]: Shubham Goyal, "Sentimental Analysis of Twitter Data using Text Mining and Hybrid Classification Approach", International Journal of Advanced Research, Ideas and Innovations in Technology, Volume2, Issue 5, 2016.

[10]: Gurneet Kaur, Abhinash Singla, "Sentimental Analysis of Flipkart reviews using Naïve Bayes and Decision Tree algorithm", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5, Issue 1, January 2016.

[11]: Zhen Hai, "Analyzing Sentiments in One Go: A Supervised Joint Topic Modeling Approach", IEEE Transactions on Knowledge and Data Engineering, Vol 29, pg: 1172-1185.

[12]: Yazhi Gao, Wenge Rong, Yikang Shen, Zhang Xiong, "Convolutional Neural Network based sentiment analysis using Adaboost combination", International Joint Conference on Neural Networks (IJCNN), 2016, Pg 1333-1338.

[13]:Shilpi Chawla, Gaurav Dubey, Ajay Rana, "Product opinion mining using sentiment analysis on smartphone reviews", 6th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), Sep. 20-22, 2017, AIIT, Amity University Uttar Pradesh, Noida, India, Pg 378-383.

[14]: Xiaojiang Lei, Xueming Qian and Guoshuai Zhao, "Rating Forecast Dependent on Social Slant From Literary Surveys",Volume 18 Issue 9, September 2016 .

[15]: Bouazizi, M., & Ohtsuki, T.A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter. *IEEE Access*. https://doi.org/10.1109/ACCESS.2017.2740982

[16]: K. R. Aida-zade, S. S. Rustamov, E. E. Mustafayev, and N. T. Aliyeva, "Human-Computer Dialogue Understanding Hybrid System, " IEEE Xplore, International Symposium on Innovations in Intelligent Systems and Applications. Trabzon, Turkey, Jule 2012, pp. 1-5.

[17]:Najkar, Negin & Razzazi, Farbod & Sameti, Hossein. (2010). "A novel approach to HMM based speech recognition system using particle swarm optimization". Mathematical and Computer modeling. 52. 1910-1920. 10.1016/j.mcm.2010.03.041.

[18]: Manvee Chauhan ,Divakar Yadav , "Sentimental Analysis of Product Based Reviews Using Machine Learning Approaches", Journal of Network Communications and Emerging Technologies (JNCET) Volume 5, Special Issue 2, December (2015).

[19]: Lijuan Zheng, Hongwei Wang, Tune Gao, Sentimental feature selection for sentiment analysis of Chinese online reviews, International Journal of Machine Learning and Cybernetics, Springer

[20]:https://www.nielsen.com/in/en/insights/news/2012/social-media-report-2012-social-media-comes-of-age.html,[Accessed: May 10'19]

[21]: https://monkeylearn.com/sentiment-analysis, [Accessed: May 10'19]