

SENTIMENT ANALYSIS OF POLITICAL TWEETS

Project report submitted in partial fulfillment of the requirement for the
degree of Bachelor of Technology

in

Information Technology

By

Tashi Wangmo 131403

Under the supervision of

Ms. Nishtha Ahuja

to



Department of Computer Science & Engineering and Information
Technology
**Jaypee University of Information Technology Waknaghat, Solan-
173234, Himachal Pradesh**

Candidate's Declaration

I hereby declare that the work presented in this report entitled “**Sentiment Analysis of political tweets**” 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 August 2016 to May 2017 under the supervision of (**Ms. Nishtha Ahuja**) (Assistant Professor (Grade-I) Department of Computer Science and Engineering).

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

Tashi Wangmo (131403)

.

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

Supervisor Name: Ms. Nishtha Ahuja

Designation : Assistant Professor (Grade-I)

Department name: Department of Computer Science and Engineering

Dated:

ACKNOWLEDGEMENT

I would like to express my gratitude to all those who gave me the possibility to complete this project. I want to thank the Department of CSE & IT in JUIT for giving me the permission to commence this project in the first instance and to do the necessary research work.

I am deeply indebted to my project guide Ms. Nishtha Ahuja, whose help, stimulating suggestions and encouragement helped me in all the time of research on this project. I felt motivated and encouraged every time I got her encouragement. For her coherent guidance throughout the tenure of the project, I feel fortunate to be taught by her.

I am also grateful to **CSE Project lab staff** for their practical help and guidance.

Tashi Wangmo (131403)

TABLE OF CONTENTS

CERTIFICATE	i
ACKNOWLEDGEMENT	ii
LIST OF FIGURES.....	v
LIST OF TABLES	vi
ABSTRACT.....	vii
1.) INTRODUCTION.....	1
1.1) Introduction.....	1
1.2) Problem Statement	2
1.3) Objective	3
1.4) Methodology and Organization.....	3
2.) LITERATURE SURVEY.....	8
2.1) Political Sentiment Analysis using hybrid approach.....	8
2.2) Predicting Elections with Twitter:	9
2.3) Sentiment analysis of political tweets:	10
2.4) To what degree sentiment analysis of Twitter can forecast constituent outcomes....	11
2.5) what are the limitations of Electoral Predictions Using Twitter.....	12
2.6) Sentiment Analysis of Twitter Data.....	13
2.7) To predict Twitter Trend: Significant Factors and Models.....	14
2.8) To Learn What Hashtags Will Be Prevalent Where.....	15
2.9) To predict Bursts and Popularity of Hashtain Real-Time.....	16
2.10) Utilizing Topic Models for Twitter Hashtag Recommendation.....	17

2.11) Hashtags Performance Forecast and Multi- Report Rundown.....	18
2.12) Using Machine-Learning approach for election prediction.....	19
2.13) Predicting the EU 2014 Election Results in various Countries Using Twitter.....	20
2.14) Sentiment Analysis of Twitter Data.....	21
2.15) Suggesting Hashtags on Twitter.....	22
3.) SYSTEM DEVELOPMENT	23
3.1) Eclipse IDE Juno.....	23
3.2) Twitter.....	24
3.3) Twitter 4J	26
3.4) Hashtags.org	27
3.5) Creating a twitter app.....	28
3.6) Setting up twitter 4J.....	30
3.7) Sentiment Analysis.....	32
3.8) Algorithm.....	38
4.) PERFORMANCE ANALYSIS	40
4.1) Sun security validation certificate	40
4.2) Extracting Tweets.....	42
4.3) Sentiment analysis using Lexical approach.....	43
4.4) Result.....	44
5.) CONCLUSION.....	47
5.1) Conclusion	47
5.2) Future work.....	48
6.) REFERENCES.....	49

List of Figures

1.) Methodology Flow Chart.....	4
2.) Lexical orientation- Dictionary based approach.....	7
3.) Eclipse IDE Juno.....	23
4.) Twitter accounts.....	24
5.) Retweet example.....	25
6.) Hillary hour trend graph.....	27
7.) Trump hour trend graph.....	28
8.) Consumer Key and Consumer Secret.....	29
9.) Access token and Access Token Secret	29
10.) Adding External Jars.....	31
11.) External Jars Added.....	31
12.) SentiWordNet.....	34
13.) SentiWordNet fragment.....	35
14.) Percentage polarity.....	37
15.) Sun Security Validation Error on Execution.....	40
16.) Importing the Certificate.....	41
17.) Certificate Imported Successfull.....	41
18.) Output with Hashtag.....	42
19.) Output with Tweets Generated.....	43
20.) Output using lexical approach.....	44
21.) Pie chart showing Hillary’s positive, negative and neutral tweets.....	46
22.) Pie chart showing Trump’s positive, negative and neutral tweets.....	46

List of Tables

1.) Total number of tweets (positive, negative, neutral and spam) for each candidate	44
2.) Percentage of tweets (positive, negative and neutral) for each candidate.....	45

Abstract

Social media has proven rapid growth in becoming the most popular activities in internet and to appeal the researchers to get better insights into the understanding of the underlying sociology. In olden days traditional form of navigation and hypertext browsing was used which is now replaced by tags (used in twitter) and it is less time consuming. Twitter is one popular web application nowadays. Twitter allows users to use hash tags to extract only those tweets related to a particular category. Twitter is called a micro-blog because anyone can post short messages up to 140 characters in length. We use our past experiences when we make decisions such as to take various selections in our everyday life whilst buying, making an investment money, handling others, deciding on our favourite celebrity, celebrities, electing our ministers. It was quoted by somebody “who learns from his own mistakes is smart, but who learns from others mistake is genius”. Sentiment analysis enables us to use others past experiences to make right decisions.

In sentiment analysis we collect reviews from social networks, in my project I extracted from twitter, and analyse these reviews in a way that can convey us the writer’s opinion, which will help us in making plans in future. Sentiment analysis is a challenging task because to do the analysis it requires a deep knowledge of implicit and explicit, we need to consider the regular and irregular expressions and grammatical rules of the language. My project is based on politics; Sentiment Analysis can help the political parties when they make future policies, their party promotion and is also helpful in predicting the election results. As soon as the time for election start people also begin sharing their opinion on social networking sites and these sites during peak time of election gets flooded with the online discussions about political parties, candidates representing the parties, bunches of disputable discourses and level headed discussions are held over social networks. And when people share thoughts and discuss, it give us opportunity to use these as a resource for study and analysis. In our research we are using twitter as our data source and we are applying lexical approach of sentiment analysis. In Lexical approach it works on bag of words to extract the sentiments, it considers each word in a sentence and these words are matched with pre-set dictionary words.

CHAPTER-1 INTRODUCTION

1.1 Introduction

Sentiment Analysis is the process of which finds whether a word or sentence or a document is positive, negative or neutral. SA is also known as opinion mining. This innovation is used commonly to discover how different individuals feel about a certain topic. The applications for sentiment analysis are endless. It can be applied to customer reviews, survey responses, competitors, etc. Its benefit is popular in business analytics and mostly in situations where text needs to be analysed. Sentiment Analysis aims to discover opinions, identify sentiments and afterwards classify those sentiments into various categories. A well-defined and an accurate system for predicting sentiments could enable us, to extract sentiments from the internet and forecast social behaviour, political drifts, and evolving parties from a particular geographical location.

However when we do sentiment analysis various challenges are faced because people don't always express their feelings in a same way, from a particular instance the sentence might seem positive and might seem negative from a different point of view. More over there is possibility of wrong spellings, intensifiers and spams confuses us when we do the analysis and there are millions of ways to join sentences and to treat negations which is even more stimulating. As the growth of existing subjective text on the internet is increasing rapidly, in order to achieve more refined, realistic and subjective opinion on companies and products, people uses internet. Lexicon approach is one of the analysis technique which gives each word a sentiment score. There are three sentiment scores - positive, negative and neutral.

1.2 Problem Description

Problem Statement

Sentiment analysis is not an easy task and moreover there are many problems with it. Sentiment is naturally subjective from individual to individual and sometimes it can be absolutely absurd. If there are more than one observer to the test, it is quite obvious that opinion we share will be different from others. This is really an issue regularly experienced with sentiment analysis. What matters is the aggregate because no specific information point is essentially applicable. Sometimes an individual's feeling towards an item may be impacted by some indirect reasons; she might have an awesome day and give a positive and warm comment about something for which she generally had a quite impartial feeling about.

With a huge sample, outliers are diluted when we do the summation. It is normally essential to look at data from the perspective of time, since it is probable that the sentiment might vary over time with change in person's mood, the surrounding and situations etc. Deciphering the mind-set of a subject may also differ from a person to another. Sentiment analysis have other more issues like:

- To find ways to discover the profundity of emotions. Just to classify into positive and negative is simple but to extract emotions like amount of hatred in an opinion, amount of happiness, and sadness, etc. is quite challenging.
- To find ways that will show that the opinion is positive for that particular entity and negative for the other. For example, if you say "She won him!" this could convey a positive feeling towards her and a negative towards him, at the same time.
- To find ways to scrutinize very subjective sentences or paragraphs.

1.3 Objectives

The main aim of our project is:-

- ▶ Sentiment analysis of political tweets using lexical approach
- ▶ Particularly around election time, social media becomes a good platform for political talks. It is intriguing to perceive how precisely can Twitter inform us about the electorate's political opinion?

1.4 Methodology and Organisation

To achieve the above mentioned objective we tend to follow the following methodology:-

Phase 1: Data Collection

Data collection of the upcoming presidential elections in US (i.e to find out which parties are competing against which other parties and which candidates are representing this parties). It was the 58th US presidential election which was held on Tuesday November 8, 2016 and the latest quadrennial American presidential election. From the Republican there was businessman Donald Trump and Indiana Governor Mike Pence who won over the Democratic Party's nominees Hillary Clinton and Tim Kaine. President Barack Obama could not run again due to term limits.

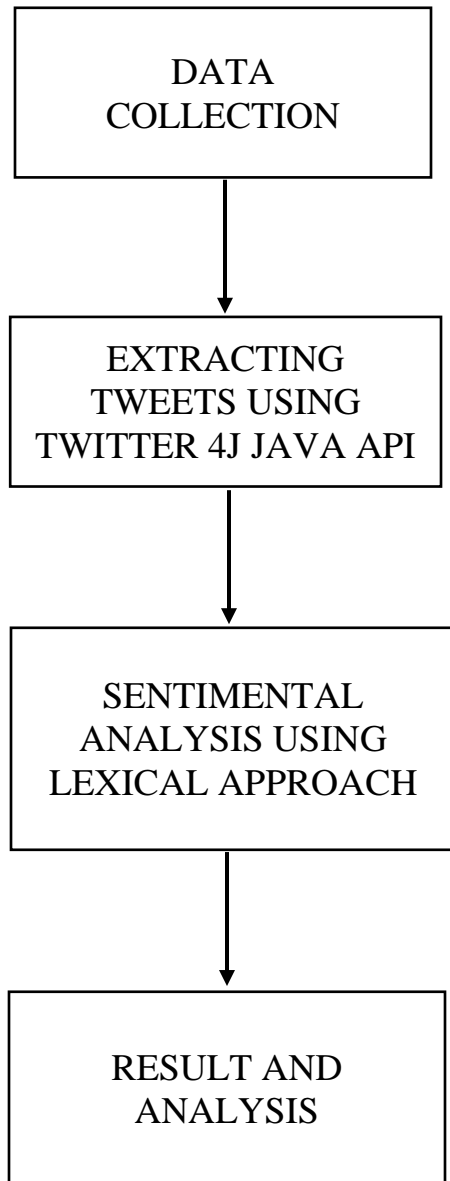


Figure 1: Methodology Flow Chart

Phase 2: Extracting tweets using twitter 4j

The data required in our project is US election tweets, so we extract filtered data from Twitter, i.e. data that corresponds to a US presidential election which will be the system input. The primary requirement for building a database to get processed is to extract the tweets. Twitter4j Java API is an unofficial third party Java library for the Twitter API which is used to extract the tweets through authenticating connection with Twitter web sites. The proposed method is used for extracting election data from Twitter.

Below are the steps that describes how to extract tweets from twitter web site

- (1) Initially for experimental setups, we had selected an election with a specific hashtag (#) which was related to that particular election. To identify the trending hash tags in the entire twitter network we used an online service called HashTags.org. This constituted the first means, to identify which keywords should be adopted in order to filter the initial amount of tweets and trace other related hashtag on the same topic.
- (2) We used Twitter4j API to connect to Twitter. Twitter4j is a viable and flexible library for getting connected to twitter and to communicate with the customer application. In order to connect to the twitter, we need to setup consumer key, consumer secret key, access tokens and access token secret key in application code. Authentication of requests to Twitter API are based on the open authentication 1.0 standard in which two pairs of keys are used in the authentication, namely, consumer key and access token and both can be obtained from the Twitter web site application section.
- (3) Twitter sends request to twitter database server and server responds to tweets within a particular period that is already setup in application.

- (4) Extracted tweets were sent to an application and stored in a database in tabular format. The user can access election tweets anytime that are stored in MySQL DB just by sending a request to DB and DB will respond to related movie tweets via application GUI.
- (5) The extracted tweets were stored into the database which was used in sentiment analysis. Further the downloaded database file (excel format) and then that file is used as input in the sentiment analysis classification model.

Phase 3: Sentimental Analysis

Sentiment analysis extracts and analyzes sentences to identify whether it conveys positive, negative and neutral opinions and is useful while making decisions. When different users exhibit their social, religious and political views on Twitter, then their tweets are one of the most valuable sources of people's opinions. To perform sentiment analysis of Twitter data, we used lexical approach.

The figure below shows the different methods to be followed in the Lexical approach; the method starts from microblogging website (i.e twitter); which is the basis of our text dataset; we collect text reviews form twitter and store those review for future processing. In the cleaning part we take data from database (tweets stored in csv format) and remove unwanted contents from it and make it ready for Natural Language Processing; where our tweets get classified into sentiments of different categories like positive, negative and neutral.

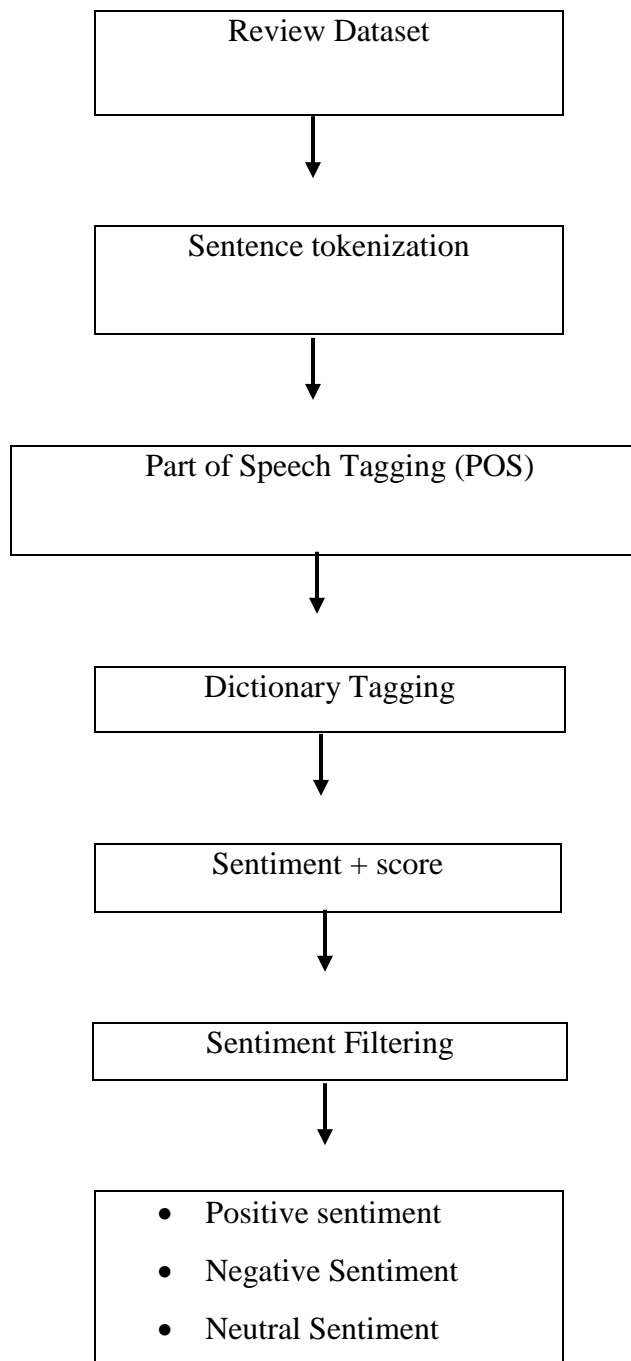


Figure 2: Lexical orientation- Dictionary based approach

CHAPTER 2 – LITERATURE SURVEY

Summary of Research Papers

Title Of Papers	Political Sentiment Analysis using hybrid approach
Authors	Vikashh Naindi and Suvyash Agrrawal
Year of Publication	2016
Technique used	Lexical approach
Summary	<p>They first compared two approaches of sentiment analysis, i.e lexical approach and machine learning feature extraction. They found some problems with each of this approach, so they took a hybrid of the two approaches to accomplish more efficiency.</p>

Title of Papers	To predict elections using twitter:
Authors	Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner, Isabell M. Welp
Year of Publication	2010
Techniques Used	Linguistic Inquiry and Word Count
Summary	In this paper they did analysis on almost 104,003 tweets related to German election to see if Twitter can be used as a paradigm for political deliberation .They also investigated if the tweets on internet could sincerely show political sentiments in the real world. And they discovered many shades of the election crusade from the political sentiments they obtained from the tweets.

Title Of Papers	Sentiment analysis of political tweets.
Authors	AkshatBakliwal, JenniferFoster, Jennifer van der Puil, RonO'Brie, LamiaTounsi and Markhes
Year of Publication	2013
Publishing Details	In proceedings of the seminar on Language in Social Media
Summary	They utilized a supervised learning and a list of capabilities comprising of subjectivity dictionary construct scores in light of an arrangement of 2620 tweets created from an Irish decision in 2011. They removed those tweets that was labeled as sarcastic and yet it represented a problematic test set. They got an accuracy of 61.6% as a result.

Title of Papers	To what degree sentiment analysis of Twitter can forecast constituent outcomes?
Authors	Luigi Curini – Università degli Studi di Milano, luigi.curini@unimi.it Andrea Ceron
Year of Publication	2013
Techniques used	Hopkins and King (HK) method
Summary	They implemented a supervised sentiment analysis on social networks in three different political scenarios (Italy, France, and the United State) This method improved the traditional methods yielding more accurate results. They discovered that their investigation shows a consistent correlation between outcomes from social media and the outcomes they obtain from traditional ways of mass surveys. They also discovered that a prominent talent of social media to forecast electoral results.

Title of Papers	What are the limitations of Electoral Predictions Using Twitter
Authors	Danieel Gayo-Ayvello , Panagiotis T. Metaxas and Eni Murstafaraji
Year of Publication	2011
Techniques used	Twitter streaming API, polarity lexicon
Summary	They used polarity lexicon to find positive, negative, and neutral words. They did not include those tweets which mention both opposing candidates. They have noted that the intricacy of professional polling cannot be replicated by sampling social media data (because not every adult who has the right to vote will perform it.) Moreover social media is influenced by spammers and propagandists. This research has revealed that data from Twitter was no very much dependable in predicting results in the last US elections.

Title of Papers	Sentiment Analysis of Twitter Data
Authors	Johan Bollen ,Huina Mao ,Alberto Pepe
Year of Publication	2011
Techniques used	Profile of Mood States (POMS)
Summary	<p>They did a research on how public mood patterns relate to variations in macroscopic social and economic displays. They presented results for sentiment analysis on Twitter. They utilized beforehand proposed best in class unigram display as their model and revealed a general pick up of more than 4% for two arrangement errands:</p>

Title of Papers	To predict Twitter Trend: Significant Factors and Models
Authors	Pehng Zhiong, Xufei Wanig and Boxin Lei
Year Of Publication	2011
Publishing Details	In Proceedings of the twentieth global meeting on World Wide Web.
Summary	They concentrated two essential issues in data incline forecast, i.e. vital variables and proper models. They composed elements of various pattern variables from both tweet substance and system setting. The likewise researched demonstrate classifications as the blend of two fundamental properties, i.e. (non)- linearity and (non)-state-space.

Title Of Papers	To Learn What Hashtags Will Be Prevalent Where
Authors	Kisshna Y. Karmath and Jamesy Cavelee
Year of Publication	2011
Publishing Details	In Proceedings of the thirteenth global meeting on World Wide Web.
Summary	They handled the issue of foreseeing what online images will be mainstream in what areas. In particular, they created information driven methodologies expanding on the worldwide impression of 755 million geo-labeled hashtags spread by means of Twitter. Their proposed techniques show the geo-spatial engendering of online data spread to distinguish which hashtags will wind up noticeably well known in particular areas. In tests, they found that the proposed strategy beats elective straight relapse based techniques.

Title Of Papers	To predict Bursts and Popularity of Hashtain Real-Time
Authors	Shouibin Kung, Qaozhiu Meii, Linu Fenlg
Year of Publication	2011
Publishing Details	In Proceedings of the thirteenth global meeting on World Wide Web.
Summary	Hashtags have been broadly used to comment on themes in tweets. In this paper, they studied the complications in real-time prediction of abounding hashtags. In light of observational examination of information gathered from Twitter, they suggested solutions to these thought-provoking problems of real time prediction. They even evaluated the performance of various features and probable solutions.

Title Of Papers	Utilizing Topic Models for Twitter Hashtag Recommendation
Authors	Frédéric Godiin, Vikytor Slakovikj and Wesleyy De Neive
Year of Publication	2011
Publishing Details	In Proceedings of the thirteenth global meeting on World Wide Web.
Summary	They proposed a method which recommends hashtags for tweets. These hashtags were unsupervised and content based. They also modelled the fundamental topic assignment of tweets which were classified in accordance to a particular language. This approach suggests common hashtags with the use of a topic distribution.

Title Of Papers	Hashtags Performance Forecast and Multi- Report Rundown
Authors	Ron Deveud and Floriian Bodin
Year of Publication	2011
Publishing Details	Reports of the Second International Conference on Human Language Technology Study
Summary	They utilized an approach which is made out of three principle segments: preprocessing, wikipedia articles recovery and multi-report rundown. They however exploited a bigger utilization of hashtags in the points and utilized them to upgrade the recovery of pertinent Wikipedia articles. They even exploited the preparation cases from a year ago which enabled them to take in the weights of each sentence determination include.

Title of Papers	Using Machine-Learning approach for election prediction
Authors	Colletto, Luccheise, S. Orlanndo, and R. Periego
Year of Publication	2011
Techniques Used	-Twitter API -Data cleansing -Mean Absolute Error(MAE) content-based analysis
Summary	In this paper they investigated how to exploit and improve those indicators (based on tweet and user volumes) to reduce the biasness of the Twitter user's sample. Customary methodologies display the engendering of thoughts in online networking by breaking down the topology of the social diagram. They used a hybrid method in predicting the spread of ideas conferring to their content.

Title of Papers	Predicting the EU 2014 Election Results in Various Countries Using Twitter
Authors	Adam Tsakalidis, Symeon Papadopoulos, Alexandra Cristea, & Yiannis Kompatsiaris
Year of Publication	2011
Techniques Used	-Linear Regression
Summary	In this paper they took a shot at anticipating the 2014 European Union (EU) Election brings about Germany, the Netherlands and Greece. They likewise regard the clients' voting goals as time-variation highlights.

Title of Papers	Sentiment Analysis of Twitter Data
Authors	BrendanO'Connor,RamnathBalasubramanyan, BryanR.Routledge, NoahA.Smith
Year of Publication	2012
Techniques Used	Text Analysis
Summary	In this paper they dealt with finding moderately straightforward estimation locator in light of Twitter information duplicates buyer certainty and presidential occupation endorsement surveys.

Title Of Papers	Suggesting Hashtags on Twitter
Authors	Allie Mazzia, and James Juett
Year of Publication	2011
Publishing Details	In Proceedings of the thirteenth global meeting on World Wide Web.
Summary	<p>Twitter is a micro-blogging site that continue to grow in fame, they were presented with the problem of how to effectively classify and search for posts. When they took a preview of Twitter, they found out that clients might sort their posts retaining hashtags, and any word or expression might be utilized as the classification. Opening investigations suggest their dataset is satisfactorily rich to separate instructive assumptions of words for some hashtags that will inspire a credulous Bayes show for hashtag suggestion given an inquiry post.</p>

CHAPTER 3 – SYSTEM DEVELOPMENT

3.1 Eclipse IDE Juno

Eclipse is the most comprehensively used Java integrated development environment (IDE). It is a platform for constructing integrated web and application progress tooling. It can run on numerous operating systems while giving an energetic integration with each underlying operating system. The key use of eclipse is that it can be used for evolving Java applications, however it can be utilized to create applications in other programming vernaculars. C, C++, JavaScript, PHP, Python, R etc.

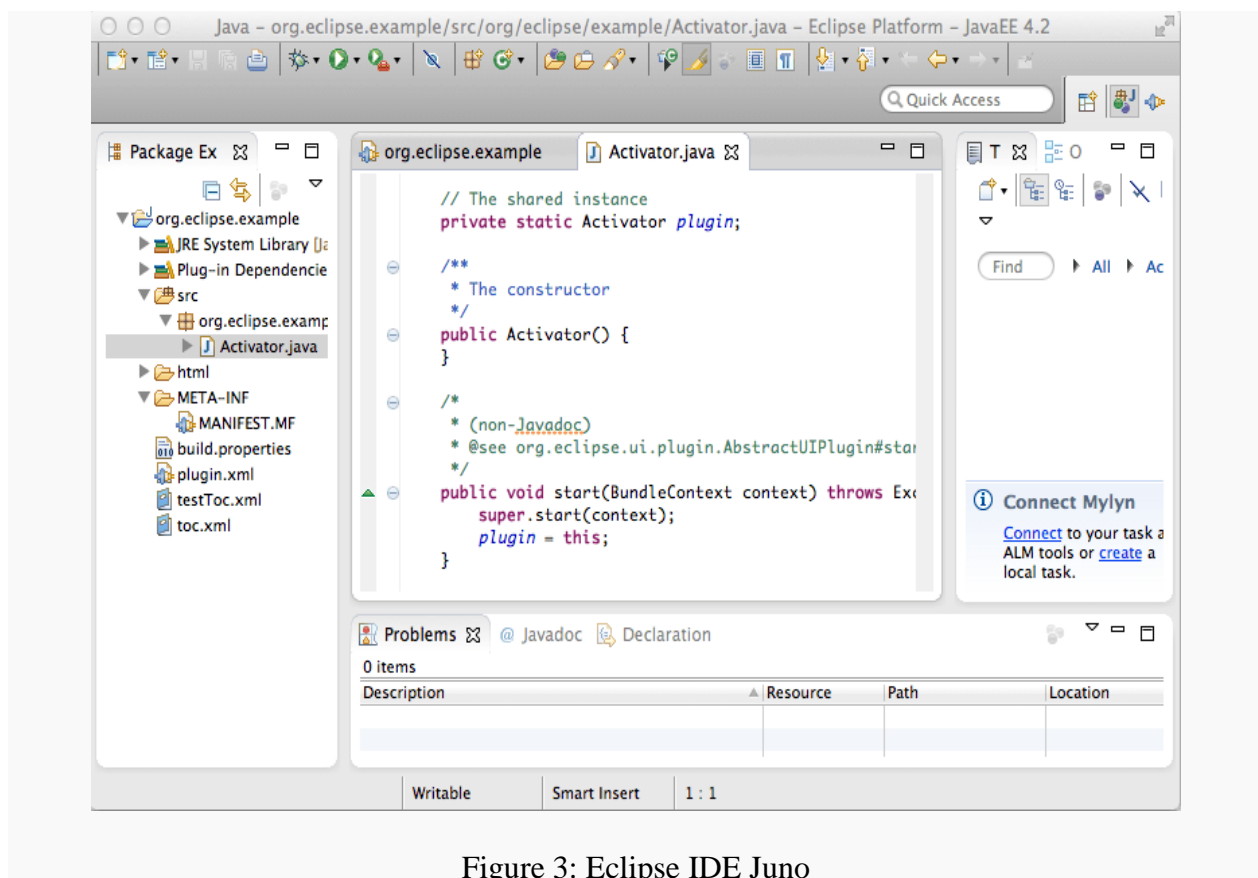


Figure 3: Eclipse IDE Juno

There are additionally various new communication and ease of use enhancements, including the capacity to blend perspectives and editors: The new 4.x workbench brings a wholly new programming model .The workbench is signified as an EMF model, and is dynamically condensed. This entirely separates the User Interface concerns from the application logic and allows for a much more vibrant workbench and a plug-able exhibition engine.

3.2 Twitter

Twitter and to tweet is to let the world know what you are up to every day by posting short messages which can be seen by many other people, with the anticipation that what you have messaged can be valuable and captivating to somebody. On the other hand, Twitter is likewise about finding fascinating individuals on the web, and taking after their burst messages for whatever length of time that they are intriguing.

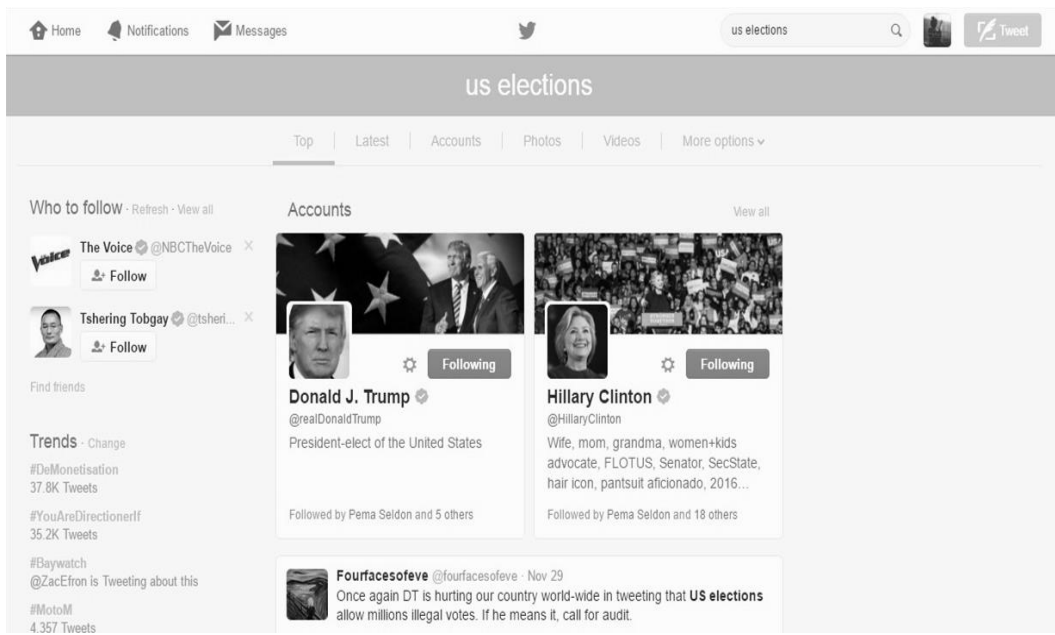


Figure 4: Twitter accounts

Twitter is exceptionally easy to use as telecaster or beneficiary. You join with a free record and Twitter name. At that point you send broadcasts every day, or even hourly. There is a 'What's Happening' box, we have to click there and categorize 140 characters or less, and click 'Tweet'. To get Twitter encourages, you essentially discover somebody fascinating (VIPs included), and "tail" them to subscribe to their tweet microblogs. Once a man ends up noticeably uninteresting to you, you basically "unfollow" them.

Retweet:

When we re-post a tweet, it is called retweet. Retweet is a feature in twitter that allows you and others to speedily share your views and thoughts through tweets with all of your followers. Anybody can retweets their own tweets or Tweets posted by somebody else. At times individuals sort RT toward the start of a Tweet to show that they are re-posting another person's substance.



Figure 5: retweet example

3.3 Twitter 4j

Twitter4J is an unlicensed Java library for the Twitter API.

Twitter4J is valuable since makes us simpler to coordinate our Java application with the Twitter benefit.

Twitter4J has the following features:

- It is Purely Java (It works on any Java Platform version 5 or advanced)
- It has an android platform and Google App Engine prepared
- It has no dependency and doesn't require extra jars
- It has a built-in OAuth support
- It supports with Out-of-the-box gzip
- It is completely Twitter API 1.1 compatible

System Necessities

Operating System: Windows or any kind of UNIX that assists Java.

Java Virtual Machine: we can have Java 5 or more advanced

Way to use

Twitter4j-core-4.0.4.jar is included to our application class path.

3.4 Hashtags.org

A hashtag is an unmistakable plan to arrange messages on Twitter and is viewed as an exceptionally powerful showcasing device by organizations. The # symbol leads the tag followed by one or more keywords that will appropriately lead individuals to conversations and discussions relating to a specific topic or theme. There are many benefits that come with using hashtags.

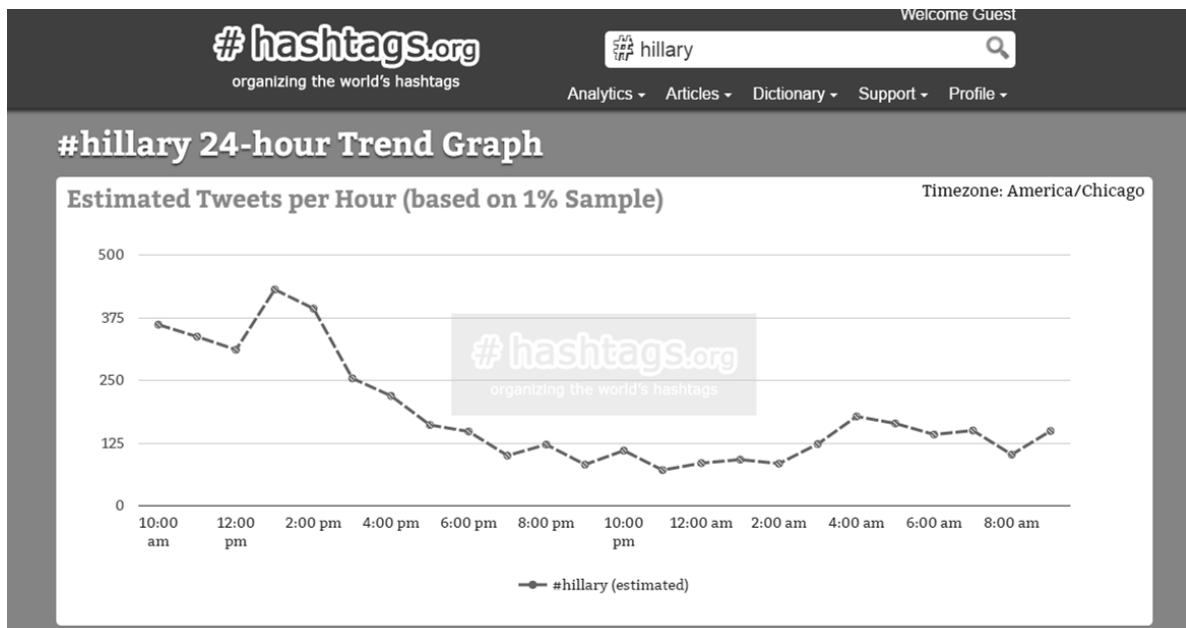


Figure 6: Hillary hour trend graph

Preparatory investigation of hashtags utilization in a Twitter information accumulation, acquired by an arrangement of seek questions demonstrates that 86% of novel hashtags are utilized under 5 times inside 3,209,281 tweets with hashtags. #jobs, #nowplaying, #zodiacfacts, and #news and #fb are the widespread hashtags and it is present in 8% of all tweets with hashtags. There are however there are some hashtags which are used seriously while the greater part of alternate hashtags are used meagerly.

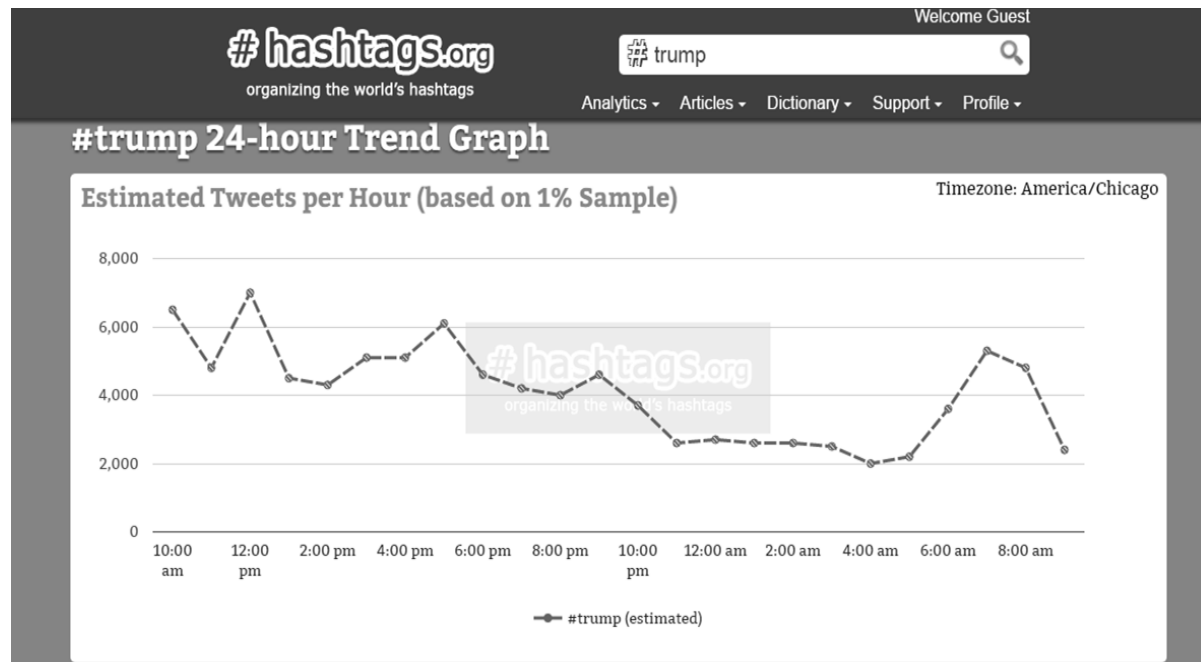


Figure 7: Trump hour trend graph

3.5 Creating a Twitter App

- Firstly we have to go to twitter's developer site and there is "My Applications" which will come up when we click on the top-right portion of the screen where our profile picture is.
- After that we have to click "Create new app" and fill in the details required.
- And then we go to that particular section in our app that has the API Keys, scroll down and click "Create my access token". Until we get the required Access token data on screen we have the keep on refreshing the page.

Project Module app

Test OAuth

Details Settings **Keys and Access Tokens** Permissions

Application Settings

Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.

Consumer Key (API Key)	FYQ5Avl03dlkCnd5NuiAzTSp4
Consumer Secret (API Secret)	XPcf2yFBRPFrsX1xhjYWMrvlg1aOC3yVlzABylmmJa9PtcD1oux
Access Level	Read, write, and direct messages (modify app permissions)
Owner	tashey95
Owner ID	1560949351

Application Actions

Regenerate Consumer Key and Secret Change App Permissions

Figure 8: Consumer Key and Consumer Secret

Your Access Token

This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.

Access Token	1560949351-Mxr7rcFsF98Tjnj4qSiQaWn25JWFRMRn32XZJ16
Access Token Secret	PYbBpnUtedopiAOm7TFo5oKrZrgdOEzpPf9zrG3BG7Dtj
Access Level	Read, write, and direct messages
Owner	tashey95
Owner ID	1560949351

Token Actions

Regenerate My Access Token and Token Secret Revoke Token Access

Figure 9: Access token and Access Token Secret

- After that in eclipse we will create a Java project and furthermore create a file named twitter4j.properties inside that venture.
- Finally we need to assimilate the data given below in a similar configuration in our code.

```
.setOauth.consumerKey=< consumer key of your app>  
.setOauth.consumerSecret=< consumer secret of your app >  
.setOoauth.accessToken=< Access token>  
.setOauth.accessTokenSecret=<Access token secret >
```

3.6 Setting up Twitter4J

- The initial step is in that Eclipse extend we just made, a “lib” folder is created.
- Then latest stable version of Twitter4J is downloaded.
- After that we extract files, click on the libs folder, ”twitter4j-core-xxx.jar” jar file is copied into the lib folder of our project
- Then in Java Build Path we have to add this jar file (Steps - We have to right click - project , Properties , Java Build Path ,Libraries tab , Add Jars)

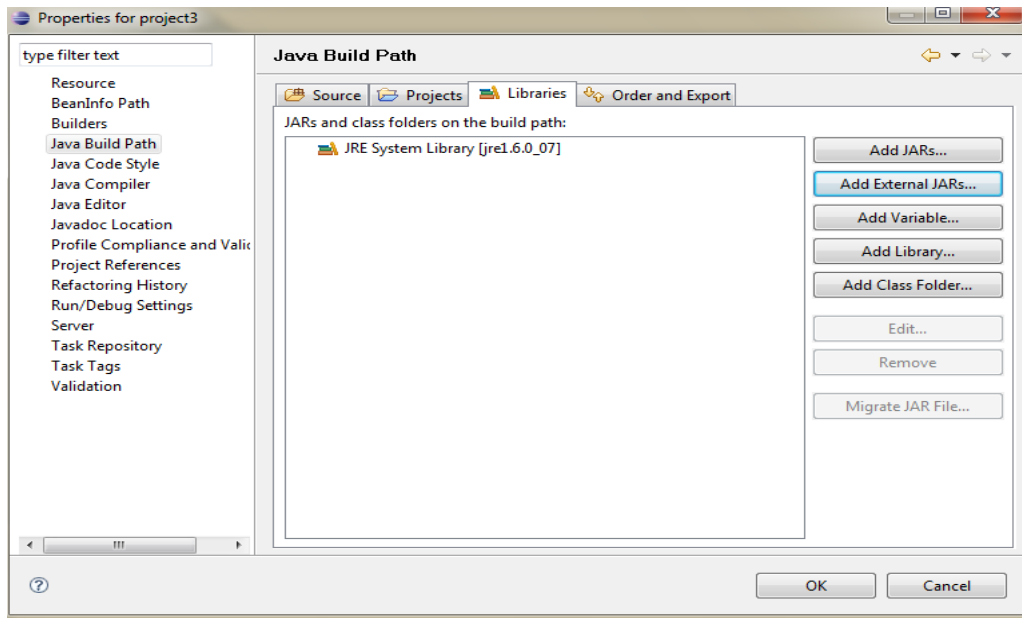


Figure 10: Adding External Jars

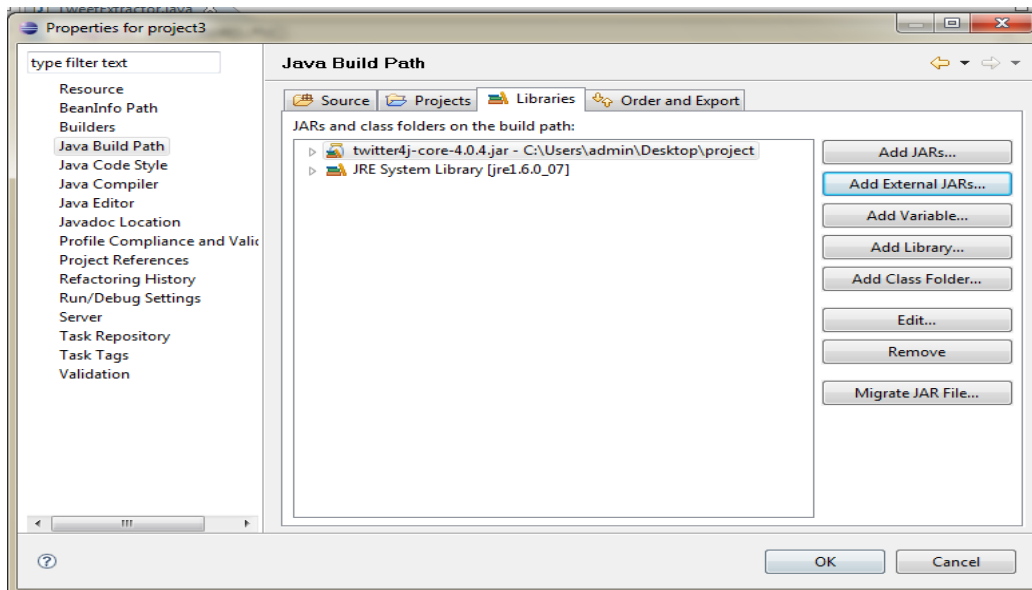


Figure 11: External Jars Added

- Now we are ready to pull out some tweets

3.7 Sentiment Analysis

Sentiment Analysis is the process of which finds whether a word or sentence or a document is positive, negative or neutral. SA is also known as opinion mining. This innovation is exploited to decide how individuals feel about a topic like election, film or a book and so forth. The applications for sentiment analysis are endless. It can be used to track customer reviews, survey responses, competitors, etc.

The need for sentiment analysis is increasing rapidly because SA is efficient. To extract sentiments and opinion, sentiment analysis evaluates thousands of text documents within few seconds when equated to groups of people who take hours of time to manually analyze each word or sentence. Text and sentiment analysis is embraced by many business and tries to use it into their processes because of its efficiency.

Sentiment analysis of social media can be an impressive foundation of information and can deliver insights that can:

- Define “marketing strategy”
- Enhance “campaign success”
- Enhance “product messaging”
- Enhance “customer service”

In a nutshell, sentiment analysis of social media can improve your company if done properly. However an inaccurate sentiment analysis data while deciding can give unsatisfactory result.

Sentiment Analysis have three classification stages:

1. Document level sentiment analysis.
2. Sentence level sentiment analysis.
3. Aspect level sentiment analysis.

In document-level sentiment analysis it classifies a document as containing a positive or negative sentiment. It considers the whole document as an elementary information unit. Each sentence is classified into “positive” and “negative” in sentence-level sentiment analysis by the sentiments conveyed by it. In SA the very initial phase is to discover whether the sentence is “subjective” or “objective”. On the off chance that the sentence is subjective, sentence-level sentiment analysis will conclude whether the sentence delivers a positive or negative opinions Wilson et al. said that the sentiment expressions are not necessarily subjective in nature. There is no distinct categorization between document and sentence level categorizations as sentences make up a document. Finally we have aspect-level sentiment analysis in which it considers particular features of objects to sort the sentiments. The first step is to identify the entities and their aspects.

Lexical -based Approach

Lexicon approach is a method in which each word in a sentence is given a sentiment polarity. This approach works on a bag of words. It has three sentiment scores which unfolds how positive, negative and neutral the words present in the synset are. The value of positive, negative and neutral score varies between ‘-1 to 1’, sum of the scores is 1 for each synset.

SentiWordNet word values have been calculated by employing the semi-supervised strategy portrayed in cases of sentiment scores related to SentiWordNet sections as appeared in the given figure beneath.

```

a    03154649  0.1250  tittering#1 thoriated#1  being or
sounding of nervous or suppressed laughter
a    03154786  0.1250  transpiring#1  that is passing
through; "transpiring gas"
a    03154886  0  0  sought#1  that is looked for; "the
long sought relatives"
a    03154986  0  0  closed-captioned#1  broadcast with
captions that are seen only on receivers having special
equipment; can be provided for hard-of-hearing viewers
a    03155193  0  0  saponified#1  converted into soap;
"saponified oils"
a    03155306  0.25  0.375  unsaponified#1  not converted into
soap; "unsaponified fat"
n    00001740  0  0  entity#1  that which is perceived or
known or inferred to have its own distinct existence (living or
nonliving)
n    00001930  0  0  physical_entity#1  an entity that
has physical existence
n    00002137  0  0  abstraction#6 abstract_entity#1  a
general concept formed by extracting common features from
specific examples
n    00002452  0  0  thing#12  a separate and self-
contained entity
n    00002684  0  0  physical_object#1 object#1  a tangible
and visible entity; an entity that can cast a shadow; "it was
full of rackets, balls and other objects"
n    00003553  0  0  whole#2 unit#6  an assemblage of
parts that is regarded as a single entity; "how big is that part
compared to the whole?"; "the team is a unit"

```

Figure 12: SentiWordNet

The records contain the parts of speech (POS) type of each word presented i.e whether it is an adjective - 'a', noun - 'n', verb - 'v' or an adverb - 'r'. Each word is given an id and some synonyms of the word share the common id. Then the entries' positivity score, negativity score, the list of synonyms for the entry and a brief description of the words.

The following image shows the various synset associated to the words “good” and “bad”. There are 4 values of the noun “good” 21 values of the adjective “good” and 2 values of the adverb “good” in WordNet. There is one value of the noun “bad”, 14 values of the adjective “bad”, and 2 values of the adverb “bad” in WordNet.

Category	WNT Number	pos	neg	Synonyms
A	14283456	0.815	0	good#1
A	01237812	0	0	good#1 full#4
A	02134741	0	0.615	bad#1
A	03401124	0.25	0.25	big#3 bad#1
N	03617701	0	0	trade_good#1 good#2 commodity#1

Figure 13: SentiWordNet fragment

After the tweets are pulled out and cleaned, sentence tokenization is done. In tokenization every sentence is separated into various components each representing an instantaneous tokens and afterward this token is sent to a part of speech tagger. It is also recognized as lexical classifications or word classes. The process in which we categorize words into their parts of speech and to tag them consequently is known as “part-of-speech tagging”, “POS-tagging,” or simply “tagging”. In Part of Speech (POS) tagging the tagger associates “the significance and role of each word in syntactic setting”. The tagger is useful once we obtain sentiments because it tells the classifiers the doable elements of a content where the sentiments might exists. Natural Language Tool-Kit (NLTK) is employed to train various a part of speech (POS) taggers.

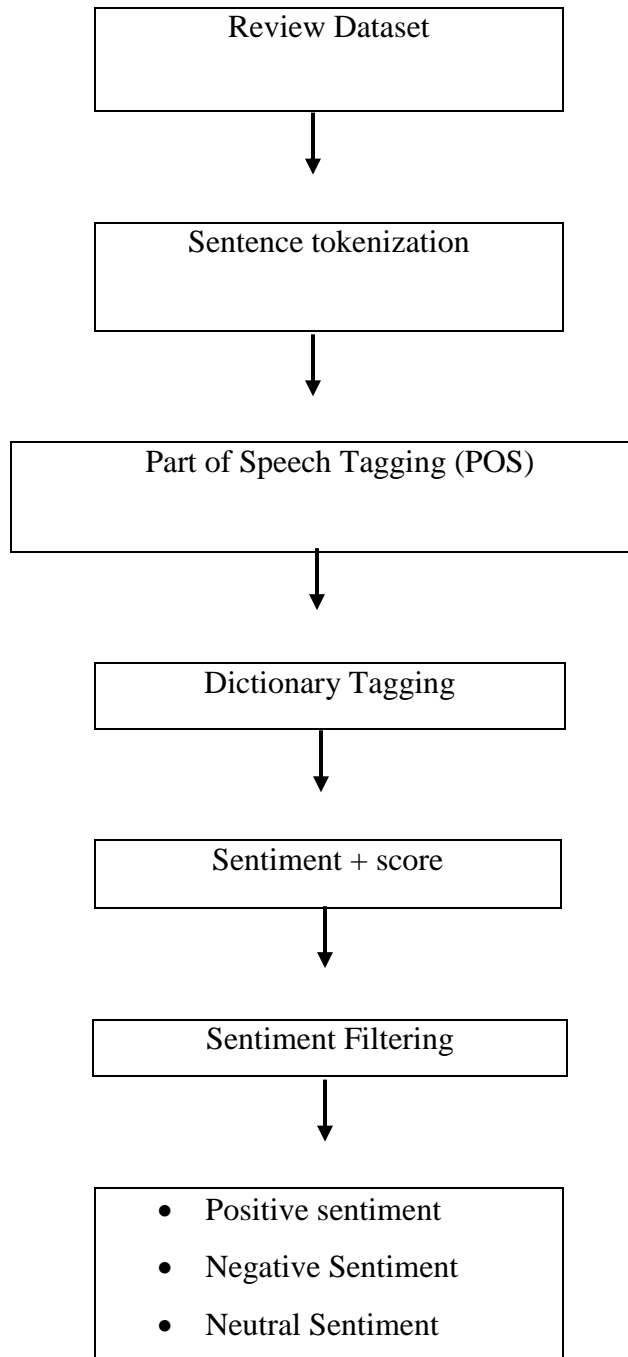


Figure 2: Lexical orientation- Dictionary based approach

POS tagging is followed by the dictionary tagging in which it matches words with dictionary objects and if any matching conditions become true then the word is given a sentiment score (which is “positive score – negative score”). Dictionary tagging is a crucial step as it contains words carrying sentiments which will be matched with the dataset. The dictionary has “the word”, “an opinion score” and “a synset score”. This process is repeated for every sentences in the whole document. Polarity score is number arranging between ‘-1’ to ‘1’ .If a word shows strong sentiment and is very intense then as a result a number nearby to 1 will be allotted to that word and if a word is less intense having weak sentiment then accordingly a number near to -1 is allotted to that. Sentiment score “0” signifies that word is neutral that it doesn’t have any sentiment.

After every sentence a sentiment score is assigned which is used to sieve these sentences into four unique classes as Positive, Negative, Neutral and Spam. If the sentences have either values greater than ‘-1’ or ‘1’, then these sentences are considered spam and the main purpose to unfold these spam is to intensely support or intensely oppose any issue and their motive is to unfold rumor at intervals the social media. We computed the sentiment for every tweet, initially the tweets were tokenized and every tokenized word is matched with dictionary words and then an appropriate sentiment score is given to the tokens (from dictionary). We summed the score for every tweet to search for (abundant|what proportion|what quantity) positive and negative tweets are there for every candidates collaborating within the election. To exemplify sentiments in an effective way we calculated percentage polarity.

$$\begin{aligned} \text{Positive sentiment\%} &= \frac{\text{Number of positive tweets}}{\text{Total no. of tweets} - \text{number of spam tweets}} * 100\% \\ \text{Negative sentiment\%} &= \frac{\text{Number of negative tweets}}{\text{Total no. of tweets} - \text{number of spam tweets}} * 100\% \\ \text{Neutral sentiment\%} &= \frac{\text{Number of neutral tweets}}{\text{Total no. of tweets} - \text{number of spam tweet}} * 100\% \end{aligned}$$

Figure 14: Percentage polarity

3.8 Algorithm

Initialize a variable say line as a string type.

Split the given string and store into an array say words for w in words

Split w wrt #

// Example line:

// POS ID PosS NegS SynsetTerm#sensenum Desc

// a 00009618 0.5 0.25 spartan#4 austere#3 ascetical#2

// ascetic#2 practicing great self-denial;...etc

Add the obtained string to a vector say v

Initialize a variable of double type say score = 0.0

Initialize a variable of double type say sum = 0.0

// Calculate weighted average.

//Weigh the synsets according to their rank.

// Score= $1/2 * \text{first} + 1/3 * \text{second} + 1/4 * \text{third} \dots$ etc.

// Sum = $1/1 + 1/2 + 1/3 \dots$

For i from 0 to size of vector v

 score = score + v.get(i)/i+1

For i from 1 to size of the vector v

 sum = sum + 1/i

Calculate score/sum as store the result in score

Initialize a variable of string type say sent

If score is greater than or equal to 0.75

 Then set sent as strong_positive

Else if score is greater than 0.25 and less than or equal to 0.5

Then set sent as positive

Else if score is greater than 0 and less than equal to 0.25

Then set sent as weak_positive

Else if score is less than 0 and greater than or equal to -0.25

Then set sent as weak_negative

Else if score is less than -0.25 and greater than or equal to -0.5

Then set sent as negative

Else if score is less than or equal to -0.75

Then set sent as strong_negative

Then initialize a dictionary as push the obtained word and score as the key value pair into that dictionary.

Catch the exceptions if any

Initialize a new variable of double type say total

Extract the word from the dictionary one by one

If dict.get(word+"#n") is not equal to null

Then set total = total + dict.get(word+"#n")

If dict.get(word+"#a") is not equal to null

Then set total = total + dict.get(word+"#a")

If dict.get(word+"#r") is not equal to null

Then set total = total + dict.get(word+"#r")

If dict.get(word+"#v") is not equal to null

Then set total = total + dict.get(word+"#v")

Return the obtained value of total.

CHAPTER 4 – PERFORMANCE ANALYSIS

Our project is on “sentiment analysis of political tweets”. The data required in our project is US election tweets, so we extract filtered data from Twitter, i.e. data that corresponds to a US presidential election which will be the system input. Then we did sentiment analysis on these extracted tweets using lexical approach.

4.1 Sun validation certificate

Sometimes we may get an error 'unable to find valid certification path to requested target' when functioning on a client that works with an SSL allowed server in https protocol, if the server certificate is not issued by certification authority(CA), but a self-signed or issued by a private CMS.

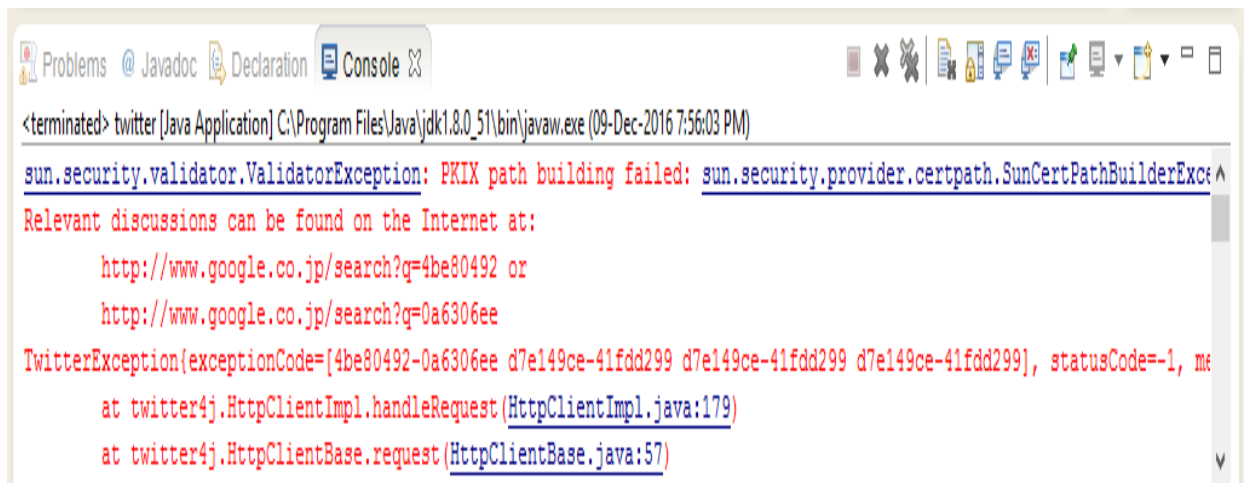
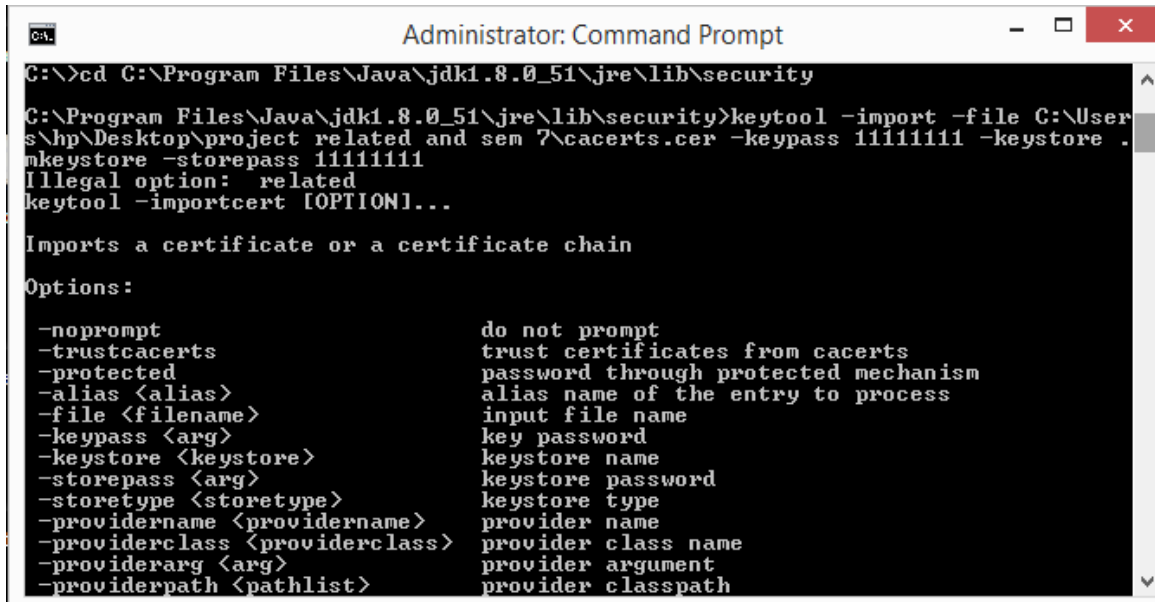


Figure 15: Sun Security Validation Error on Execution

There is a file called cacerts located in Program Files/Java/jdk/jre/lib/security which we have to copy to desktop. Then we clicked on the link(http://) of the twitter and clicked on “view certificate”.

After we copy the cacerts on the desktop it's replaced with the new exported certificate. Now that the server certificate is installed on the computer, the browser doesn't warn us anymore when we visit the same site again.

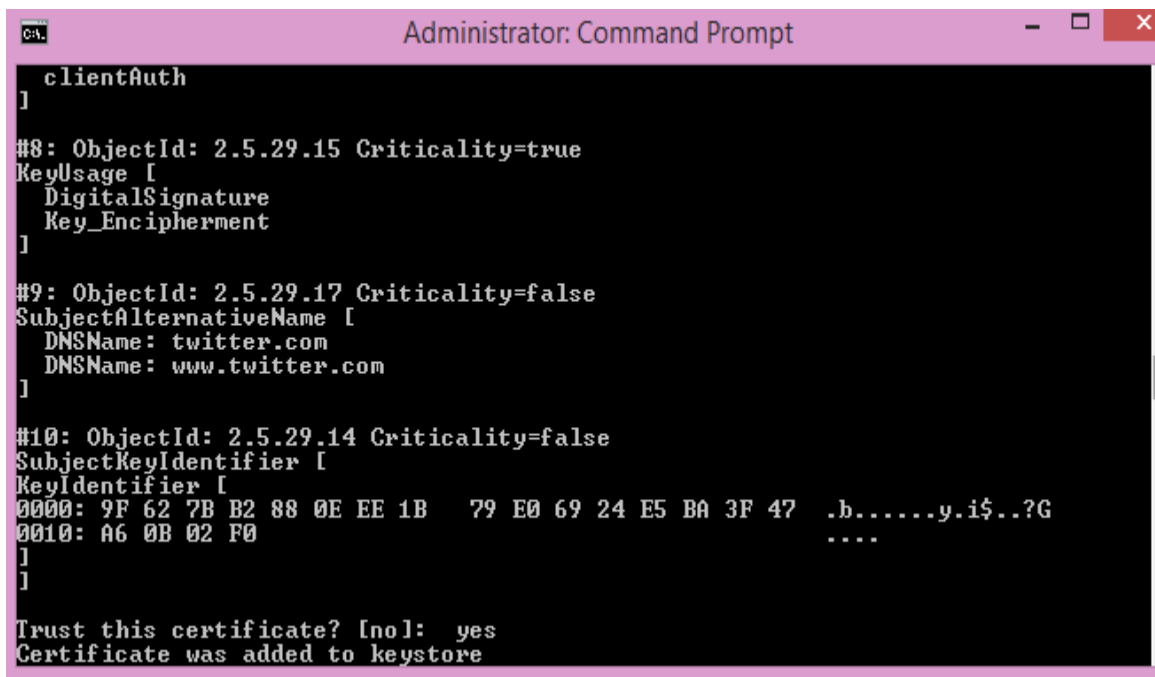


```
Administrator: Command Prompt
C:\>cd C:\Program Files\Java\jdk1.8.0_51\jre\lib\security
C:\Program Files\Java\jdk1.8.0_51\jre\lib\security>keytool -import -file C:\User
s\hp\Desktop\project related and sem 7\cacerts.cer -keypass 11111111 -keystore .
\mkeystore -storepass 11111111
Illegal option: related
keytool -importcert [OPTION]...

Imports a certificate or a certificate chain

Options:
-noprompt                do not prompt
-trustcacerts            trust certificates from cacerts
-protected               password through protected mechanism
-alias <alias>          alias name of the entry to process
-file <filename>        input file name
-keypass <arg>          key password
-keystore <keystore>    keystore name
-storepass <arg>        keystore password
-storetype <storetype>  keystore type
-providername <providername> provider name
-providerclass <providerclass> provider class name
-providerarg <arg>      provider argument
-providerpath <pathlist> provider classpath
```

Figure 16: Importing the Certificate



```
Administrator: Command Prompt
clientAuth
]
#8: ObjectId: 2.5.29.15 Criticality=true
KeyUsage [
  DigitalSignature
  Key_Encipherment
]
#9: ObjectId: 2.5.29.17 Criticality=false
SubjectAlternativeName [
  DNSName: twitter.com
  DNSName: www.twitter.com
]
#10: ObjectId: 2.5.29.14 Criticality=false
SubjectKeyIdentifier [
KeyIdentifier [
0000: 9F 62 7B B2 88 0E EE 1B 79 E0 69 24 E5 BA 3F 47 .b.....y.i$..?G
0010: A6 0B 02 F0 .....
]
]
Trust this certificate? [no]: yes
Certificate was added to keystore
```

Figure 17: Certificate Imported Successful

4.2 Extracting Tweets

After we have successfully added the certificate to the key store, we did not get the sun validation certificate exception anymore. Rather the code was compiled without any errors.

Then after the code was executed, we got the output as shown in the figure below:

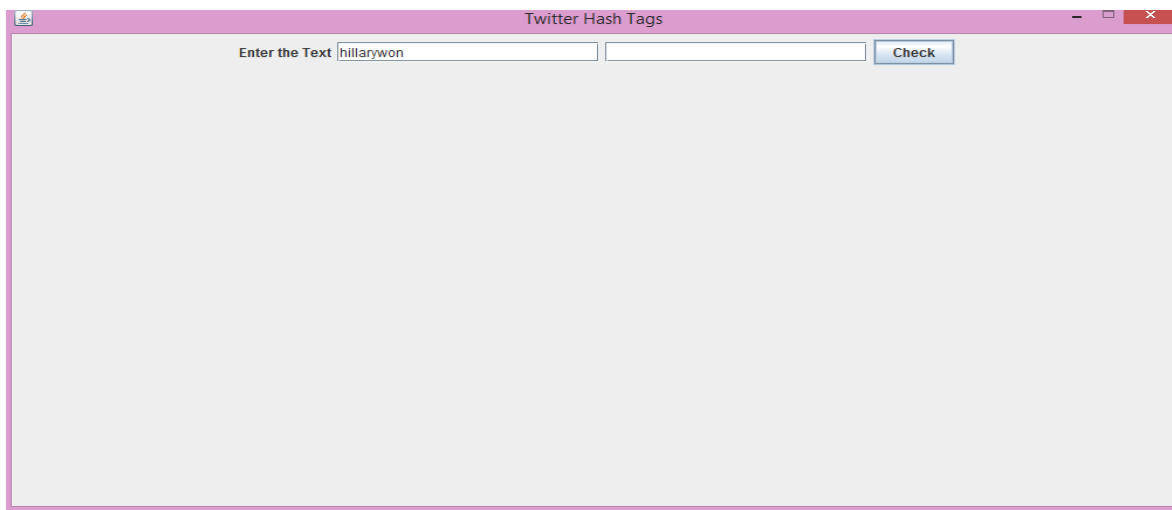


Figure 18: Output with Hashtag

In the output displayed, we have entered the hash tag. It took some time to authenticate and after the authentication the tweets related to the hash tags were generated along with the user name and the date when the tweets were tweeted.

SN	User	MSG	Date	Score
0	JJohnson2u	RT @dawnsto...	Mon Feb 01 1...	1
1	JJohnson2u	RT @dawnsto...	Mon Feb 01 1...	1
2	LisaHauser8	RT @OfficialN...	Sun May 29 0...	1
3	smiley1085	RT @OfficialN...	Wed Jul 02 05...	1
4	OfficialNMP	awww whaaa...	Mon Feb 29 1...	1
5	FuQDonaldTr...	RT @silva198...	Tue Dec 30 0...	1
6	gpg408	RT @cit92: @...	Sun Apr 17 08...	1
7	Nikki206	Yay @Kellogg...	Wed Mar 11 0...	1
8	JJohnson2u	RT @dawnsto...	Mon Feb 01 1...	1
9	levylieu	#crookedTru...	Thu Jan 22 08...	1
10	solleytolley	@solleytolley t...	Mon Oct 10 06...	1
11	lbutlr	RT @myownp...	Sat Jun 09 06...	1
12	JJohnson2u	RT @dawnsto...	Mon Feb 01 1...	2
13	dawnstottlar	For real @m...	Fri Oct 11 01...	2
14	myownpetard	I'm not shoppi...	Sun Jul 12 00...	1
15	NameGirl	RT @UckfayR...	Wed Aug 20 1...	1
16	rodone_keep...	RT @intlbeco...	Fri Nov 11 02...	1
17	designergrace	#dumptrump ...	Sun Jun 27 0...	2
18	BoycottHamil...	@Above_Kas...	Mon Nov 21 0...	1
19	FreedomNow...	@huitecoutur...	Wed Nov 16 1...	1
20	patts54	RT @PennDr...	Thu May 10 0...	1
21	cj_protzman	@chrishayes ...	Wed Nov 16 0...	1
22	FoikWestRoad	RT @stillgails...	Fri Jun 22 15...	1
23	PennDragonArt	#hamiltonelec...	Fri Apr 17 16...	1
24	tinaljacobs	@cherbp @L...	Sun Jan 03 1...	1

Figure 19: Output with Tweets Generated

4.3 Sentiment analysis using Lexical approach

In Lexical Based approach we add up the polarities (which is measured in weighted average form) of individual words or phrases and as a result it generates the polarity of the entire document. The Statistical orientation approach of Sentiment analysis is “unsupervised learning” because it does not require any preceding training to classify the data. Rather it computes whether a particular word is more probable to have positive and negative sentiment.

Pros: The advantage of this technique is that it can be applied to a text with no training data required (It can be an alternative where the training data doesn’t exist or when dealing with multiple domain).

Cons: On the other hand this technique consumes heavier sentence manipulation and requires lexicons which are already constructed beforehand.

We have gathered almost 20,000 tweets and did the sentiment analysis on this tweets. Each tweet is given a sentiment score ranging from ‘-1’ to ‘1’. If the sentence have score closer to 1 then it carries strong or positive sentiment and if the score is closer to -1 it carries negative sentiment and scores greater than ‘1’ and less than ‘-1’ are considered spam.

4.4 Result

1	UserName	message	Score
2	LawsonDavidn	RT @LPNational: Hillary Clinton and Tim Kaine will refuse to let American's invest in their own retirement #VPDebate #VPDebate2016 #Debates?	-0.12232
3	thenelipas	RT @true_pundit: Hillary Clinton on her proposed Drone Strike on Julian Assange: "I don't recall" @wikileaks #TruePundit #OctoberSurprise #?	0.096633
4	clark7950	RT @LisaMThornburg: Facebook is already trying to block this. Retweet ASAP. #Guccifer2.0 Hacked Clinton Foundation https://t.co/z2FA0X9WP3	-0.37714
5	hassan_k82	RT @timkaine: Hillary Clinton's life has always been about putting others first?Donald Trump's always been about putting himself first.	0.182399
6	juliettebodo	I can see why Crooked Hillary Clinton picked Kaine for VP Pick their 2 of a kind.Pence is more of a man for VP. https://t.co/sCycX8mp4n	-0.01899
7	LindaSuhler	RT @NBCNews: Pence: People don't trust Hillary Clinton "because they're paying attention." #VPDebate #Decision2016 https://t.co/LuZRxP9vFY	0.566952
8	margaritateresa	RT @PoliticosSarah: Pence says Clinton ran the country into the ground Kaine smiles and talks about historic job growth. #VPDebate	0.093546
9	aragonac	RT @MSNBC: Pence: People don't trust Hillary Clinton "because they're paying attention." #VPDebate #Decision2016 https://t.co/iSnKhtYtQ0g	0.566952
10	Len21878832	RT @TeamTrump: There's no doubt Obama-Clinton's leadership has WEAKENED America's stance on the world stage. #TrumpPence16 will put #Amer	-0.28405
11	kkarop	RT @TeamTrump: "The Clinton Foundation accepted foreign contributions from foreign donors and foreign govts WHILE SHE WAS SECRETARY OF STAT?	-0.10219
12	RhondaWimbish	RT @OurVoice911: mike_pence: Hillary Clinton's economic plan would send our economy into a tailspin. #BigLeagueTruth https://t.co/VtyU4wxxDI	-0.18811
13	BeckCapital	@realDonaldTrump paid millions in payroll tax and property tax. @HillaryClinton deducted almost \$200k in donations to the Clinton Foundation	0.356918
14	marktsword	RT @TeamTrump: #VPDebate: Over Half Of Clinton's Non-Gov Meetings At State Were With Foundation Donors https://t.co/RNfRdckFaV	0
15	jakobk9	everytime pence tries to talk Kaine just speaks over him and they let it happen. this election is so biased towards the Clinton campaign.	0.842594
16	user239956	@mike_pence "Your running with Hillary Clinton who wants to RAISE taxes!" #TrumpPence16 #maga ??????????????	0.023866
17	neerajamb	RT @timkaine: Hillary Clinton's life has always been about putting others first?Donald Trump's always been about putting himself first.	0.182399
18	stmichael36	RT @voxdotcom: Donald Trump's tax plan makes impossible promises. This cartoon explains. https://t.co/ofvUgXhNj6 #VPDebate	-0.18182
19	FortuneMagazine	Clinton Foundation denies its server was hacked by Guccifer https://t.co/a2l20l7o9m https://t.co/67Dx1BiWlj	-0.09091
20	MatpieWan	A Bill Clinton gaffe on health care and Trump pounces: WASHINGTON: Bill Clinton created fresh problems for h... https://t.co/LLxRI1U1Vj	0.363186
21	jaycamikenGL	Kaine is so obnoxious - all he does is interrupt and spin the same old Clinton lines that deflect divert & deceive... #VPDebate	-0.23751
22	GreillyPost	RT @SopanDeb: It should be noted - the withdrawal of troops that Pence is criticizing Obama/Clinton for is the same one that his own runnin?	0.225424
23	RobertJ74458844	@IngrahamAngle Sorry- Kaine got the questions before the debate. Clinton also got the questions. This is a scam.	-0.22799
24	RobertJ74458844	Why is the moderator not stopping Kaine from interrupting? @realDonaldTrump's VP Clinton brought questions?	0.60077

Figure 20: Output using Lexical approach

After calculating the sentiments of each tweets, we are going to find the number of positive, negative, neutral and spam tweets for each candidate participating in US elections.

	Total Tweets	Positive	Negative	Neutral	Spam
Hillary	9300	4178	3477	1207	438
Trump	9300	4353	3222	1324	401

Table 1: Total number of tweets (positive, negative, neutral and spam) for each candidate

$$\text{Positive sentiment\%} = \frac{\text{Number of positive tweets}}{\text{Total no. of tweets} - \text{number of spam tweets}} * 100\%$$

$$\text{Negative sentiment\%} = \frac{\text{Number of negative tweets}}{\text{Total no. of tweets} - \text{number of spam tweets}} * 100\%$$

$$\text{Neutral sentiment\%} = \frac{\text{Number of neutral tweets}}{\text{Total no. of tweets} - \text{number of spam tweets}} * 100\%$$

From these we calculated the output of the lexical approach in terms of percentage.

Approach	Candidates	Positive %	Negative %	Neutral %
Lexical Approach	Hillary	47.15	39.23	13.62
	Trump	48.91	36.21	14.88

Table 2: Percentage of tweets (positive, negative and neutral) for each candidate.

This table shows that by applying sentiment analysis on a file having 9300 tweets for each candidate (Hillary Clinton and Donald Trump) using lexical approach to calculate percentage involvement of each type we got 47.15% positive tweets, 39.23% negative tweets, 13.62% neutral tweets for Hillary. And we got 48.91% positive tweets, 36.21% negative tweets, 14.88% neutral tweets for Trump.

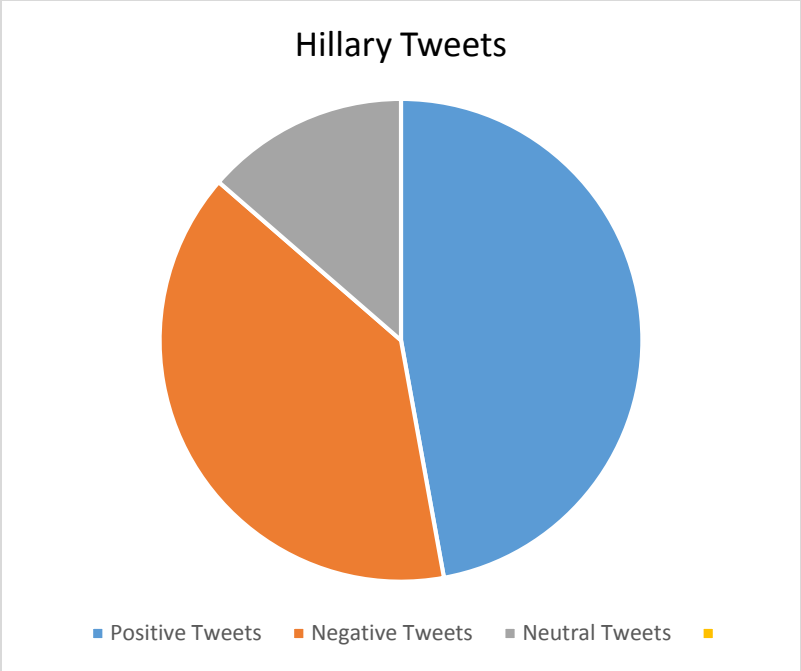


Figure 21: Pie chart showing Hillary’s positive, negative and neutral tweets

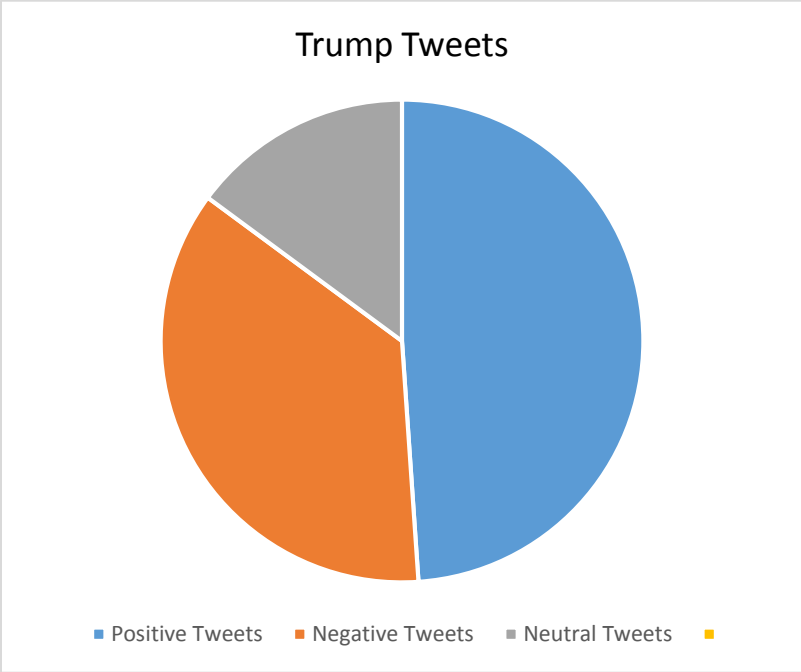


Figure 22: Pie chart showing Trump’s positive, negative and neutral tweets

CHAPTER 5 - CONCLUSIONS

5.1 Conclusions

Twitter is a micro blogging facility that has more than 500 million messages on a daily basis. Scholars have been utilizing Twitter to monitor people reactions in political activities, such as debates and campaigns. By doing so, some of them claim that a forecast or prediction to an election can be made. We have built a model to sentiment analysis of political tweets collected from Twitter. We implemented the model on the presidential election in United States and collected data based on the candidates competing for each other. Although there were many limitations of the data, we believe that this data, if analyzed thoroughly, can potentially tell us a lot, and importantly can empirically test some of the assumptions made when building the model.

We chose this project because sentiment analysis is in demand these days due to its efficiency. It evaluates huge amount of text documents for sentiment and other features within few seconds when equated to groups of people who take hours of time to manually analyze each word or sentence. Text and sentiment analysis is embraced by many business and tries to use it into their processes because of its efficiency. There are different techniques to do analysis but we chose lexical approach which provides sentiment polarity to every word in a sentence. The advantage of this technique was that it can be applied to a text with no training data required (It can be an alternative where the training data doesn't exist or when dealing with multiple domain).

On the other hand this technique consumes heavier sentence manipulation and requires lexicons which are already constructed beforehand. More over in this technique , there is a step called dictionary tagging where every word is matched with dictionary object and given a score, but if the word is not present in the dictionary , it gives a score zero even when there is chances that the sentence might carry some sentiments.

However we got the result of the sentiment analysis done on almost 20,000 tweets related to US presidential election that we have collected from twitter. We found out that Hillary have 47.15% of positive tweets while Trump have 48.91% positive tweets and Hillary have 39.23% negative tweets while Trump have 36.21% of negative tweets, so we can conclude that though the difference between the two candidate is by few percentile , there is chances that Trump might win the election.

5.2 Future Scope

Sentiment analysis is an area which is evolving rapidly from a simple classification of positive, negative and neutral to more deep learning and understanding. As the future work, we plan to add more words in the dictionary to reduce the number of sentences which are given zero sentiment score even though when they convey some opinion, due to lack of matching words in the dictionary. We also plan to adapt our sentiment analysis system to languages other than English so that people can do analysis on the dataset written in their own language. We will also try to use an early spam detection technique to make the spam filtering method more effective, so that we will get only valid information.

References

- [1] Tumasjan, Andranik, et al. "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment." ICWSM 10 (2010): 178-185.
- [2] Vikash Nandi and Suyash Agrawal. "Political Sentiment Analysis using Hybrid Approach",2016
- [3] Gayo Avello, Daniel, Panagiotis T. Metaxas, and Eni Mustafaraj. "Limits of electoral predictions using twitter." Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. Association for the Advancement of Artificial Intelligence, 2011.
- [4] Hofmann, Thomas. "Probabilistic latent semantic indexing." Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 1999.
- [5] Roman Dovgopool and Matt Nohelty,"Twitter Hash tag Recommendation" in *University of Minnesota*.
- [6] Daniel Preot, iuc-Pietro and Trevor Cohn(et al.)," A temporal model of text periodicities using Gaussian Processes" in *University of Sheffield in Portobello Street Sheffield*.
- [7] Maximilian Jenders, Gjergji Kasneci and Felix Naumann(et al.) "Analyzing and Predicting Viral Tweets" in *Hasso Plattner Institute in Potsdam, Germany*.
- [8] Wei Feng and Jianyong Wang ," Connecting Tweets to Explicit Topics" in *Tsinghua University in Beijing, China*.
- [9] Romain Deveaud and Florian Boudin," Effective Tweet Contextualization with Hashtags Performance Prediction and Multi-Document Summarization" in *University of Avignon*.
- [10] Frédéric Godin (et al.), "Using Topic Models for Twitter Hashtag Recommendation" in *Ghent University, Ghent, Belgium*.
- [11] Oren Tsur and Ari Rappoport, "Content based Prediction of the Spread of Ideas in Microblogging Communities" in *The Hebrew University*.
- [12] Su Mon Kywe (et al.), "On Recommending Hashtags in Twitter Networks" in *Singapore Management University*.

- [13] Shoubin Kong (et al.), “On the Real-time Prediction Problems of Bursting Hashtags in Twitter”.
- [14] Shoubin Kong(et al.) , “Predicting Bursts and Popularity of Hashtags in Real-Time” in *Tsinghua University*.
- [15] Siddharth Bora(et al.), “On the Role of Conductance, Geography and Topology in Predicting Hashtag Virality” in *Indian Institute of Technology, Delhi*.
- [16] Jason Weston(et al.) ,” Semantic Embeddings from Hashtags” .
- [17] Krishna Y. Kamath and James Caverlee ,” Spatio-Temporal Meme Prediction:Learning What Hashtags Will Be Popular Where” in *Texas A&M University*.
- [18] Allie Mazzia and James Juett, “Suggesting Hashtags on Twitter” in *University of Michigan*.
- [19] Bo Han (et al.), “Text-Based Twitter User Geolocation Prediction” in *The University of Melbourne*.
- [20] Lisa Posch(et al.),” Meaning as Collective Use: Predicting Semantic Hashtag Categories on Twitter” in *Graz University of Technology*.
- [21] Zongyang Ma (et al.),” On Predicting the Popularity of Newly Emerging Hashtags in Twitter” in *Nanyang Technological University*.
- [22] Tuan Anh Hoang (et al.) , “On Recommending Hashtags in Twitter Networks” in *Singapore Management University*.
- [23] Florian Kunneman(et al.), “The (Un)Predictability of Emotional Hashtags in Twitter” in *Radboud University Nijmegen*.
- [24] Qi Zhang (et al.) ,” Time-aware Personalized Hashtag Recommendation on Social Media” in *Fudan University*.
- [25] Peng Zhang (et al.),” On Predicting Twitter Trend: Important Factors and Models” in *Arizona State University*.
- [26] Tianxi Li and Yu Wu ,” Twitter Hash Tag Prediction Algorithm” in *Stanford University*.
- [27] Gaikar, Dipak Damodar, Bijith Marakarkandy, and Chandan Dasgupta. "Using Twitter data to predict the performance of Bollywood movies." *Industrial Management & Data Systems* 115.9 (2015): 1604-1621.