

TWITTER SENTIMENTS ANALYSIS

**A THESIS SUBMITTED TO THE GRADUATE
SCHOOL OF APPLIED SCIENCES
OF
NEAR EAST UNIVERSITY**

**By
MUNTAZAR MAHDI CHANDIO**

**In Partial Fulfillment of the Requirements for
the Degree of Master of Science
in
Computer Engineering**

NICOSIA, 2019

**MUNTAZAR MAHDI
CHANDIO**

TWITTER SENTIMENTS ANALYSIS

**NEU
2019**

TWITTER SENTIMENTS ANALYSIS

**A THESIS SUBMITTED TO THE GRADUATE
SCHOOL OF APPLIED SCIENCES OF
NEAR EAST UNIVERSITY**

**By
MUNTAZAR MAHDI CHANDIO**

**In Partial Fulfillment of the Requirements for
the Degree of Master of Science
in
Computer Engineering**

NICOSIA, 2019

Muntazar Mahdi Chandio: TWITTER SENTIMENTS ANALYSIS

**Approval of Director of Graduate School of
Applied Sciences**

Prof.Dr.Nadire CAVUS

**We certified this thesis is satisfactory for the award of the degree of Master of Science in
Computer Engineering**

Examining Committee in Charge:

Prof.Dr.RahibAbhiyev

Committee Chairman, Department of Computer
Engineer, NEU

Assist.Prof.Dr.Umit ILHAN

Department of Computer Engineering, NEU

Assoc.Prof. Dr. Melikesah

Supervisor, Department of Computer Engineering,
NEU

Assoc.Prof.Dr.YoneyKirsal

Department of Software Engineering, NEU

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name:

Signature:

Date:

ACKNOWLEDGMENTS

First and foremost, I give my thanks to an understanding supervisor Assoc. Prof. Dr. MelikeSahDirekoglu for her support, directions and for providing me guidance to start and complete this research.

I would like to express my deepest appreciation to my family especially my sister Muntazar Fatima and mother Amir Fatima, for providing me unfailing support, encouragement and always stand with me.

Thank you.

Muntazar Mahdi,

To my Family and Friends...

ABSTRACT

This study, we aim to analyze the effectiveness of social sentiments in the field of politics prediction. Twitter is the essential social network for sentiments analyzing and it provides useful information of mining data. In this study, we analyze whether social sentiments can be utilized for the prediction of election results. In particular, we analyze Twitter sentiments about Brexit and United Kingdom (UK) politicians as well as Pakistan politicians. Through periods, we collected Twitter data about Brexit, UK and Pakistan politicians using Twitter Application Program interface (API). First, we cleaned and pre-processed Tweet data for sentiment analysis. Then, we create a Twitter search and sentiment visualization interface using Python. Python provides useful libraries for sentiment analysis and graphical presentations. Finally, we analyze the changing opinions about Brexit, UK and Pakistan politicians using sentiments. In particular, in advance, we were able to correctly predict the UK parliament voting results in January 2019. In this thesis, we discuss Twitter data collection, Twitter sentiment search/visualization interface and detailed sentiment analysis results about Brexit, UK and Pakistan politicians.

Keywords: Twitter; sentiment analysis; Brexit; graph visualization; natural language processing; python; social media

ÖZET

Bu çalışmada, sosyal medya üyelerinin politik tahmin alanındaki etkinliğini analiz etmeye çalışılmaktadır. Twitter, duygusal analizleri için temel sosyal medya verimadenciliği için faydalı veriler sağlar. Bu çalışmada, seçim sonuçlarının tahmini için sosyal medya uygulamalarının kullanılması amaçlanmaktadır. Özellikle, Brexit ve Birleşik Krallık (İngiltere) siyasetçilerinin yanı sıra Pakistan siyasetçileri hakkındaki Twitter duygularını analiz ediyoruz. Dönemler boyunca Twitter Uygulama Programı arayüzünü (API) kullanarak Brexit, İngiltere ve Pakistanlı politikacılar hakkında Twitter verilerini topladık. İlk olarak, duyarlılık analizi için Tweet verilerini temizledik ve önceden izledik. Ardından Python kullanarak bir Twitter araması ve duyarlılık görselleştirme arayüzü oluşturuyoruz. Python, duyarlılık analizi ve grafik sunumları için faydalı kütüphaneler sağlar. Son olarak, Brexit, İngiltere ve Pakistanlı politikacılar hakkındaki deneyim ve görüşleri duyarlılık kullanarak analiz ettik. Özellikle, Ocak 2019'da İngiltere parlamentosu oylama sonuçlarını da bir ekil tahmin edebildik. Bu tez çalışmasında Twitter veri toplama, Twitter duyarlılık arama / görselleştirme arayüzü ve Brexit, UK ve Pakistan politikacıları hakkında ayrıntılı duyarlılık analiz sonuçları ele alıyoruz.

Anahtar Kelimeler: Twitter; duyarlılık analizi; Brexit; grafik görselleştirme; doğal dil işleme; Python; sosyal medya

TABLE OF CONTENTS

ACKNOWLEDGMENTS	ii
ABSTRACT	iv
ÖZET	v
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS	x
CHAPTER 1:INTRODUCTION	
1.1 Background.....	1
1.2 Aims and Objects.....	2
1.3 Motivation.....	3
CHAPTER 2:LITERATURE REVIEW	
2.1 Data Science.....	4
2.2 Social Media	5
2.3 Twitter.....	6
2.4 Python	7
2.5 Sentiment analysis research	9
CHAPTER 3:SENTIMENTS ANALYSIS USING PYTHON	
3.1 Sentiments Analysis.....	11
3.2 Natural Language Processing (NLP)	11
3.3 System Architecture.....	13
3.4 Python Libraries and Modules	13
3.5 Textblob Library	14
3.6 NLTK (Natural Language ToolKit) Library	15
3.6.1 Sentiments Analyzer.....	16

3.7 Matplotlib Library.....	16
3.8 Pandas Library	17
3.9 CSV Module	19
3.10 OS-Miscellaneous Operating System Interface Module.....	19
3.11 Sys Module	19
3.12 Tweepy Module	19

CHAPTER 4:CASE STUDIES ON TWITTER FOR SENTIMENT ANALYSIS OF POLITICAL ELECTIONS

4.1 Data Gathering	20
4.1.1 Pakistan Tweets Datasets.....	20
4.1.2 UK and Brexit Tweets Datasets.....	21
4.2 Data Collection from Twitter.....	21
4.3 Code processing	23
4.4 Case Study of UK	23
4.4.1 Visualization of Six Days Analysis UK Brexit Twitter Datasets	28
4.4.2 Evaluation Analysis	32
4.5 Case Study of Pakistan.....	34
4.5.1 Pakistan Case Study Analysis.....	36

CHAPTER 5 CHALLENGES AND CONCLUSION 38

REFERENCES..... 39

APPENDIX..... 40

Sentiments Analysis Code

Data Streaming Code

Report.....

LIST OF TABLES

Table 4.1: Pakistan collected tweets.....	20
Table 4.2: UK collected tweets	21
Table 4.3: Sentiments analysis results.....	27
Table 4.4: Sentiments analysis evaluation results of six days.....	33
Table 4.5: Sentiments analysis results of Pakistan case.....	37

LIST OF FIGURES

Figure 2.1: Data Science Interaction.....	5
Figure 2.2: Social media survey report.....	6
Figure 2.3: Twitter Statistical Survey.....	7
Figure 2.4: Python Fundamentals.....	8
Figure 2.5: Python Hierarchy	9
Figure 3.1: NLP Architecture.....	12
Figure 3.2: System Architecture working model.....	13
Figure 3.3: Interface for Input keywords.....	14
Figure 3.4: Polarity results of input keyword.....	15
Figure 3.5: Pie Chart of Keyword Love.....	17
Figure 3.6: Bar Chart of Keyword Love with horizontal and vertical series	18
Figure 4.1: Data streaming from twitter on Python idle.....	22
Figure 4.2: Collected data CSV file.....	22
Figure 4.3: Cleaned tweets data.....	22
Figure 4.4: Sentiment analysis for “Brexit” search term.....	24
Figure 4.5: Sentiment analysis for “EU” search term.....	25
Figure 4.6: Sentiment analysis for “Theresa” search term.....	25
Figure 4.7: Sentiment analysis for “Jeremy” search term.....	26
Figure 4.8: Six day of sentiments analysis evolution of keyword EU.....	29
Figure 4.9: Six day of sentiments analysis evolution of keyword Brexit.....	30
Figure 4.10: Six day of sentiments analysis evolution of keyword Theresa.....	31
Figure 4.11: Six day of sentiments analysis evolution of keyword Jeremy.....	32
Figure 4.12: Sentiments analysis evolution of keyword Imran.....	34
Figure 4.13: Sentiments analysis evolution of keyword Nawaz.....	35
Figure 4.14: Sentiments analysis evolution of keyword Bilawal.....	36

LIST OF ABBREVIATIONS

API:	Application programing interface
BREXIT:	British exit
PNN:	Positive Negative Neutral
NLP:	Natural language Processing
NLTK:	Natural language toolkit
ML:	Machine learning
CSV:	Comma-separated values
TSV:	Tab-separated values
EU:	Europe Union
UK:	United Kingdom
P:	Positive
N:	Negative
T:	Total
PA:	Positive average
NA:	Negative average
MXP:	Maximum positive
MNP:	Minimum positive
MXN:	Maximum negative
MNN:	Minimum negative

CHAPTER 1

INTRODUCTION

1.1 Background

This is a new era of technology which connects people to each other no matter how far they are. This credit goes to social media. Social media is the platform of sharing and receiving information, data, as well as communication system of people. They share their psychology, thinking, ideas, behaviors and sentiments. It is very powerful weapon of increasing literature and business. People use social media to gain education and power for a better life and health. There are many useful social media platform but twitter is the most reliable platform for sentiment analysis because there are more the 336 million worldwide active users (statista.com), more than 100 million daily active users (twitter-statistics last update 6-24-18) and 500 million posts every day (last update 6-24-18). People show their opinions and they are participated on different topics through the twitter posts (tweets) which is useful knowledge base for sentiment analysis. The Twitter data can be received from Twitter in a secure and easy way. We can receive the bulk amount of data through twitter API (Application Programing Interface).

In this research, we analyze opinions, thoughts and perspectives of people about the general topics, politics and political parties. We use worldwide twitter data for general opinion analysis and political perspective and specific data for analysis of political parties such as Pakistan and UK politics. We created an interface to search for specific keywords in a particular Twitter datasets. Then, according to the matching tweets to the Twitter search, we present the sentiments analysis of those tweets (such as sentiments results are presented as pie chart and bar chat). In particular, we analyze the sentiments of people through twitter data about Pakistan current ruling party in parliament and opposition of parliament. What kind of thoughts people have about the Pakistan politics and their political parties. We compare them and show which one is the strongest party in Pakistan. In this research, we analyze twitter post (tweets) and show the result in three ways; positive, negative and neutral. The analysis results are compared for different political parties and visualized as pie chart and bar chart graphs. Result show the total number of tweets, number of positive, number of negative and number of neutral tweet posts for each searched keyword in the political domain. In addition, we also the results of the sentiment

analysis as bar charts; the percentage of tweets positive in green, negative in red and neutral in blue color. In this way, the analyzed tweets show the fairness of the elections based on the post-tweets data. We apply the sentiment analysis of post-tweets data on Pakistan and UK election tweets data. We focused on British exit (Brexit) from the Europe Union and see reaction and support of people. Now a days Brexit is the hard issue and people want to see which policy England will apply for separation. We know that the parliament already reject the bill of Prime minister of UK and showed disagreement. That the reason we choose this topic for research and noted the reaction of people. The form of research is very reliable and authentic because the back hand coding is very strong and its shows appropriate result?Because we use the open source python code with different module in single class (single program). The module with name of “Textblob” is used for sentiments analysis for calculating polarity of tweets.The second module is “Matplotlib”, which is stilted for calculating the percentage and drawing the pie chart graphin three different colors. “Pandas” library providing the series on bar chart graph. These libraries work together first show the polarity of each tweet and then calculates all tweets PNN (positive, negative and neutral) and then calculates the percentage of each part on the pie chart. Bar charts show the total number of tweets and the number of positive, negative and neutral tweets. In this research, we only analyzed three political party of Pakistan and two political party of UK that has the most seats in the last election.

1.2 Aims and Objects

This thesis is based on analyzing the concept, fairness and stand position of the election based on twitter data.

- One of the basic purpose to apply the python code on tweets for analyzing sentiments of user.
- Second show the participation of local people and interest of local people in the politics.
- What kind of thoughts people carrying for the political parties leaders.
- To investigate the twitter data and collecting beneficial information about political parties. In particular, in certain internals, we collected Twitter data about UK (Brexit) and Pakistan politics during our study.

- To analyze if the prediction process is reliable or not and also prove the fairness and fakeness of the last election from the current tweets data based on the sentiment of tweets.
- To create a search interface for searching keywords in twitter dataset and analyze those matching tweets only for sentiment analysis; showing the polarity of the tweet posts and draw the pie chart and bar chart graphs.
- The pie chart draws the percentage of searching keywords and bar chart show the number of tweets.
- Pie charts and bar charts are divided into different colors green for positive, red for negative, blue for neutral and brown for total number of tweets. On the bar chart, the y-axis shows the total number of tweets and x-axis shows positive, negative and neutral tweets.

1.3 Motivation

Nowadays one of the most interesting topics is analysis of people sentiments which is an attractive study in order to draw the future planning and creation of new ideas. Generally, existing works focus on sentiments about the religion and politics. This thesis refers to the political ideas and way of political campaigns on the social media. Micro-blogging is the easy way of accessing writing of sentiments through the twitter posts (tweets) because this is the era of technology and smart phones. The most interesting in sentiments analysis we can predict everything like political views, interesting brand of electronics, sports, boutiques, hotel and resorts, stock exchange, movies, beautiful countries nature wise and technology wise, biggest events and many more.

CHAPTER 2

LITERATURE REVIEW

In this chapter, the following topics are discussed; the data science, social media, twitter, Python and related work on twitter sentiments analysis

2.1 Data Science

In this thesis, we work on the analysis of tweet data. And it is the part of data sciences (Hayashi, Chikio, 1998). Data sciences now a days is one of the fastest growing field in the world. The area studies how to extract the data from different disciplines and interact between each other like mathematics (statistics and algorithms), software engineering and data communication. Data science is divided into different parts such as data collection, knowledge extraction from data, data preparation (cleaning the data and transformation of the data), exploration of the data (what can be done with the gathered data and how to use it), modeling the extracted knowledge with effective tools (we used python), visualization and communication which can be one of the most trickiest part of the data since it is challenging in the thesis how to visualize and how to convey the data for other people ? And finally testing of that data through the tool. Data science also called interdisciplinary field that used scientist in methods, process, algorithms and systems said by (VasantDhar, 2013). (Fig 2.1) The process of extracting meaningful information from the big raw data. Data science is the fourth “paradigm” of science that “everything of science is changing because of the impact of information technology” said in his article the name “fourth paradigm of science” by (Stewart Tansly, el, 2009). There are three different categories for data, first data analyst between the data communication and statistics, second data engineering between software engineer and mathematics and then the data scientist which analyzing every field. (Davenport Thomas H, el, Oct 2012) he said that the data scientist the attractive job of twenty first century and statistics the attractive topic around the world.

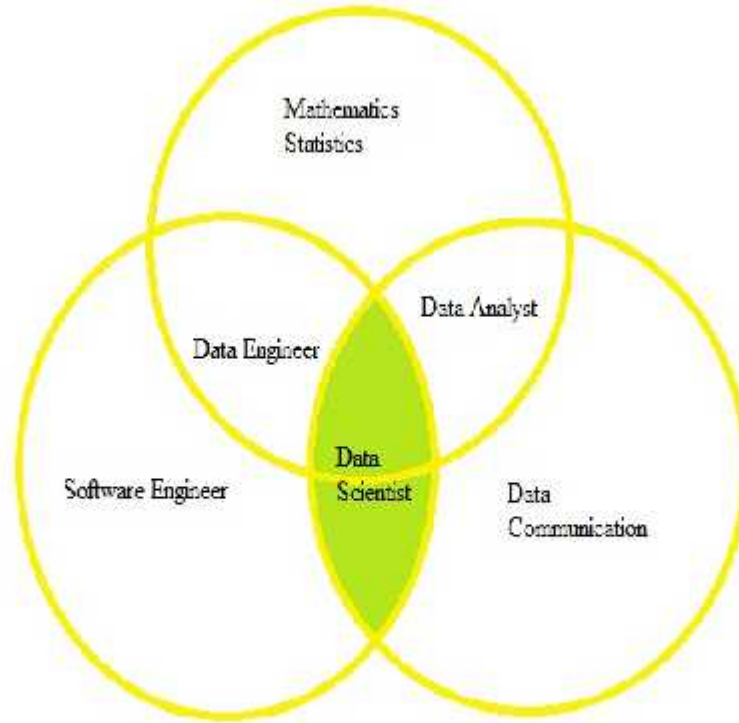


Figure 2.1: Data Science Interaction

2.2 Social Media

In this thesis the social media play the main role. The social media is digital world where people meets together without their presence. But Wikipedia write something like this “social media (noun) is website and application that enable users to create and share the content or participate in a social network” this is formal definition which everyone knows. The word social comes from society and society is living place where the human living the proper way with rules and regulation, make community interaction each other (buildings, roads and meeting clubs) which sometimes harmful for nature. In other word the place where human living and breeding together. Media is the band of communication (acquired and spread knowledge) between the people. Social media is digital platform where people communicate (sharing information or data) with each other to sharing their ideas for the benefit of new generation. There are many social networks like Twitter, Facebook, Instagram, Snapchat and etc. (Obar, el, 2015), where social media is referred as web 2.0 based interaction application. The social media influenced by 1840’s introduction of telegraph in USA, which connecting the country (the Daily Dot, 2016).

With the time apparently increase the number social media users. (Global Social Media, 2018) in the new survey shows, 3.1 billion active social media users (see Fig 2.2).

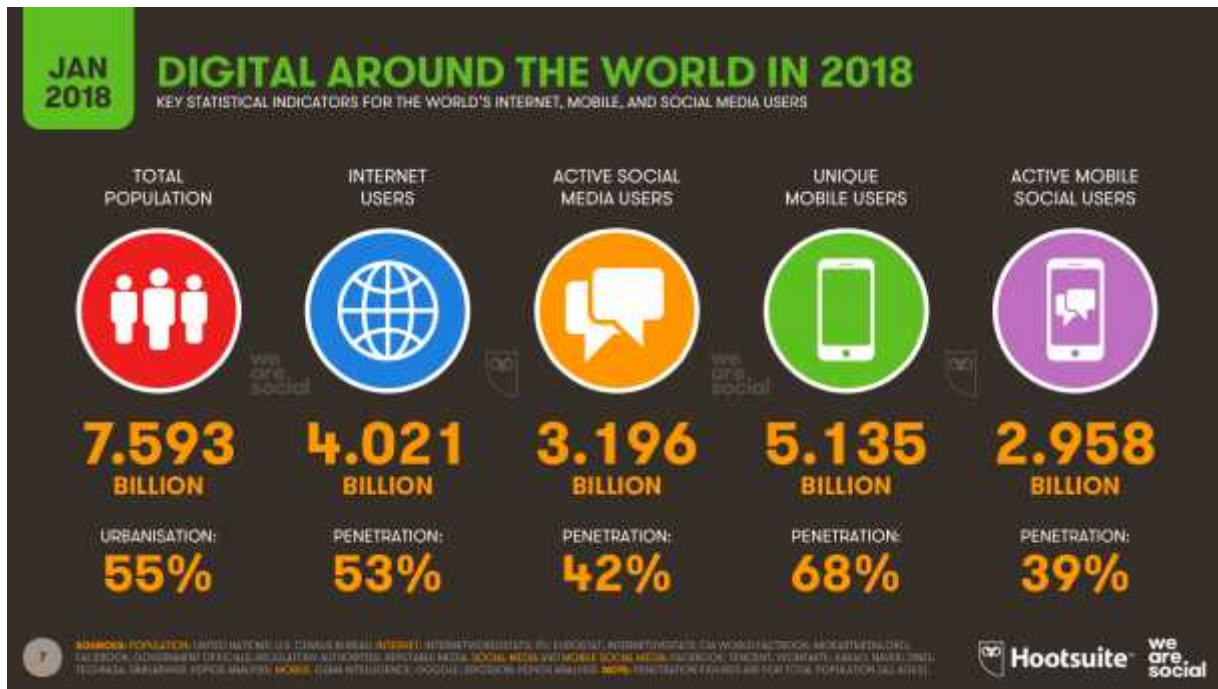


Figure 2.2: Social media survey report

2.3 Twitter

Twitter is one of the biggest social media networks in the world. Twitter is the treasure trove of sentiments people around the world, since people update thousands of actions, opinions, on every topic on every second of the day. It is called one of the biggest psychological database which always being updated and we can analyze the millions of data through the machine learning. Twitter stands on good position in social media networks. Twitter was created in March 2006 founded by Jack Dorsey, Noah Glass, Biz Stone, and Evan Williams (Way back Machine, 2012). Twitter has 336 million active users and more the 100 million daily active users which posts every day more than 500 million posts which contains maximum 280 characters (Statista, 2018). Twitter has opened the most powerful API for developers which recognized as top 10 API of the world. Twitter has two type of accounts one for normal users and other one is developer accounts (using API).

The normal users share and read the information (tweets) but the developer accounts have access to Twitter data through the API (Application program interface). In developer accounts data can

be collected through keys which is provided by Twitter. There are four types of keys, such as consumer key, consumer secret key, token key and token secret key. These keys are unique and different which are used in different programming language to collect tweet data. Twitter is also a big hub for business and advertising. (Gilbertson, et al, 2011) said Twitter uses authentication for account security through the “SMS” service. Twitter is also an open source platform (twitter, 2013).

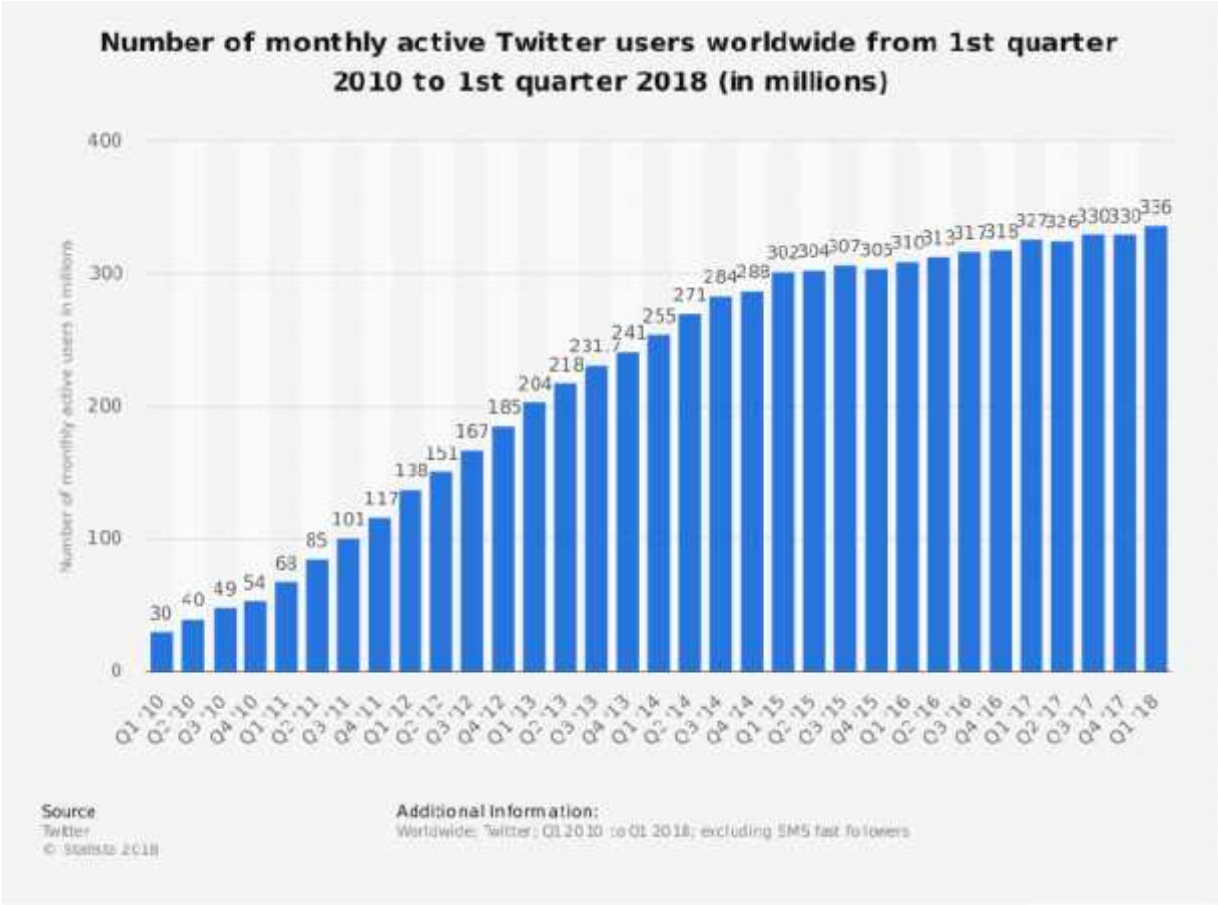


Figure 2.3: Twitter Statistical Survey

2.4 Python

Python is one of the most fast growing programming language in terms of number of developers. Developers mostly use python because it is one of the platform for easiest and fast coding and compilation. Python has huge number of libraries (scientific computing and data sciences) and many big companies use python, such as Google, Yahoo, YouTube, Dropbox and NASA. Python also supports machine learning, GUI, software developing and web developing, which are some

of the reasons, it is used in this thesis. Python is general purpose, interpreter, object oriented and high level language. Python is also multi-paradigms programming language like functional, imperative, object oriented and reflective language. Python consists of different syntax and semantics such as Indentation, Statement and control flow, Expression, Methods, Typing and mathematics.

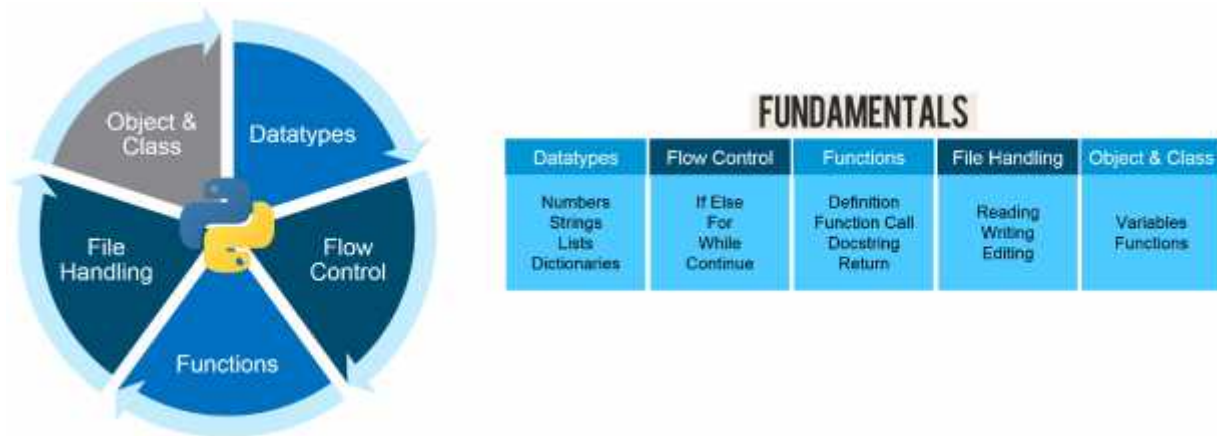


Figure 2.4: Python Fundamentals

There are three different typing ducks, dynamic and gradual. Duck typing is an object can be used for a particular purpose. With normal typing, suitability is determined by an object type (python 3.7.1, 2018). Dynamic is the set of rule properties that are called types to the various constructs of computer program such variable, function, expression or module. Programming language can effectively detect program errors at compile-time. Moreover, some recent studies have indicated that the use of types can lead to significant enhancement of program performance at run-time (Xi, et, 1998). Siek and Jeremy said that the gradual typing is a type system in which some variables and expression may be given type and the correctness. Gradual typing allows software developers to choose either type paradigm as appropriate, from within a single language.

Python's built-in type hierarchy

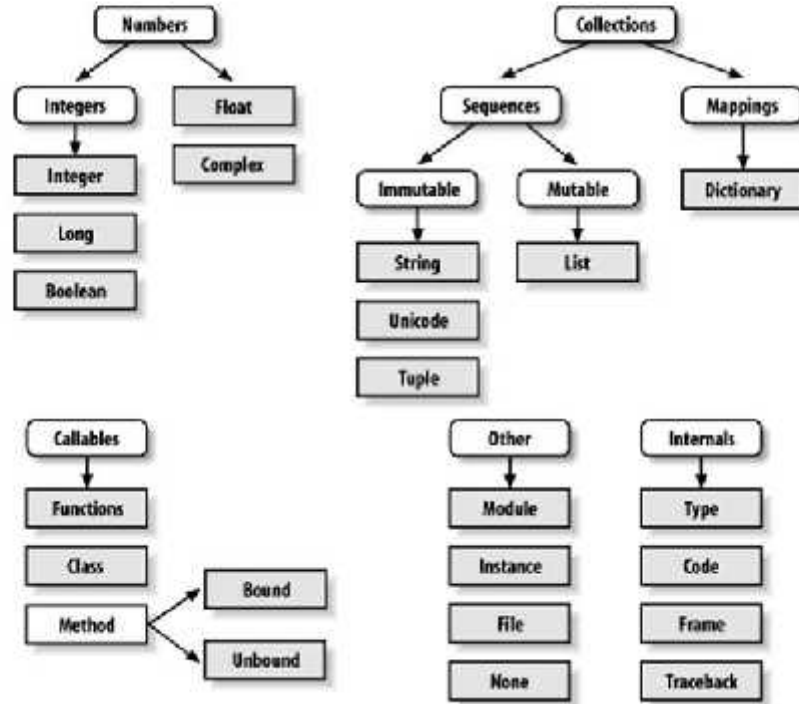


Figure 2.1: Python Hierarchy

Python is a general-purpose high level programming language that is widely used in data science and for producing deep learning algorithms. Python and its libraries like Numpy, Scipy, Pandas, Matplotlib; frameworks like Theano, TensorFlow, Keras for deep learning.

2.5 Sentiment analysis research

There are many research paper and articles about the social media and election prediction through the microblogging sites and twitter but no one can give an easy results and visualization to understand for normal people. In this study we can see the positivity and negativity of the people about their political leaders and see opinions about them. These kind of results help us to predict the election results and popularity of the politicians. There are many papers about election prediction such as (KokilJaidka, Saifuddin Ahmed, et al, 2018) the election prediction of three different countries India, Pakistan and Malaysia. The accuracy of results is awesome. They only shows volumetric performance, Supervised and unsupervised model, And Show the

result on histogram graph chart and expression but not gave an open result that an average people can understand. But in our research shows the clear number of results and clear number of tweets. Some research paper compares two or more than two parties like USA (Alexandre Bovet, et, 2016) Trump versus Clinton, they have large scale of twitter data 0.73 million and gave good results and prediction but results are reversed, Clinton being more popular than Trump. They cannot show number of tweets for each candidate and also gave line graph which do not show number of tweets. There is another paper with low fragmentation and there is no clear approaches, even not given separation or comparison of candidates in histogram chart of United States (Livne, et, 2011). United Kingdom (Boutet, et, 2012) as same problem like (livne, et, 2011) and even not gave an enough information about the prediction of results. Ireland (Birmingham & Smeaton, 2011) in this research researcher gave very low datasets, unclear approach and low fragmentation on line graph and histogram chart graph. As compare to these paper we gave cleared visualization and cleared number of tweets on pie chart and analyze tweet by tweet that mean the average people can easily read and understand the approach.

CHAPTER 3

SENTIMENTS ANALYSIS USING PYTHON

3.1 Sentiments Analysis

Sentiments analysis is the invented science of psychology and sociology and both are the scientific study of people emotions, relationships, opinions, and behaviors (wiki). Psychologist apply sentiments process through the hypothesis but data scientist apply through the data. In other words, it is the computational process which identifies and categories the opinions, thoughts and ideas through the text data. The sentiments analysis process also refer the NLP (Natural language processing). It is internal action process between human and computer. It also analyzes the treasure of natural language data. Sentiments analysis are expressed in two different categories: polarity and subjectivity. The polarity measure the text data is positive (>0) or negative (<0) or neutral (0). Classifying a sentence as subjective or objective, known as subjectivity classification (monkeylearn.com). Subjectivity measures from (0.0 to 1.0). Where 0.0 is very objective and 1.0 is very subjective. But In this thesis we calculate only the sentiments polarity from twitter data (tweets data is in CSV format). Polarity showed three different colors positive for green color, negative in red color and neutral in blue color. Polarity calculated through the python code using library of Textblob and python module Natural Language Tool Kit (NLTK) which explained later.

3.2 Natural Language Processing (NLP)

NLP is the subfield of computer science, information engineering and Artificial intelligence. It is a way of interaction between humans and computers. It is a program which processes and analyzes the large amount of natural language data (Wikipedia). This is the way which makes computers closer to humans because computer cannot understand the feeling and emotions. That is why humans build the NLP because computer work faster the humans. That is being said, recent advances in Machine Learning (ML) have enabled computers to do quite a lot of useful things with natural language. Deep Learning has enabled us to write programs to perform things like language translation, semantic understanding and text summarization. All of these things add

real-world value, making it easy for us to understand and perform computations on large blocks of text without the manual effort (George Seif, 2017).

With the help of Machine learning (ML) and python we make this task easy otherwise it is very hard for computer to understand human language. For example “in the last night Messi on fire” human know the Messi is sports man and he played well but computer take thing literally as it and it translate like this “Messi burned with fire”. So ML is the best Option in NLP. There are different steps in process to analyze the data. First preparation of documents in a proper format like text, etc. Second tokenization the data that the computer understand very well. Third negation and detection it means target the keyword in data if find return “True” for affirmed otherwise “False” for negated. If value come “True” then dependency parser analyzes the grammatically structure of sentence. Co-reference parser which analyze the expression it is the main object in NLP. Finally downstream the result.

Typical Components of an NLP Application

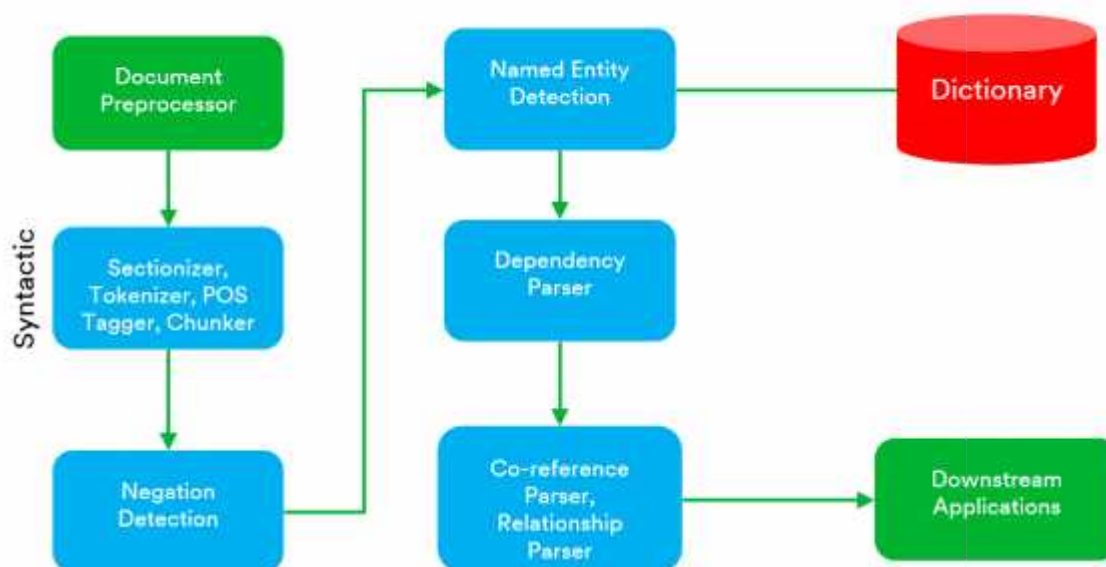


Figure 3.2: NLP Architecture

3.3 System Architecture

When we go for sentiment analysis there are many option and tools. The most popular tools are MATLAB, Python, and Java and C # and due to huge no of libraries available in python and easiest in code so mostly researcher used python because it is sensible and suitable choice. The sentiments analysis algorithm consist of 4 modules. The procedure in each model starts with importing data with pandas, since the powerfulness of pandas for processes and data preprocessing. Then used NLTK and Textblob for analyzing the text of CSV file and calculate the polarity of each text separately and output is a numeric format (-1 to +1). In this research, first collected the tweets from Twitter with given keyword and then analyze the whole text and gave the result, then Matplotlib plotting the result on the pie chat and bar chat with different colors and different formats positive, negative and neutral (greater than zero, less than zero and equal to zero). This program only those text analyze when the required keyword is founded.

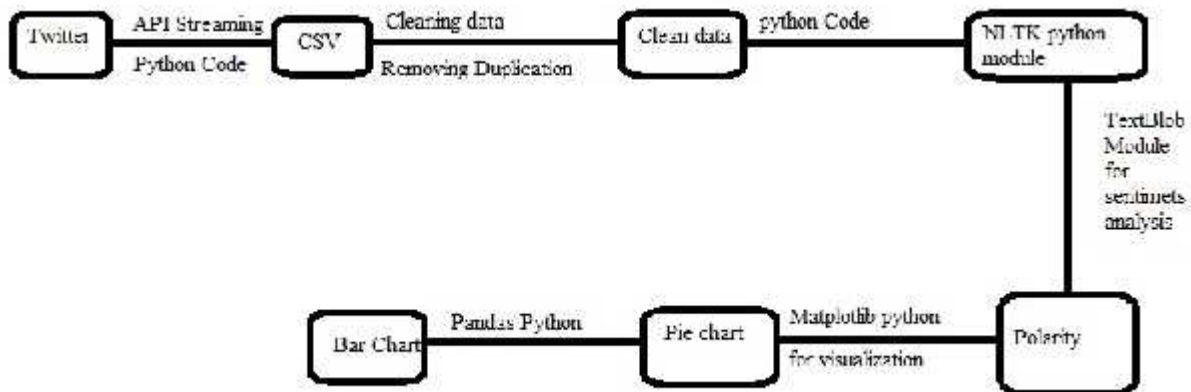


Figure 3.3: System Architecture working model

3.4 Python Libraries and Modules

Libraries which make python easy and fast. Python is the choice of developers only for the libraries of python code. Python modules are files consisting on python code which defines functions, variables and classes (Kuhlman, et, 2012). The created modules can be saved in python library which used on different projects through import files. One module can be used on

different project at the same time, the modules make us work easy because only one time created and many time used. Modules stored in the python library. The modules allow us to logically organize our python code.

3.5 Textblob Library

Textblob is the python library which process the textual data. Textblob provide API to access its methods and easily perform NLP task. The main reason behind the usage of Textblob is it's like a python string easy to use without worrying the syntax. Textblob consist on different function like part of speech, noun phrase, sentiment analysis, tokenization, word inflection and lemmatization, wordlist, spelling correction, translation or language detection and N-gram.

The textblob work and play with all kind of texts. Textblob support all kind of text formats. This also important module of python which use for sentiment analyzing and also classifying the data which part is positive of which part is negative (Steven loria, 2018).

This is the main part of research and textblob is the key library for sentiments analyzing. For example in this section first it takes keyword and collect the keyword tweets from the CSV file and then analyzing the sentiments polarity of each tweet.

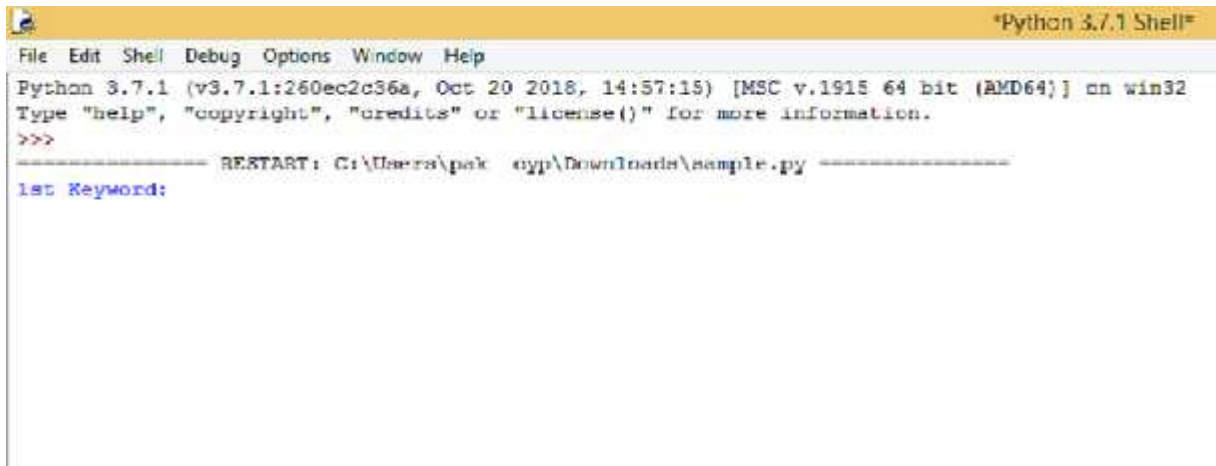


Figure 3.4: Interface for Input keywords

```

Python 3.7.1 Shell
File Edit Shell Debug Options Window Help
-----X
text:"Oz Arzu knows my love for Pakistan And its Army \u2026 https://t.co/5hk0Wdscrt"
Polarity = 0.25
-----X

text:"RT @MudRafid: Pak rulers always fake peace talk& support terrorism hand n glove.Destroy relation with neighbour"
Polarity = -0.1
-----X

text:"RT @MudRafid: Official! You and your love for Khan Khan and your self proclamation of being her Lawyer is enough for everyone to see your inte!
Polarity = 0.25
-----X

text:"RT @sarpit: An expression of love and friendship by Turkish Restaurant owner who served us Bread (Naan) writing Pakistan on it. \u2026 https://t.co/5hk0Wdscrt"
Polarity = 0.5
-----X

text:"If I have only one choice then he is : Aayan Ali's Lover boy & Pakistan's ascend biggest money laundering genius\u2026 https://t.co/5hk0Wdscrt"
Polarity = 0.0
-----X

```

Figure 3.5: Polarity results of input keyword

3.6 NLTK (Natural Language ToolKit) Library

Natural language toolkit it is also called NLTK. It is a suitcase of libraries such as symbolic and statistical natural language process which support Python English written programs. This toolkit have different classifications like sentiments, metrics, parse, tags, tokenization, chat, chunk, classify, translate, twitter, interface, draw, cluster and etc. NLTK includes graphical demonstrations and sample data. It is accompanied by a book that explains the underlying concepts behind the language processing tasks supported by the toolkit, plus a cookbook (Bird, Edward, et, 2009). NLTK used the Python platform for building the programs of natural language text (human language) for using statistical natural language processing. NLTK is open source library for python which used on any platform, such as windows, Mac, Linux and many other platforms. In our thesis, we used Textblob library for sentiment analysis which import the NLTK module and sentiment analyzer is the subclass of NLTK module.

3.6.1 Sentiments Analyzer

A sentiment analyzer is a tool to implement and facilitate sentiment analysis task Using NLTK features and classifiers, especially for teaching and demonstrative purposes. A sentiment analysis tool based on machine learning approaches.

3.7 Matplotlib Library

Matplotlib is the good visualization library and one of the most popular than other libraries. Major libraries based on Matplotlib like Seaborn and etc. it produce different types of plots and charts Matplotlib is the library of python used for 2D graphics to visualize the data and its support all kind of graphic user interface. There are different function in the library which support different bunch of commands like

-) Matplotlib.pyplot: pyplot gave full control of line style font properties, axes properties and etc. Pyplot created a blank chart then add one element at a time like title, axes, curve, bars, annotation and etc. Pyplot is collection of commands function which makes Matplotlib work like Matlab. The function of “Pyplot” make changes in the images and plotting the figure on different area. It works on different axes (negative or positive) with help “Pyplot” we plot the number on x-axes and y-axes with help of single line array plot generate automatically x-axes numbers. In this function plot the different formats of the chart.
-) Draw the line plot with text label
-) Draw the multiple sub-plot on the different charts
-) Display the image through the library image function this function also used in CT scan.
-) Display the two dimension image data with help of “pcolormesh”
-) Generate histogram with the usage of “hist” function
-) Created arbitrary path using “.path” module and also generate 3D graph surface, bar chart, wireframe and etc.
-) Generated bar char with customized values
-) Draw pie chart with different colors and calculation of percentage
-) Created tables of values which is given in data

-) Make scatter plots with size and colors
-) Filled the curves and shapes in the chart
-) Customized the time series with the availability of vacancy. It means handling the dates.
-) Polar plot, notation, mathematical expression, text objects sketch style and etc.

The matplotlib use for the sentiments visualization and it shows the total number of positive, negative and neutral tweets from the total number of tweets.

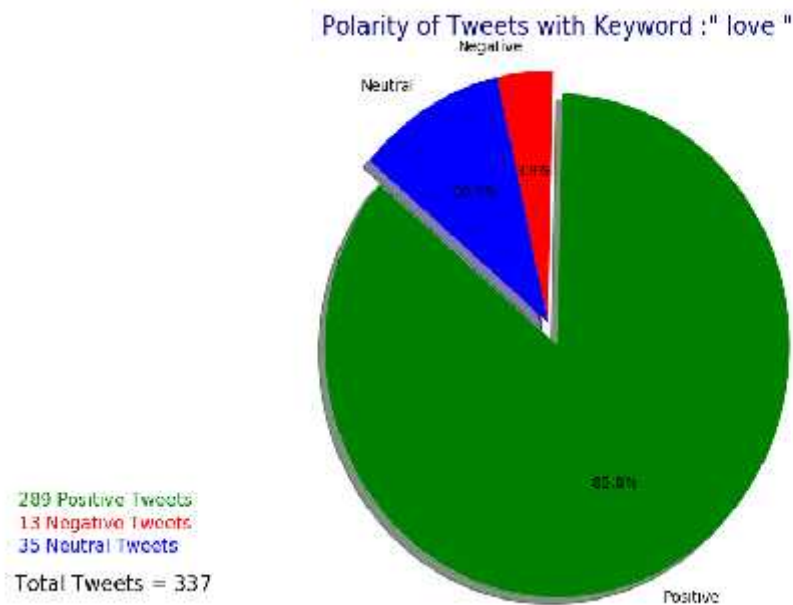


Figure 3.6: Pie Chart of Keyword Love

3.8 Pandas Library

This is although an open source library which provides data structures and data analysis tools. The important note about pandas is its high performance and easy to use especially for manipulating operations in numerical tables and time series data. Though pandas used to store the tweets data in dataframe where it then divided in X and Y dimensions and made it ready for analysing and other preprocessing operations. This library use for bar chart horizontal or vertical visualization and support the CSV files of tweets data. The total number of tweets showed on the y-axis and others are showed on x-axis. The last update of Pandas 3 august 2018

(pandas.pydata.org). Pandas also take file as CSV and TSV or SQL database and it is also created file column or row wise like Excel. Pandas also support to language “R” it means we can use Pandas in language “R”. Pandas created python objects rows and columns called data frame. It is also helpful in loading and saving the data. Pandas able to perform all commands of statistics (mean, mod, max, min, correlation and count) without using formula these feature are built in Pandas. It is also change the format of data files and used for filtering and sorting the data group wise. This library also use for cleaning the data and joining or combining the data in rows or columns.

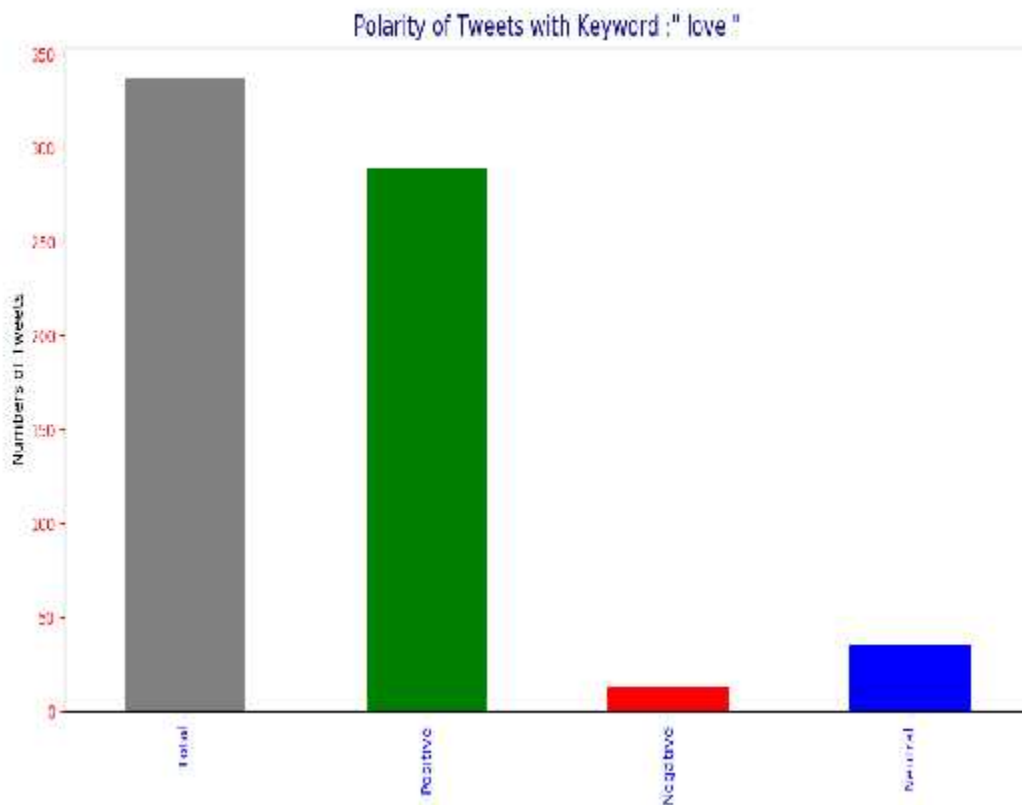


Figure 3.7:Bar Chart of Keyword Love with horizontal and vertical series

3.9 CSV Module

In this thesis used the CSV (comma separate value) format in coding this format most common format for import and export the files. CSV module used for reading and writing file in python code.

3.10 OS-Miscellaneous Operating System Interface Module

This module are portable operating system it used for reading and writing the file. We can set the path through this module and also created temporary file in this module. This module also read the common line from the all file.

3.11 Sys Module

System specific parameter function this module used for manage the size or limit of the file which used in the python code. Extract constant function and methods in python through this module. This module used as interpreter in python and this thing make special Python as compare to Pearl and Java. This module have too many commands like native byte order, tracing, mapping, copyright, clear cache, current frame and many more.

3.12 Tweepy Module

This is the most important module in our thesis work. Without this module we cannot collect the twitter posts (tweets) from the twitter API. This is the open sourced library which are connected with twitter through API. This is also efficient library of python like others. Tweepy support authentication keys provided by twitter. Consumer, consumer secret, token, and token secret keys these keys are unique for every user or API. Through these keys we extract the data from twitter on different topics. Tweepy to connect to twitter streaming API and downloading the data.

CHAPTER 4

CASE STUDIES ON TWITTER FOR SENTIMENT ANALYSIS OF POLITICAL ELECTIONS

This chapter is designed to describe the methods and tools used to forecast the currency exchange rates in the Forex market. Firstly the used tools are presented with clarifications of how they being used within the research. Then the data cleaning, preprocessing and algorithms are discussed with a brief conclusion and summary of the mentioned issues at the end.

4.1 Data Gathering

The datasets download from the twitter through twitter API. Four different datasets about the political view of England and Pakistan. Two datasets for current and two datasets for earlier sentiment analysis. These datasets consist on more than two millions tweets. Some datasets are big but some datasets are small in capacity of tweets. During the streaming data the Twitter disconnect the connection of API, it means no more data with searching the same keywords every day and also find more duplication in tweets so after cleaning the tweets data are too smaller in size.

4.1.1 Pakistan Tweets Datasets

These tweets datasets downloaded or streamed from the twitter through API. These thousands of tweets consist on special keywords like the name of political leaders and these tweets belongs to Pakistani peoples about their leaders and it shows the emotions, sentiments and opinion of Pakistani people. There are two kinds of tweets datasets of two different months.

Table 4. 1:Pakistan collected tweets

Date	Number	of
12.2018	29327	tweets
01.2019	1119	tweets
Total	30446	tweets

4.1.2 UK and Brexit Tweets Datasets

In this case study, there are eleven type of different tweets datasets. Two tweets datasets downloaded from (data.world) and we collected nine tweets datasets through from Twitter API. In January 2019 (before/after the Brexit parliament voting in 15.01.2019). These millions of tweets data consist on the opinions of UK and Europe people. These tweets datasets are gave reliable results about the UK parliament.

Table 4. 2:UK Collected tweets

Dates	Number of tweets data
30.05.2017	418328 tweets
31.05.2017	1048576 tweets
13 Jan to 20 Jan 2019	333510 tweets
04 Feb to 06 Feb 2019	23517 tweets
13.01.2019	15937 tweets
14.01.2019	4939 tweets
15.01.2019	1513 tweets
16.01.2019	24289 tweets
17.01.2019	98588 tweets
18.01.2019	15727 tweets
20.01.2019	16945 tweets
Total	1995417 tweets

4.2 Data Collection from Twitter

This is the basic thing of research or just said the core thing of research without data is nothing. There are many ways to collecting the data from twitter but in our suggestion, python is the easiest and simple way to collection the data. Using the python code library Tweepy we access the data through twitter API. API provide the keys for accessing the data of twitter. There are four keys which used for authentication and accessing the twitter account consumer key, consumer secret key, token key, token secret key. There are three different steps to collect the cleaned data from Twitter. First streaming the data from Twitter and saved in CSV file. Second collect the tweets text from one CSV file to save in another CSV file. Third removed duplication from tweets data. These steps are also showed in the (Fig 4.1, Fig 4.2 and Fig 4.3).

4.3 Code processing

After the datasets, are collected next step is code execution or code processing. Import the datasets in code and execute the program. The program based on three steps one input the keywords which you want, two search the input keywords and three results of the giving keywords. These steps are compulsory in program if we don't close the first result then we are not able to process second result so first close the executed then apply for next results or keywords

4.4 Case Study of UK

In our work, we analyze sentiment analysis of two political leaders of Britain one Theresa May and second Jeremy Corbyn. We also analyze changing opinions of people about the Europe Union and Brexit. There are four different kind of datasets, two datasets downloaded from (dataworld.com) and two datasets are collected from the twitter API. In particular, we queried the Tweeter API with a combination of keywords such as Brexit, Theresa May, Jeremy Corby, European Union (EU) and collected daily tweet data in January 2019 and February 2019. It can be seen that on the day of Parliament voting (17th of January), the tweet activities were increased considerably. In Figures 4.4 to 4.7, we demonstrate visual analysis of sentiments about Brexit, European Union (EU) and UK politicians. Each keyword search is applied to four time intervals that we collected tweet data. In particular, we present four pie charts as oppose to tweet data in 30 May 2017, in 31 May 2017, in January 2019 (January tweets are combined together) and in February 2019 (February tweets are combined together).

We observe that people were more positive about Brexit in 2017, whereas in January 2019 and especially after parliament voting in February 2019, their positivity was dropped around 5%. Similarly, even after the Brexit referendum, positivity about EU was high around 38% in 2017. However, before British parliament voting in January 2019, the positivity was dropped around 3%, and after the parliament voting in January 2019, the positivity about EU was also dropped and kept around 30%. When we observe changes in UK politicians, we observe that Theresa May tweet sentiments were dropped considerably. In 2017, people were more positive about Theresa May around 30%. Before and after the parliament voting in January 2019, the positivity about Theresa May dropped to 28% and 23% respectively. The opposition party leader Jeremy Corbyn

also has more positive sentiments in 2017 with around 40%. In January positivity about Jeremy Corby was kept stable with around 40%. But after the parliament voting in February 2019, the positivity about Jeremy Corby was dropped to 29%.

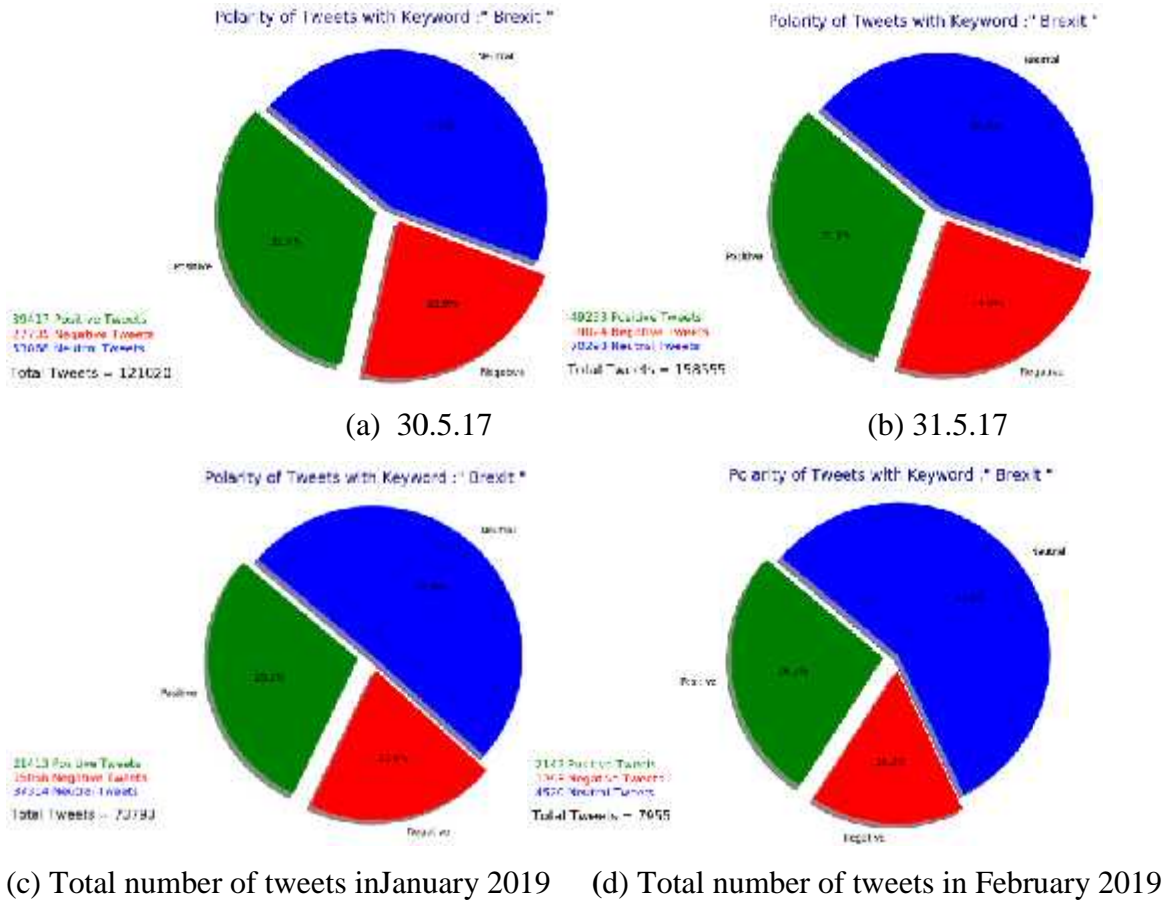
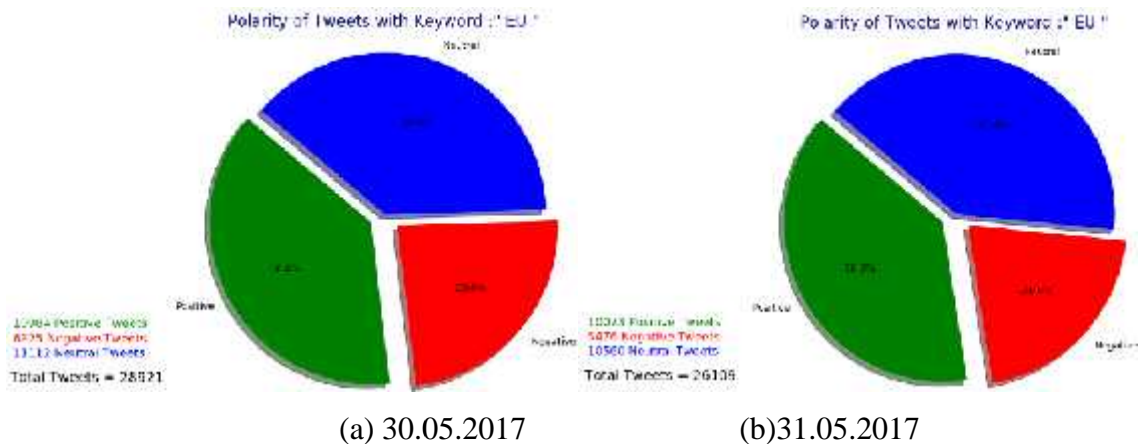
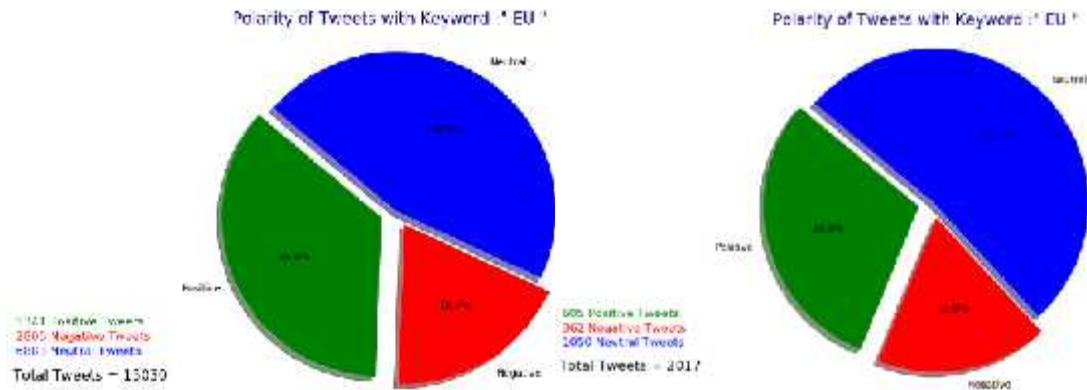
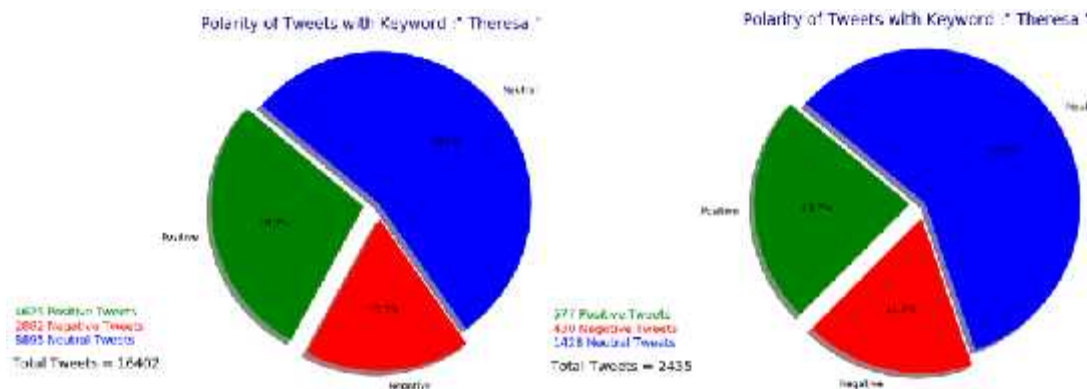
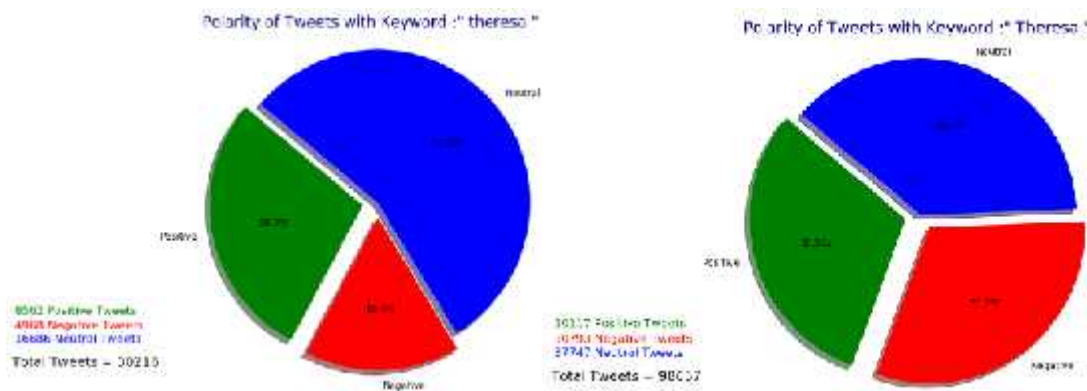


Fig. 4.4. Sentiment analysis for “Brexit” search term





(c) Total number of tweets in January 2019 (d) Total number of tweets in February 2019
Fig. 4.5. Sentiment analysis for “EU” search term



(c) Total number of tweets in January 2019 (d) Total number of tweets in February 2019
Fig. 4.6. Sentiment analysis for “Theresa May” search term

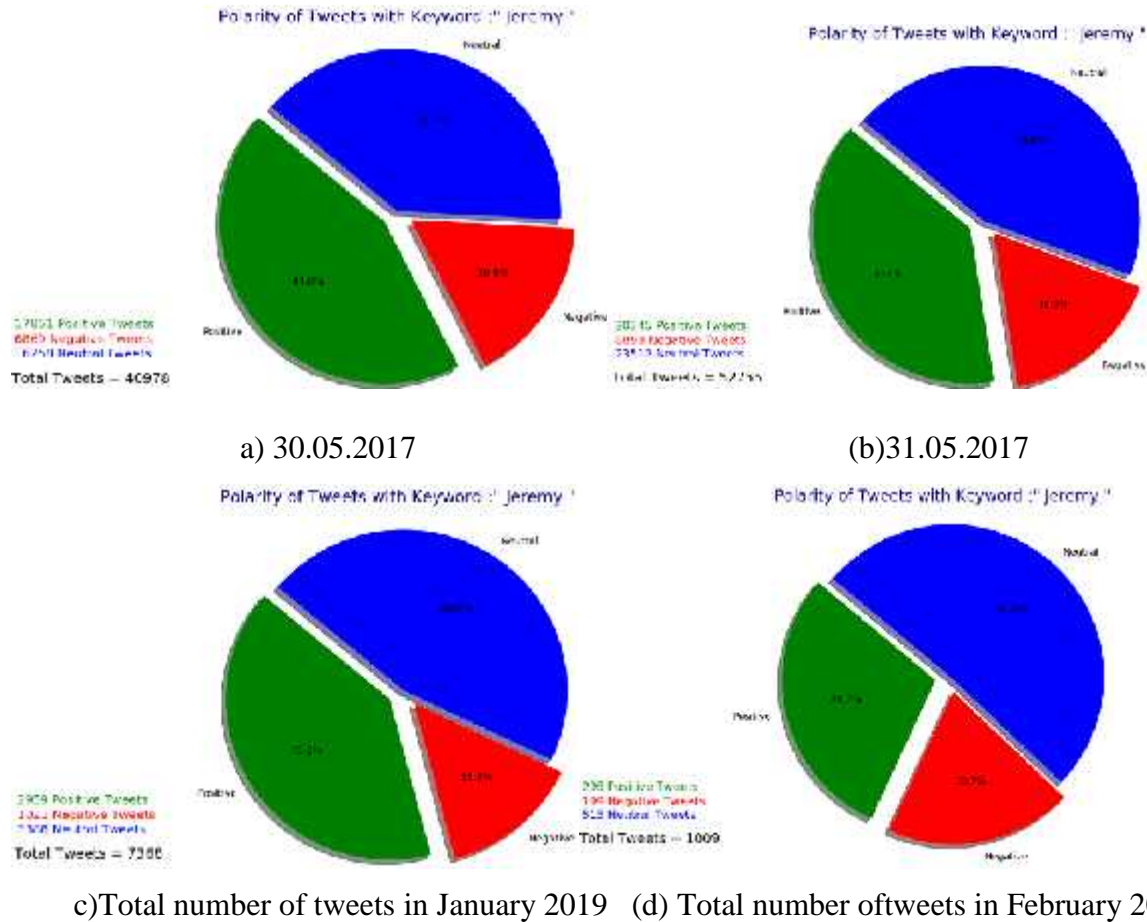


Fig. 4.7. Sentiment analysis for “Jeremy Corby” search term

In Table 4.3, we also analyze sentiment data quantitatively. ‘P’ represents positive, ‘N’ represented negative, ‘NT’ represents neutral, ‘T’ for total, ‘PA’ for positive average, ‘MXP’ for maximum positivity, ‘MNP’ for minimum positivity, ‘NA’ negative average, and ‘MXN’ maximum negativity and ‘MNN’ minimum negativity.

$$P = \frac{m \cdot p_i}{t_i \cdot t} * 100 \quad (1)$$

$$NA = \frac{m \cdot n_i}{t_i \cdot t} * 100 \quad (2)$$

Table 4.3.Quantitative tweet sentiment analysis.

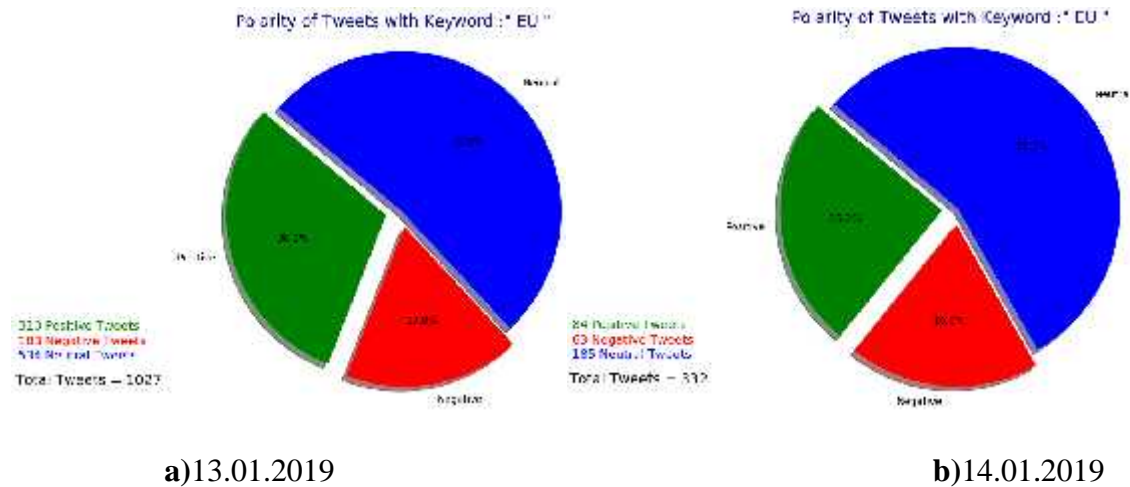
Dates	Brexit	EU (Europe Union)	Theresa May	Jeremy Corbyn
30.5.2017	P=32.6=39417 N=22.9=27735 NT=44.5=53868 T=100=121020	P=38=10984 N=23.6=6825 NT=38.4=11112 T=100=28921	P=28.3=8562 N=16.4=4968 NT=55.2=16686 T=100=30216	P=43.6=17851 N=16.8=6869 NT=39.7=16258 T=100=40978
31.5.2017	P=31.1=49233(1.5) ↓ N=24.6=39024 NT=44.3=70298 T=100=158555	P=38.6=10073(0.6) ↑ N=21=5476 NT=40.4=10560 T=100=26109	P=30.5=30117(2.2) ↑ N=31.2=30793 NT=38.3=37747 T=100=98657	P=38.6=20345(5) ↓ N=16.9=8898 NT=44.6=23512 T=100=52755
01.2019	P=29=21413(3.6) ↓ N=20.4=15066 NT=50.6=37314 T=100=73793	P=35.5=5341(2.5) ↓ N=18.7=2806 NT=45.8=6883 T=100=15030	P=28.2=4625(0.1) ↓ N=17.6=2882 NT=54.2=8895 T=100=16402	P=40.2=2959(3.4) ↓ N=13.9=1021 NT=46=3388 T=100=7268
02.2019	P=26.9=2142(5.7) ↓ N=16.3=1293 NT=56.8=4520 T=100=7955	P=30=605(8) ↓ N=17.9=362 NT=52.1=1050 T=100=2017	P=23.7=577(4.5) ↓ N=17.7=430 NT=58.6=1428 T=100=2435	P=29.2=295(13.8) ↓ N=17.9=199 NT=51=515 T=100=1009
Overall	MXP=32.6 MNP=26.9 PA=31.05 MXN=24.6 MNN=16.3 NA=23	MXP=38.6 MNP=30 PA=37.46 MXN=24.6 MNN=16.3 NA=21.4	MXP=30.5 MNP=23.7 PA=29.7 MXN=31.2 MNN=16.4 NA=26.45	MXP=43.6 MNP=29.2 PA=40.6 MXN=17.9 MNN=13.9 NA=16.65

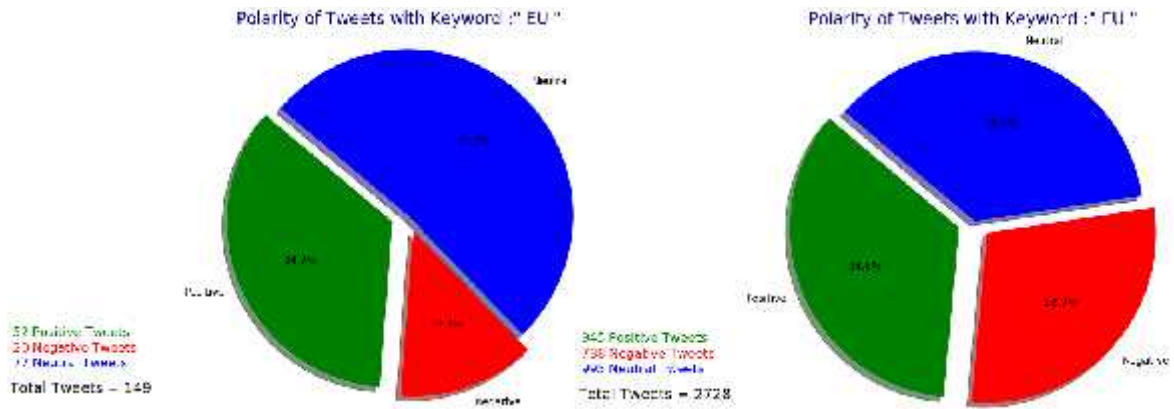
With the help of equations (1) and (2), we can calculate the positive average and negative average of the sentiments analysis results which shown in given Table 4.3. Analysis of the results on these datasets shows that Theresa May received a maximum of positive average 29.7% and maximum negative average 26.45% tweets compared to maximum positive average 40.6% and maximum negative average 16.65% of Jeremy. We observed that positive and negative

average of Jeremy is better than Theresa. In other case EU receives 37.46% maximum positive and 21.4 maximum negative average as compared to Brexit 31% maximum positive and 23% maximum negative average which is not good as compared to EU.

4.4.1 Visualization of Six Days Analysis UK Brexit Twitter Datasets

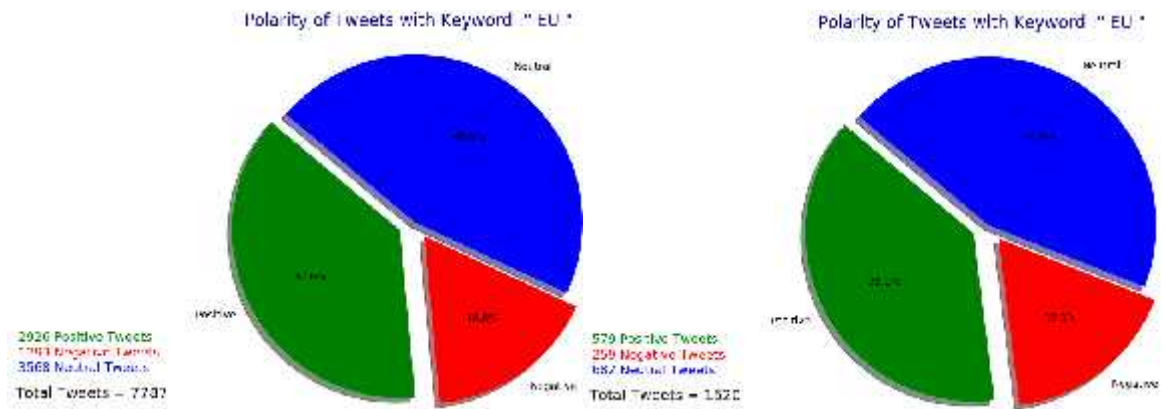
This research shows the assessments of British parliaments voting through the twitter tweets. In this research we analyze tweets of different dates and find different sentiments of the people. In the behalf of result we easily judge the voting results and who the strongest candidate of UK parliament is and also see the sentiments about Brexit from EU. This evaluation assessment of different datasets of different dates which directly collect from the twitter through twitter API. First shows the results of keywords date by date and the shows the differentiation between them. This evaluation shows the maximum, minimum and average percentage of each keyword. The all results are visualized on the pie charts for easy to understand.





c)15.01.2019

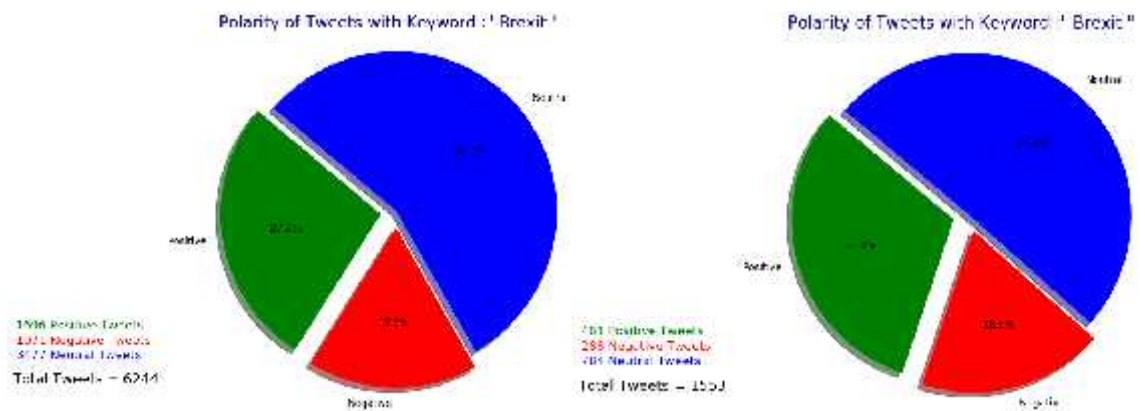
d)16.01.2019



e)17.01.2019

f)18.01.2019

Figure 4.10:Six day of sentiments analysis evolution of keyword EU



a)13.10.2019

b) 14.01.2019

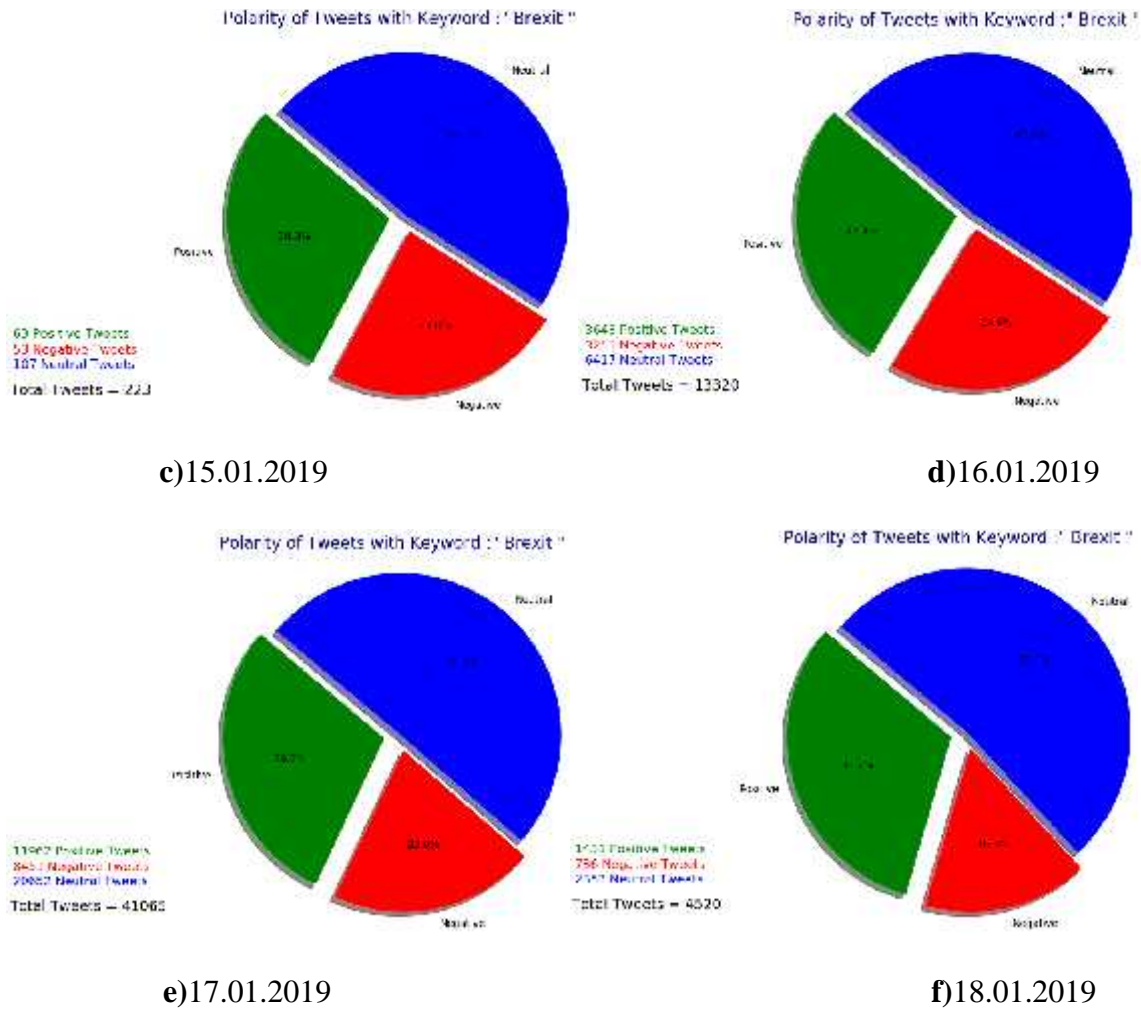
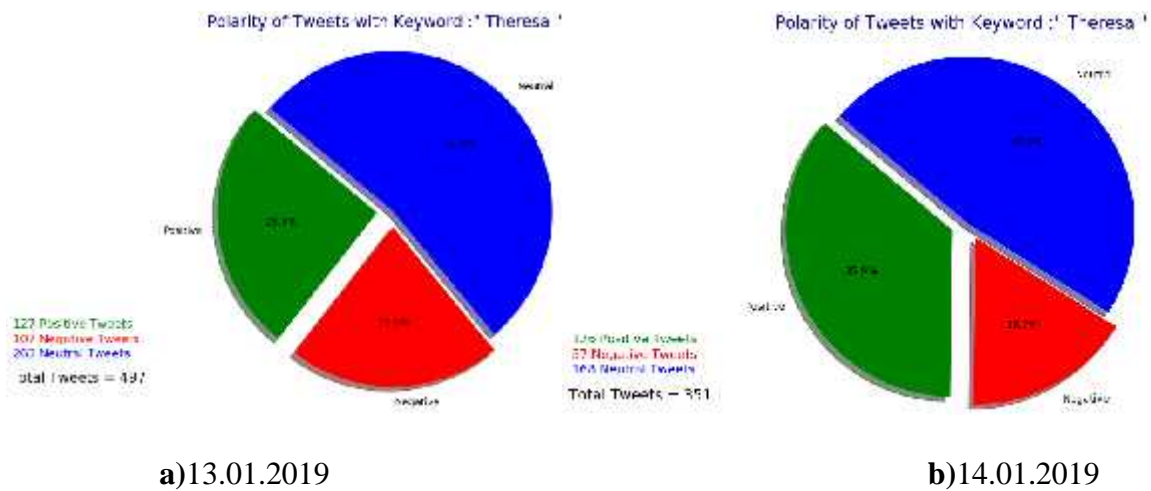
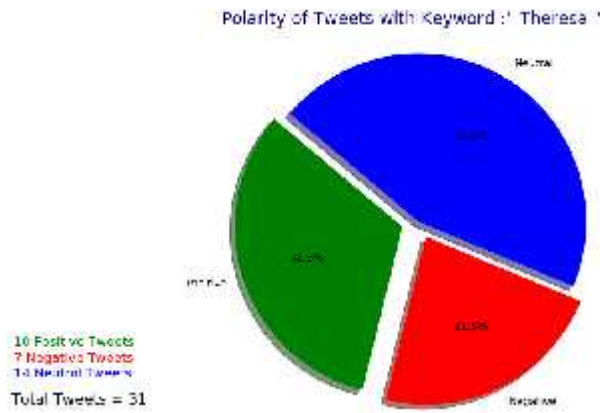
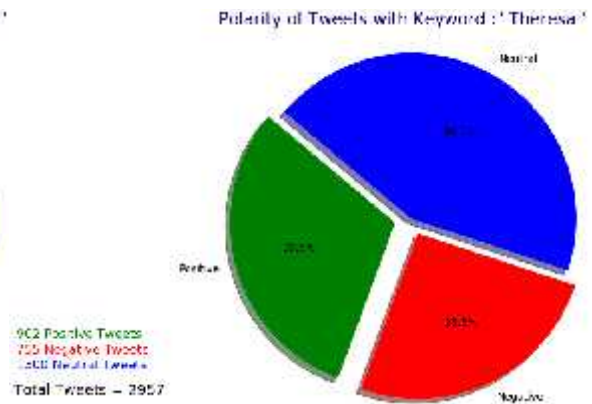


Figure 4.11: Six day of sentiments analysis evolution of keyword Brexit

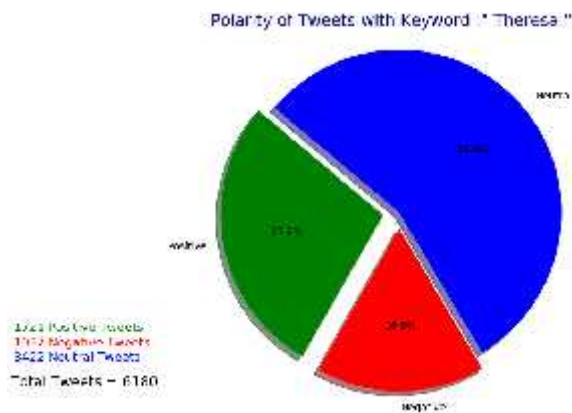




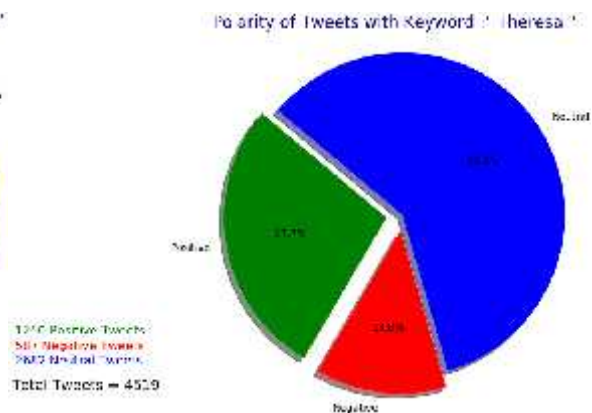
e)15.01.2019



d)16.01.2019

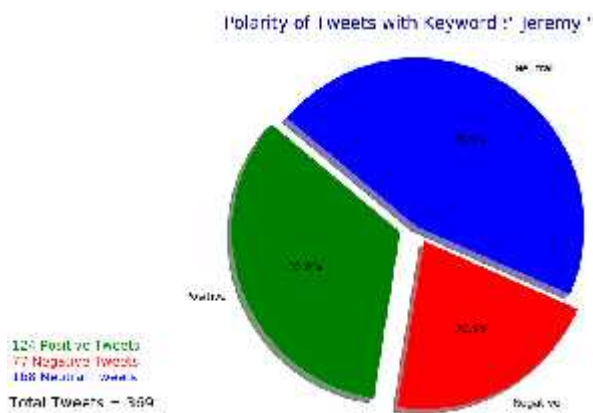


e)17.01.2019

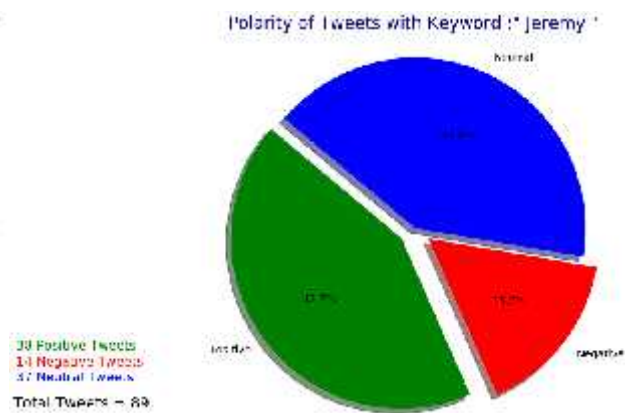


f)18.01.2019

Figure 4.12:Six day of sentiments analysis evolution of keyword Theresa



a)13.01.2019



b)14.01.2019

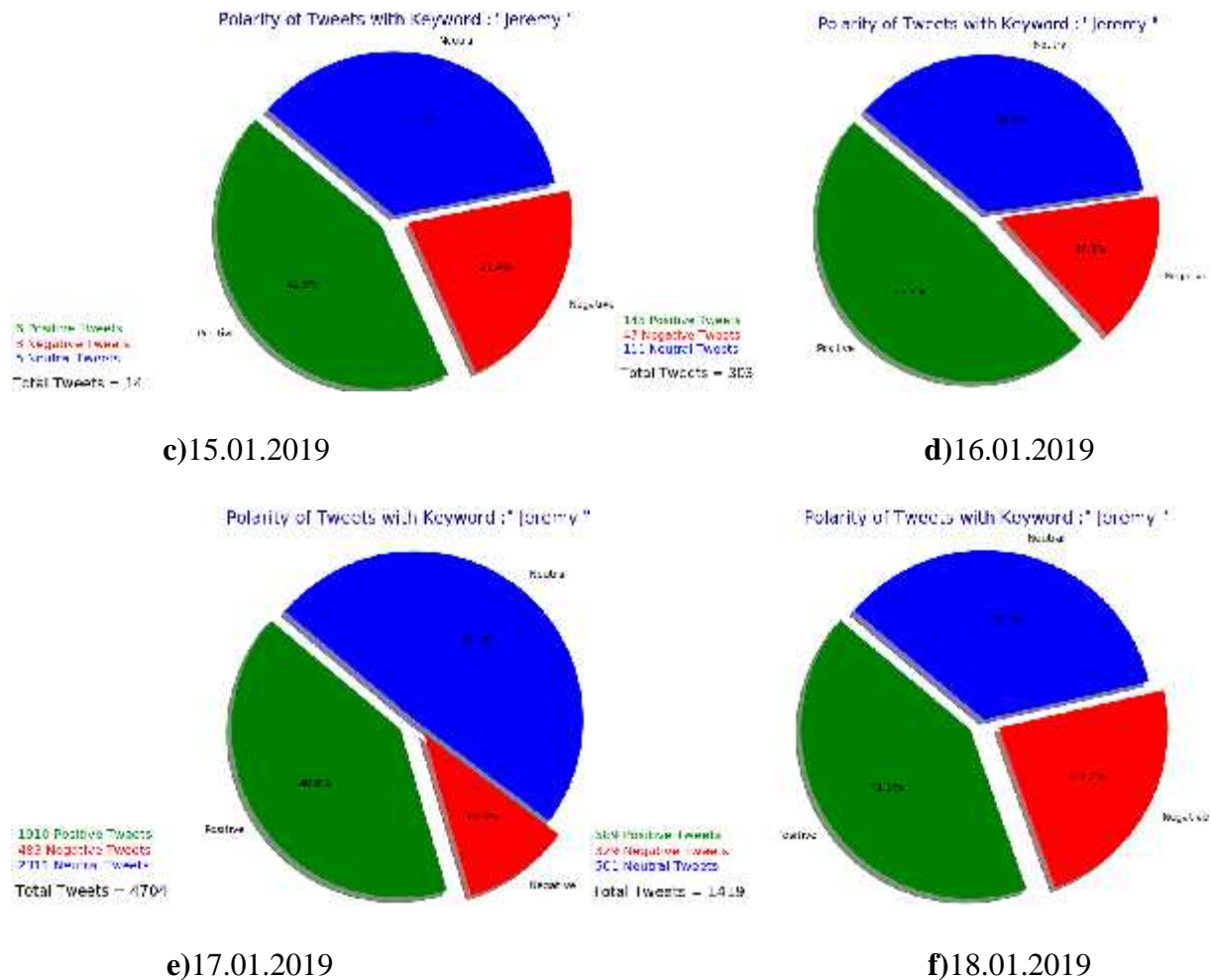


Figure 4.13:Six day of sentiments analysis evolution of keyword Jeremy

4.4.2 Evaluation Analysis

Shows the differentiation between the keyword ‘P’ used for positive, ‘N’ used for negative, ‘NT’ used for neutral, ‘T’ used for total, ‘MXP’ used for maximum positive, ‘MNP’ used for minimum positive, ‘PA’ used for positive average, ‘MXN’ used for maximum negative, ‘MNP’ used for minimum negativity and ‘NA’ used for negative average in the table. Up and down arrow shows the results up and down, these all arrows compare with first result of the date. Analysis of the results on these datasets shows that Theresa May received positive average of 28.45% and negative of 17.45%, tweets as compared to Jeremy Corbyn received positive average of 40.76% and negative average of 13.8%. From these results we observed that Theresa May losing power in UK politics and Jeremy is strong politician as compared to Theresa. It means in upcoming election the Theresa political party face more difficult situation instead of Jeremy. In

other case EU received positive average of 36.15%, negative average of 19.24% tweets, Brexit received positive average of 28.8% and negative average of 20.7%. From these results we analyze that EU is going more positive outlook as compared to Brexit. In future may be UK face economical problem because UK losing his supports.

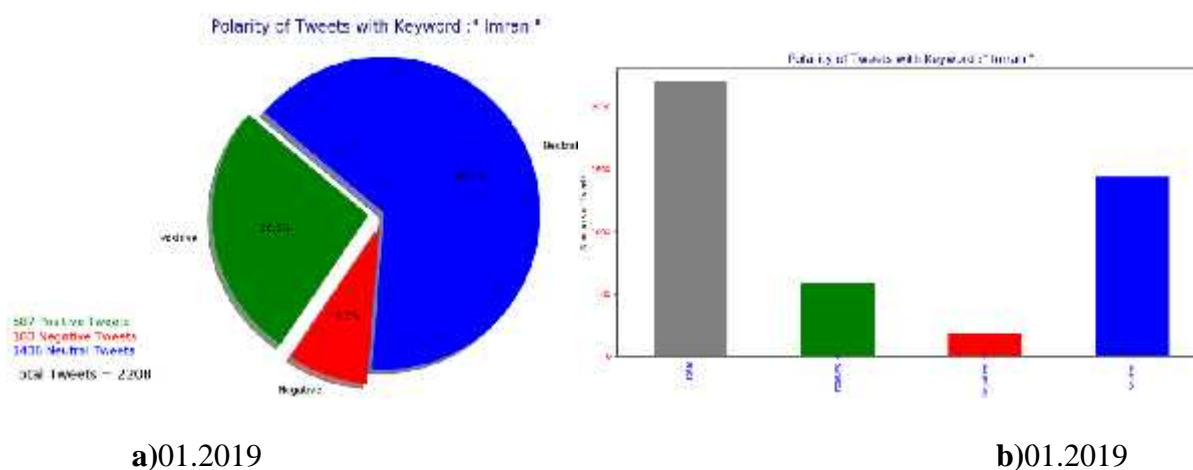
Table 4.3: Sentiments analysis evaluation results of six days

Date	EU	Brexit	Theresa	Jeremy
13.1.19	P=30.2=310 N=17.8=183 NT=52=534 T=100=1027	P=27.2=1696 N=17.2=1071 NT=55.7=3477 T=100=6244	P=25.6=127 N=21.5=107 NT=52.9=263 T=100=497	P=33.6=124 N=20.9=77 NT=45.5=168 T=100=369
14.1.19	P=25.3=84(4.9) ↓ N=19=63 NT=55.7=185 T=100=332	P=31=481(3.8) ↓ N=18.5=288 NT=50.5=784 T=100=1553	↑ P=35.9=126(10.3) ↑ N=16.2=57 NT=47.9=168 T=100=351	P=42.7=38(9.1) ↑ N=15.7=14 NT=41.6=37 T=100=89
15.1.19	P=34.9=52(4.7) ↑ N=13.4=20 NT=51.7=77 T=100=149	P=28.3=63(1.1) ↑ N=23.7=53 NT=48=107 T=100=223	↑ P=32.3=10(6.7) ↑ N=22.6=7 NT=45.2=14 T=100=31	↑ P=42.9=6(9.3) ↑ N=21.4=3 NT=35.7=5 T=100=14
16.1.19	P=34.6=945(4.4) ↑ N=28.9=788 NT=36.5=995 T=100=2728	P=27.4=3648(.2) ↑ N=24.4=3255 NT=48.2=6417 T=100=13320	↑ P=30.5=902(5.9) ↑ N=25.5=755 NT=44=1300 T=100=2957	↑ P=47.9=145(14.3) ↑ N=15.5=47 NT=36.6=111 T=100=303
17.1.19	P=37.6=2926(7.4) ↑ N=16.6=1293 NT=45.8=3568 T=100=7787	P=29.1=11962(1.9) ↑ N=20.6=8451 NT=50.3=20652 T=100=41065	↑ P=27.8=1721(3.2) ↑ N=16.8=1037 NT=55.4=3422 T=100=6180	↑ P=40.6=1910(7) ↑ N=10.3=483 NT=49.1=2311 T=100=4704
18.1.19	P=38.1=579(7.9) ↑ N=17=259 NT=44.9=682 T=100=1520	P=31.7=1431(4.5) ↑ N=16.3=736 NT=52.1=2353 T=100=4520	↑ P=27.7=1250(3.1) ↑ N=13=587 NT=59.3=2682 T=100=4519	↑ P=41.5=589(7.9) ↑ N=23.2=329 NT=35.3=501 T=100=1419

	MXP=38.1	MXP=31.7	MXP=35.9	MXP=42.9
Overall	MNP=25.3	MNP=27.2	MNP=25.6	MNP=33.6
	PA=36.15	PA=28.8	PA=28.45	PA=40.76
	MXN=28.9	MXN=24.4	MXN=25.5	MXN=23.2
	MNN=13.4	MNN=16.3	MNN=13	MNN=10.3
	NA=19.24	NA=20.7	NA=17.54	NA=13.8

4.5 Case Study of Pakistan

This case study little bit different from the UK case study behind that two reason first we do not streaming (downloading) the tweets as much as required the reasons are too much duplication and less usage of twitter in Pakistan as compare to Facebook. Second the non-English tweets because much amount of people used Urdu in English alphabet. So that why the datasets consist on low amount of tweets data. After all the program find the good results. In this study there are three political leaders of Pakistan one the current prime minister of Pakistan, two the second seats taker of general election of Pakistan and three the Pakistan most famous and one off the oldest party and third seats taker party. The all sentiments of people shared in table (4.5) with the percentage of positive, negative and neutral and number of tweets positive negative and neutral.



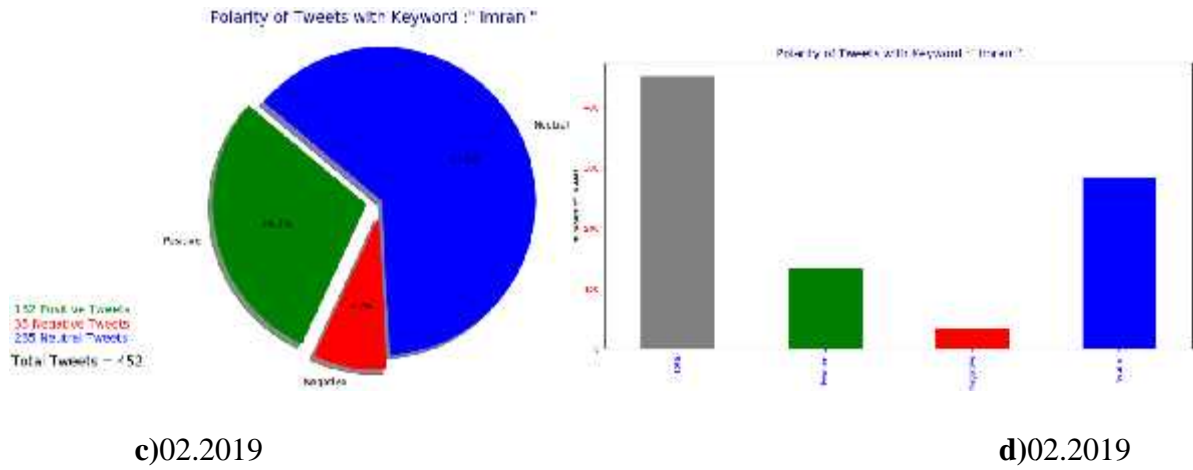


Figure 4.14:Sentiments analysis evolution of keyword Imran

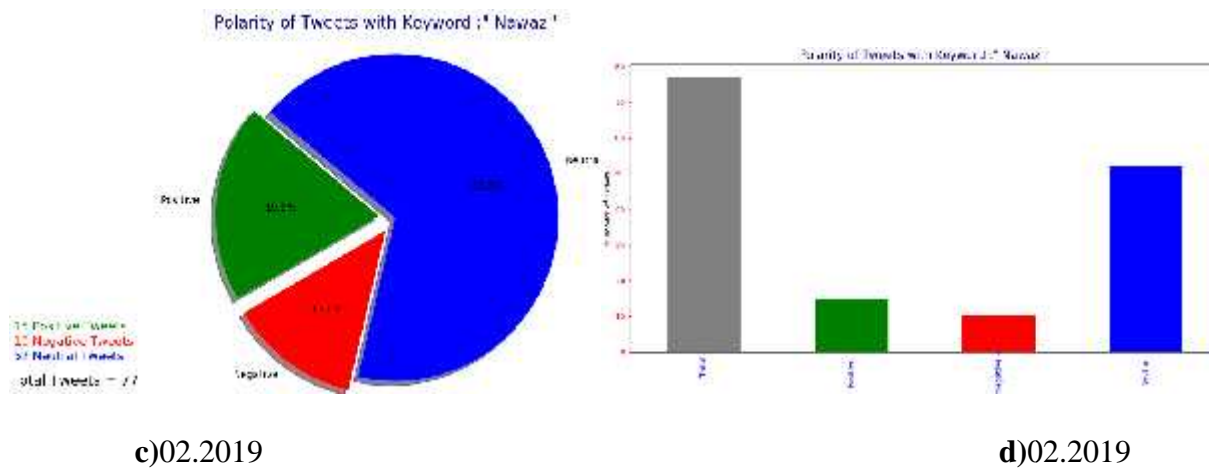
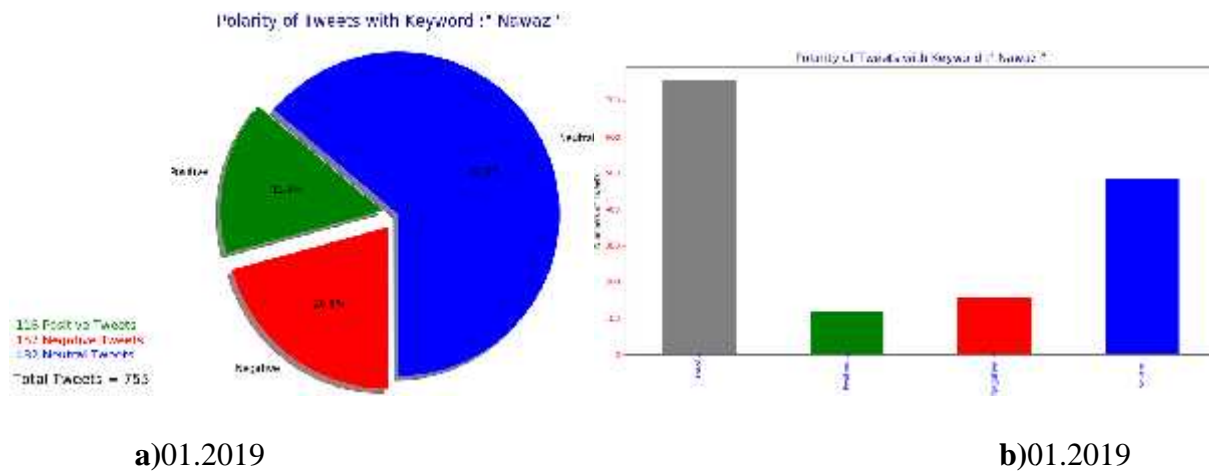


Figure 4.15:Sentiments analysis evolution of keyword Nawaz

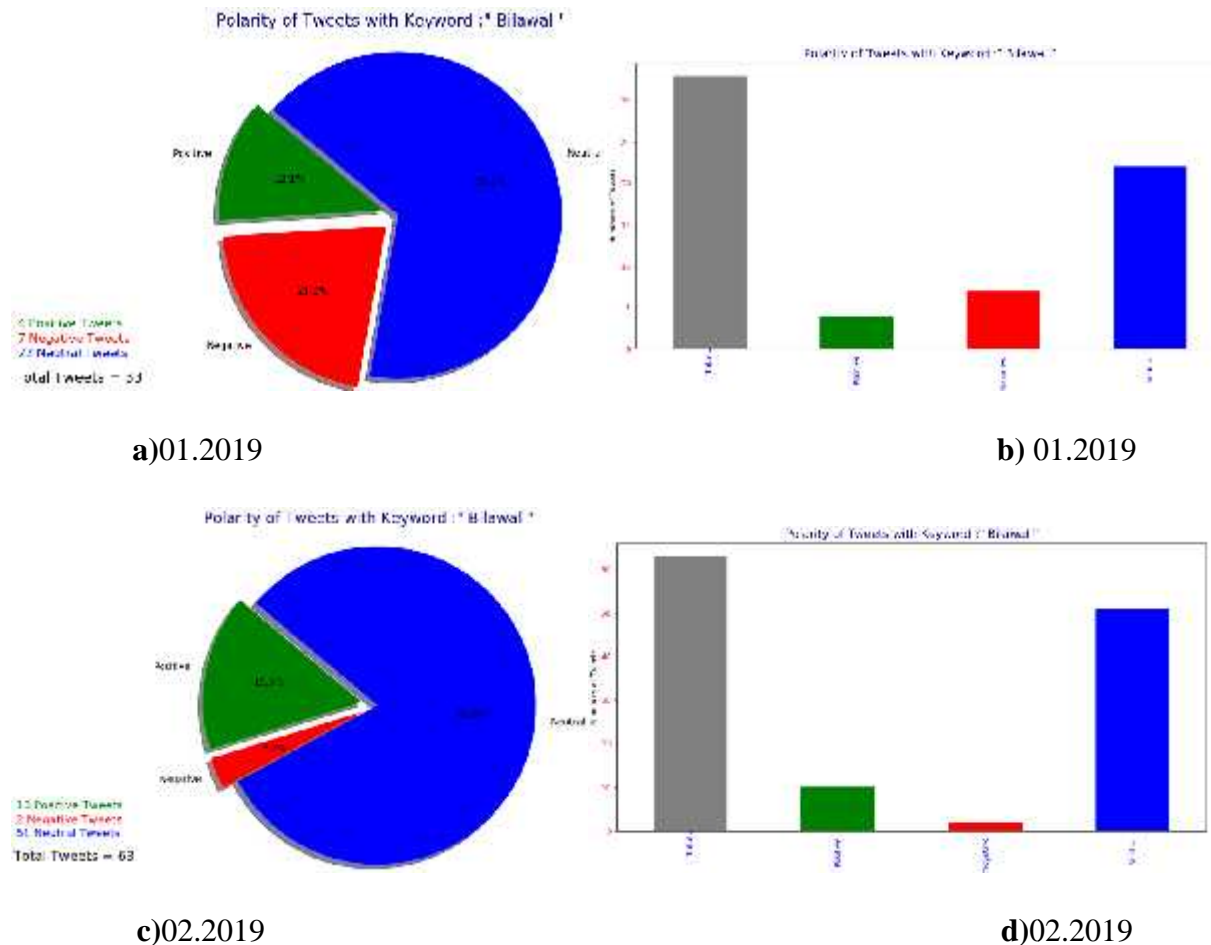


Figure 4.14: Sentiments analysis evolution of keyword Bilawal

4.5.1 Pakistan Case Study Analysis

These research results show the more positivity about the current prime minister of Pakistan as compared to the other two opposition party. Mostly people like the current prime minister of Pakistan Imran Khan and gave positive tweets about him and a greater number of tweets as compared to opposition (Nawas Sharif and Bilawal Bhutto). The all results shows in the table (4.5) with the keywords ‘P’ used for positive, ‘N’ used for negative, ‘NT’ used for neutral, ‘T’ used for total, ‘MXP’ used for maximum positive, ‘MNP’ used for minimum positive, ‘PA’ used for positive average, ‘MXN’ used for maximum negative, ‘MNP’ used for minimum negativity and ‘NA’ used for negative average of the sentiments percentage. Analysis of the results from these datasets shows that Imran Khan positive 27% and negative 8.1% average of the results as compared to Nawaz Sharif positive of 15.74% and negative of 20% average of results and

Bilawal Bhutto positive of 14.58% and negative of 9.3% average of results which is lesser than Imran Khan.

Table 4.4: Sentiments analysis results of Pakistan case

Date	Imran Khan	Nawaz Sharif	Bilawal Bhutto
12.2018	P=26.6=587 N=8.3=183 NT=65.1=1438 T=100=2208	P=15.4=116 N=20.8=157 NT=63.8=482 T=100=755	P=12.1=4 N=21.2=7 NT=66.7=22 T=100=33
01.2019	P=29.2=132(2.6) ↑ N=7.7=35 NT=63.1=285 T=100=452	P=19.5=15(3.9) ↑ N=13=10 NT=67.5=52 T=100=77	P=15.9=10(3.8) ↑ N=3.2=2 NT=81.9=51 T=100=63
Overalla	MXP=29.2 MNP=26.6 PA=27 MXN=8.3 MNN=7.7 NA=8.1	MXP=19.5 MNP=15.4 PA=15.74 MXN=20.8 MNN=13 NA=20	MXP=15.9 MNP=12.1 PA=14.58 MXN=21.2 MNN=3.2 NA=9.3

CHAPTER 5

CHALLENGES AND CONCLUSION

In this research faced different kind of challenges and most difficult challenge is collecting the data (tweets) from twitter because when the same search keyword is queried every day, the twitter disconnects your connection and stop the data collecting process. After data collection, another challenging task is data cleaning. Duplication of tweets is also major issue, since the same tweets appear in search results and affect healthy result analysis. Multi language or non-English tweets this one is typical issue during the sentiment's analysis because non-English tweet count as neutral. As a conclusion, in this research we gave a way to predict useful results about politics because every person directly or indirectly connected with politics through votes. The microblogging is mostly using weapon in technology and twitter is useful resource for prediction. In parliament voting in January 2019, we were expecting a reject due to positive thoughts about EU as well as changing opinions about Brexit. Positive average of Brexit was 31.05% and negative average was 23 as well as positive average of EU was 37.46 and negative average was 21.4. It means that positive average of EU was around 6.4% higher than the Brexit. On the other hand, Theresa May's positive and negative average was 29.7 and 26.45 respectively compared to Jeremy Corbyn's 40.6 positive average and 16.65 negative average. It means that people are more supportive of Jeremy Corbyn as compared to Theresa May according to the Tweet datasets we collected in January and February 2019. Our findings are correlating with the rejection decision that was made in February 2019. This research is domain independent, it means we can apply on the any field not only the politics, such as social data, stock market and medically so forth. We can also apply our software not only Twitter, but any other text datasets. In future work, we will support multi keywords and multi results on the same view.

REFERENCES

- Barril, J. F. H., & Tan, Q. (2017, April). Integrating privacy in architecture design of student information system for big data analytics. *In 2017 IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)* (pp. 139-144). IEEE.
- Boutet, A., Kim, H., & Yoneki, E. (2012, May). What's in your tweets? I know who you supported in the UK 2010 general election. *In Sixth International AAAI Conference on Weblogs and Social Media*.
- Bermingham, A., & Smeaton, A. (2011). On using Twitter to monitor political sentiment and predict election results. *In Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011)* (pp. 2-10).
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. "O'Reilly Media, Inc."
- Bovet, A., Morone, F., & Makse, H. A. (2018). Validation of Twitter opinion trends with national polling aggregates: Hillary Clinton vs Donald Trump. *Scientific reports*, 8(1), 8673.
- Bhadani, A. K., & Jothimani, D. (2016). Big data: challenges, opportunities, and realities. *In Effective Big Data management and opportunities for implementation* (pp. 1-24). IGI Global.
- Gilbertson, S. (2011). Twitter Vulnerability: Spoof Caller ID To Take Over Any Account. *Webmonkey*. Retrieved February, 5.
- Jaidka, K., Ahmed, S., Skoric, M., & Hilbert, M. (2019). Predicting elections from social media: a three-country, three-method comparative study. *Asian Journal of Communication*, 29(3), 252-273.
- Kuhlman, D. (2009). *A python book: Beginning python, advanced python, and python exercises* (pp. 1-227). Lutz: Dave Kuhlman.
- Livne, A., Simmons, M., Adar, E., & Adamic, L. (2011, July). The party is over here: Structure and content in the 2010 election. *In Fifth international AAAI conference on weblogs and social media*.

- Obar, J. A., & Wildman, S. S. (2015). Social media definition and the governance challenge-an introduction to the special issue. *Obar, JA and Wildman, S.(2015). Social media definition and the governance challenge: An introduction to the special issue.Telecommunications policy, 39(9), 745-750.*
- Tansley, S., & Tolle, K. M. (2009). *The fourth paradigm: data-intensive scientific discovery* (Vol. 1). A. J. Hey (Ed.). Redmond, WA: Microsoft research.
- Wang, X., Fang, Z., &Guo, X. (2016). Tracking the digital footprints to scholarly articles from social media. *Scientometrics, 109(2), 1365-1376.*
- Stokes, W. (2018). The government of the United Kingdom: The Women’s National Commission.In *Mainstreaming gender, democratizing the state?*. Manchester University Press.
- Xi, H., &Pfenning, F. (1999, January). Dependent types in practical programming. In *Proceedings of the 26th ACM SIGPLAN-SIGACT symposium on Principles of programming languages* (pp. 214-227). ACM.
- Tosi, S. (2009). *Matplotlib for Python developers*. Packt Publishing Ltd.
- Sewalk, K. C., Tuli, G., Hswen, Y., Brownstein, J. S., & Hawkins, J. B. (2018). Using Twitter to Examine Web-Based Patient Experience Sentiments in the United States: Longitudinal Study. *Journal of medical Internet research, 20(10), e10043.*
- Agarwal, A., Xie, B., Vovsha, I., Rambow, O., &Passonneau, R. (2011). Sentiment analysis of twitter data. In *Proceedings of the Workshop on Language in Social Media (LSM 2011)* (pp. 30-38).
- Kwak, H., Lee, C., Park, H., & Moon, S. (2010, April). What is Twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web* (pp. 591-600). AcM.

APPENDIX

Sentiments Analysis Code

```
fromtextblob import *
```

```

import pandas as pd
import matplotlib.pyplot as plt
import csv
import os
while True:
total=0
positive=0
negative=0
neutral=0
path=(r'c:\\Users\\pak_cyp\Desktop\\New folder\\18.1.19.csv') #put your csv file path here
t=csv.reader(open(path,newline="", errors='ignore'),delimiter=" ")#

oneword=str(input("Keyword:"))#The word you want to search in file
csv.field_size_limit(11000000)
for row in t:
row=" ".join(str(x) for x in row)
if oneword in row:
r=TextBlob(str(row))
print("".join(row))
print("          Polarity = {}".format(r.sentiment.polarity))
print("X-----X")
print("\n")
total+=1
s=r.sentiment.polarity
if s==0:
neutral+=1
if s>0:
positive+=1

```

```

if s<0:
    negative+=1

fig=plt.figure()
ax = fig.add_subplot(111) #
fig.set_size_inches(12, 8)
plt.title('Polarity of Tweets with Keyword : " {}'.format(oneword),color="navy",fontsize=16)

ax.text(0.01,0.2, '{} Positive Tweets'.format(positive),verticalalignment='bottom',
horizontalalignment='left',transform=ax.transAxes,color='green', fontsize=12)
ax.text(0.01,0.16, '{} Negative Tweets'.format(negative),verticalalignment='bottom',
horizontalalignment='left',transform=ax.transAxes,color='red', fontsize=12)
ax.text(0.01,0.12, '{} Neutral Tweets '.format(neutral),verticalalignment='bottom',
horizontalalignment='left',transform=ax.transAxes,color='blue', fontsize=12)
ax.text(0,0.05, " Total Tweets = {}".format(total),verticalalignment='bottom',
horizontalalignment='left',transform=ax.transAxes,color='black', fontsize=14)

labels=["Positive","Negative","Neutral"]
sizes=[positive,negative,neutral]
colors=["green","red","blue"]
explode=(0.1,0.1,0)

plt.pie(sizes, explode=explode, labels=labels, colors=colors,autopct='%1.1f%%', shadow=True,
startangle=140)
plt.axis('equal')#
plt.show()

s =
pd.Series([total,positive,negative,neutral],index=["Total","Positive","Negative","Neutral"])

```

```

fig2=plt.figure(2)
fig2.set_size_inches(12, 8)
#Set descriptions:
plt.title('Polarity of Tweets with Keyword : " {} "'.format(oneword),color="navy",fontsize=15)
plt.ylabel("Numbers of Tweets")
#Set tick colors:
ax = plt.gca()
ax.tick_params(axis='x', colors='blue')
ax.tick_params(axis='y', colors='red')

#Plot the data:
my_colors = ["grey","green","red","blue"]

s.plot(kind='bar', color=my_colors)

plt.show()

```

Data Streaming Code

```

from tweepy import Stream
from tweepy import OAuthHandler
from tweepy.streaming import StreamListener
import time
import json
import csv
import datetime

consumer_key= "D5f5APKIClbG4I2dKJgb6qgIq"
consumer_secret = "nQKB9KqQURLciIAuxZh0MPx595bHr6SCAPbnWT647jYv13IfZI"
access_token = "808861882887110658-29osI1eccBAX5KlndmDv3zq8vuGHGRl"

```



```
access_token_secret = "AX79IFZA6umX1UVzYZ0qDKjG4UAo1oCs9KVUnzwKWltiv"
```

```
#This is a basic listener that just prints received tweets to stdout.
```

```
classStdOutListener(StreamListener):
```

```
defon_data(self, data):
```

```
print (data)
```

```
myfile=open(r'c:\\Users\\pak cyp\Desktop\\New folder\\bhrat.csv','a')
```

```
myfile.write(data)
```

```
myfile.close()
```

```
return True
```

```
defon_error(self, status):
```

```
print (status)
```

```
if __name__ == '__main__':
```

```
    #This handles Twitter authentication and the connection to Twitter Streaming API
```

```
    l = StdOutListener()
```

```
    auth = OAuthHandler(consumer_key, consumer_secret)
```

```
    auth.set_access_token(access_token, access_token_secret)
```

```
    stream = Stream(auth, l)
```

```
    stream.filter(track=['Modi','Gandhi'])
```

Report

The screenshot shows the Turnitin web interface. At the top, there is a navigation bar with the Turnitin logo and various menu items like 'MyInfo', 'Messages', 'Instructor', 'English', 'Community', 'Help', and 'Logout'. Below this is a secondary navigation bar with buttons for 'Assignments', 'Students', 'Grade Book', 'Library', 'Calendar', 'Discussion', and 'Preferences'. The main content area is titled 'NOW VIEWING: HOME - MASTER STUDENTS 2018 2019 - THESES'. Underneath, there is a section 'About this page' explaining how to view papers and similarity reports. The 'Thesis' section includes a 'Submit File' button and a table of submitted papers. The table has columns for 'Action', 'TITLE', 'SIMILARITY', 'GRADE', 'REVISIONS', 'FILE', 'PAPER ID', and 'DATE'. The data rows show papers submitted by 'Muntazar M' with titles like 'Abstract', 'Chapter 5', 'Chapter 4', 'Chapter 1', 'Complete Thesis', 'Chapter 3', and 'Chapter 2', along with their respective similarity percentages and dates.

Action	TITLE	SIMILARITY	GRADE	REVISIONS	FILE	PAPER ID	DATE	
<input type="checkbox"/>	Muntazar M	Abstract	0%	—	—		1129540255	10-Apr-2019
<input type="checkbox"/>	Muntazar M	Chapter 5	0%	—	—		1129540109	10-Apr-2019
<input type="checkbox"/>	Muntazar M	Chapter 4	0%	—	—		1129545838	10-Apr-2019
<input type="checkbox"/>	Muntazar M	Chapter 1	7%	—	—		1129540444	10-Apr-2019
<input type="checkbox"/>	Muntazar M	Complete Thesis	4%	—	—		1129547414	10-Apr-2019
<input type="checkbox"/>	Muntazar M	Chapter 3	6%	—	—		1129541853	10-Apr-2019
<input type="checkbox"/>	Muntazar M	Chapter 2	10%	—	—		1129540193	10-Apr-2019

Copyright © 1999 - 2019 Turnitin, LLC. All rights reserved.

[Privacy Policy](#) | [Privacy Pledge](#) | [Terms of Service](#) | [EU Data Protection Compliance](#) | [Copyright Protection](#) | [Legal FAQs](#) | [Helpdesk](#) | [Research Resources](#)