SENTIMENT ANALYSIS OF TWITTER TEXTUAL DATA USING MAP REDUCE ON DEMONETISATION OF MONEY IN INDIA

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

Computer Science and Engineering

by

Sulabh (131284)

Under the supervision of
Ravindara Bhatt

to

Department of Computer Science and Engineering and Information Technology

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

Certificate

Candidate's Declaration

I declare that work that has been done by me in this report titled **"Sentiment analysis of twitter textual data using map reduce on demonetisation of money in India"** in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering** submitted to the department of Computer Science and Engineering and Information Technology**,** Jaypee University of Information Technology is my very own real record for work done from August 2016 to May 2017 under the guidance of Ravindara Bhatt (Assistant Professor, CSE). All the data encapsulated in this report has not been used for any other degree.

Sulabh, 131284

I hereby certify that the above declaration made by the student is correct to the best of my knowledge.

Dr. Ravindara Bhatt
Assistant Professor
Department of CSE
Dated

Acknowledgement

"The successful achievement of any work would be incomplete without acknowledging the people whose constant guidance and encouragement secured us the success."
First of all, we are grateful to the Almighty for establishing us to complete this project.

We owe a debt of gratitude to our guide, **Dr. Ravindara Bhatt (**Assistant Professor**)** for suggesting us this challenging yet creative idea for Major Project. He helped me completing this project and readily provided his assistance whenever we needed it.

I also place on record, my sense of gratitude to one and all, who directly or indirectly have lent their helping hand in this venture.

We experience proud and privileged in expressing my deep feel of gratitude to all those who have helped me in presenting this project.

Table of Content

List of Abbreviations

Sr. No	Abbreviations	Term
1.	Application Programming Interface	API
2.	Hadoop Distributed File System	HDFS
3.	JavaScript Object Notation	JSON
4.	Social Network Service	SNS

List of figures

List of tables

Abstract

The fast advancement and also utilization of the second stage of development of the Internet and Online networking has led the users generating huge volumes of data. A lot of sites on the web offer users to express their opinions on various products, people and events. This leads to an opportunity for mining sentiments from large unstructured data. In this project we implemented a dictionary based algorithm (which uses a predefined dictionary instead of a classifier to determine whether a word is negative, positive or neutral) on map reduce framework that is capable of processing large amount of data. A large amount of tweets are fetched using “Twitter Developer API” to HDFS. These tweets are then pre-processed to extract relevant information and then are further analysed to determine the sentiment of tweet. The output is the time series visualisation of average sentiment about the given subject. Using this system on the subject demonetisation of money in India lead to the observation that most people have neutral sentiment towards the issue followed by positive and then negative.

Chapter-1
Introduction

1.1 Introduction

Current population of our globe is around seven billion and is exceeding rapidly. Thirty percent of this population are connected to internet. Additionally, five billion persons are in possession of some mobile device, conferring to McKinsey (2013). A consequence of this technical revolution is that this huge population of people are generating humongous amounts of data through the increased use of such devices. This huge data contains opinions or sentiments of people about a certain subject which might be of value for business and scientific works. Sentiment Analysis is called as the task to find the opinions and reviews of people about anything, products, people, events or new movie. At the time of this project duration Demonetisation of money in India was a hot topic about which a lot of people had different opinions. So this topic was chosen as the subject for the sentiment analysis program developed. A new form of blogging that is microblogging has arisen out to be the stage where people expresses their sentiments regarding certain topic, how they feel? Is 'good' or 'bad'. Of all such microblogging sites Twitter is one such platform, which offers such services imposing a word limit of 150 on each tweet. Being established in later 2000's this site has gained a huge user base of over a billion users. Almost all VIP's and politicians having impact on the culture and society have their accounts on Twitter. That's why Twitter was chosen for experimental data source for this work on predicting people emotions on Demonetisation of money in India.

1.2 Problem Statement

We propose an approach to analyse public sentiment on Demonetisation of money such that a clear, concise image of people can be known regarding new amendment and their sentiments or opinions regarding it.

1.2.1 Why Sentiment Analysis?

Everyday enormous amount of data is being created by microblogging & social networking sites. This huge data contains opinions or sentiments of people about a certain subject which might be of value for business and scientific works. This structured/unstructured data is of the order of magnitude of Zeta Bytes and can be used for various scientific and business purposes. Sentiment analysis is used to classify the text. It categories the sentiment into three types namely neutral, positive and negative, and hence reflects the opinion of author of the text.

1.3 Methodology

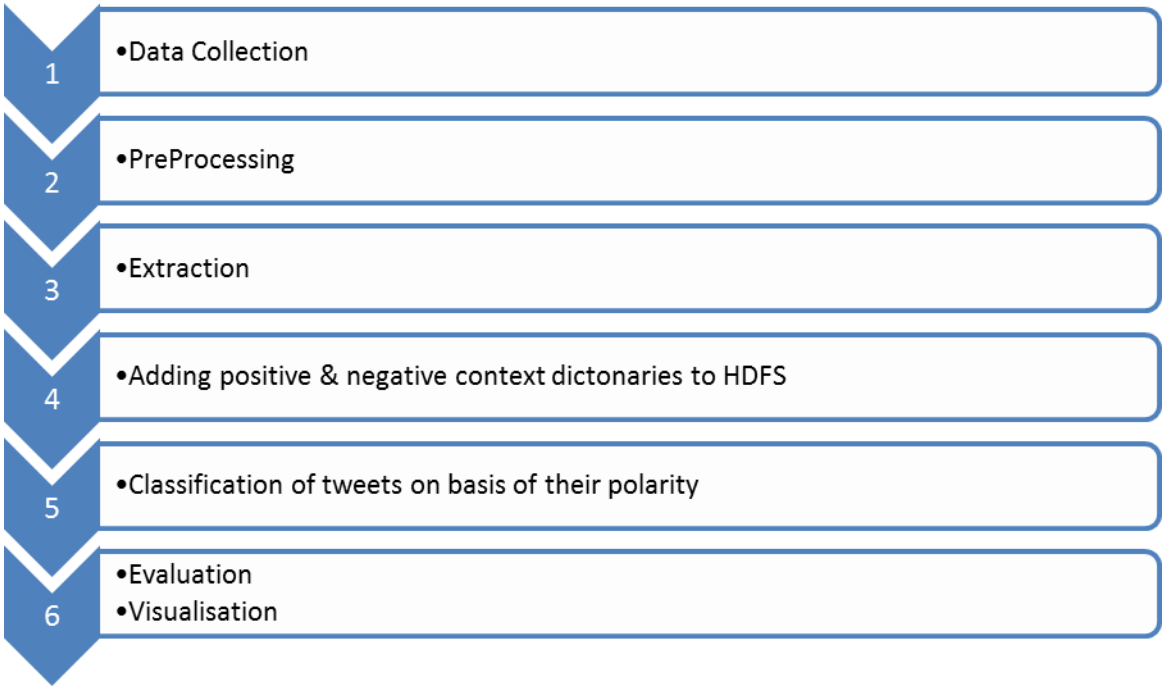


Figure 1: Depicts the flow diagram of the project.

Figure 1 depicting the flow diagram of the project. First three steps fetches the relevant data. Step four loads the necessary metadata and steps 5 and 6 focus on computing and visualising results.

1.4 **Organisation**

Chapter 1: In this chapter, the introduction to various concepts and techniques used in the implementation is covered.

Chapter 2: This includes reviewing relevant work from various research papers, books, journals and conferences. In this chapter, the extracts from assorted research papers on various situations are taken.

Chapter 3: Discusses about proposing a model suitable for sentiment analysis and implementing it. It is the key aspect of this work.

Chapter 4: Shows the simulation of implementation results with the relative performance analysis. In this chapter, the simulation results and screenshots are revealed to depict and defend the proposed work.

Chapter 5: This section concludes the whole work and also elaborates the scope of work that can be done in future which leaves an opportunity for upcoming students and scholars to further enhance this work.

Chapter-2

Literature Survey

Earlier sentiment analysis of large plain texts like articles, essays etc. was the topic of interest to the researchers but now sentiment analysis of small texts such as in micro-blogging has a become the topic of research and there is lot of scope for research in this field. As said earlier sentiment analysis of papers, blogs, articles, user reviews have seen a lot of research work but these differ from Twitter primarily because of the relatively short amount of text of about 150 char limit for each tweet. Machine learning methods such as Support Vector Machines (SVM's) and Naive Bayes have produced good results but require manual labelling and is therefore quite expensive. The work that has been done on some semi-supervised and unsupervised approaches have low accuracy. Various researchers testing new features and classification techniques often just compare their results to base-line performance. Presently there is also no formal and proper comparison and benchmarking technique for various methodologies for sentiment analysis to select the appropriate technique for various applications.

The most widely used model for sentiment classification tasks is the *bag-of-words model* because of its easiness and performance. The easiness is due to the fact that it ignores the grammar and how the words are connected to each other and views text as a collection of words. This is the reason for its popularity amongst various researchers. By using unigrams as feature you can use it in your classifier. "n-grams" means a consecutive sequence of n words independent of other words in the text. Therefore, unigrams means that each word is independent of each other. This seems like an oversimplification but it provides good results. One easy way to use unigram is to use a dictionary of pre assigned polarities and sum the polarities of all unigrams and take the average. Polarity of a word means whether a word is positive, negative or neutral. Polarities can be assigned as positive negative or neutral or we can also assign the degree of positivity negativity or neutrality. This allows a word like "good" would probably have weak positive polarity, and the word "fantastic" would strong positive polarity.

Prior polarity of unigrams can be used in mainly three ways. One way is to make usage of various available dictionaries online that map a word to its polarity. SentiWordNet and MPQA (Multi Perspective Question Answering) provides a lexicon which has a mapping of about five thousand words according to strong or weak or negative or positive subjectivity. A yet another way is to use your own dictionary constructed according to frequency in a specific class. Example if a particular word occurs frequently in the sentences labelled as positive in our data for training then the word probably belongs to a positive class rather than neutral or negative. Relatively better performance is achieved using this approach because the polarity of word is more appropriate and not general like in the previous approach. The second one is an example of supervised approach as manual labelling of training data has to be done for correct classes before calculating the sentiment of a sentence.

The last method is a central ground between the other two methods. Here in this method we create a new dictionary but not from training data as we want to avoid manual labelling. A method of achieving this proposed by Turney et al. [10,11] is to compute the polarity of a word by computing its mutual information with “poor” and take the difference of the result with the mutual information of that word with the “awesome”. To calculate the mutual information number of hits from search engines such as Google, Bing etc. of a relevant query. They use the following formula:

$$Polarity(phrase) = log_2 \frac{hits(phrase \text{ NEAR "excellent"}).hits(\text{"awesome"})}{hits(phrase \text{ NEAR "poor"}).hits(\text{"poor"})}$$

Here “hits” means number of results fetched by the search engine in the sentence whose polarity is to be computed and the word “awesome” is co-occurring. Prabowo et al. implemented this method and used one hundred twenty positive wor
ds and same amount of negative words to perform the searches on online search engine. To calculate the sentiment of the given word we calculate the nearness of that word to the seed word and take the overall average. Another method was proposed by Pak and Paroubek [12] which involves automatically collecting corpus for opinion analysis. That corpus can then be used to build classifier for classifying text as negative, positive and

neutral opinions. Their classifier was quite efficient but they didn't considered hardware for saving collecting training data. Bhattacharyya and Mukherjee [13] considered the possibility of detecting the sentiment of Twitter messages using linguistic features. They collected training data using hashtags. They studied the part of the hashtag and various elements of speech in analysing sentiment. But they did not said anything about the hardwarealso did not consider the hardware aspects related to data collection and processing.

Bhattacharyya and Mukherjee [13] used an efficient process for using discourse relations to determine the sentiment of a tweet. It integrates he bag of words model with the discourse information to increase accuracy. There are other works on analysing the sentiment. Some specifically work with opinion analysis on Twitter, for example, Go et al. [14] defined a distant supervision-based approach for classifying the opinion. They employed the use of hashtag to generate training set and a classifier to determine the sentiment. Barbosa and Feng [15] also suggested a process for analysing sentiment of tweets. They used POS-tagged hashtags and n-gram features.

Till present numerous works have been done on analysing sentiment using Hadoop. Khucet al. [16] proposed a scalable and distributed system for analysing sentiment on Hadoop. His proposed system consists of two subsystems an opinion classifier and a lexicon builder. Jeonghee Yi et al. [17], suggested a tool to extract sentiments regarding a particular subject from various online documents. It uses advanced NLP techniques. Basically what it does is it searches all the references on the subject and determine the polarity of each one. It uses an opinion lexicon and a pattern sentiment database for purpose of association. They analysed reviews from online shopping sites with nice results.

Chapter-3

System Development

3.1 HDFS (Hadoop Distributed File System)

HDFS is a distributed file system quite dissimilar from other file systems having a very unique set of features. HDFS provides great reliability and fault tolerance and is designed to operate at ordinary low cost PC's.

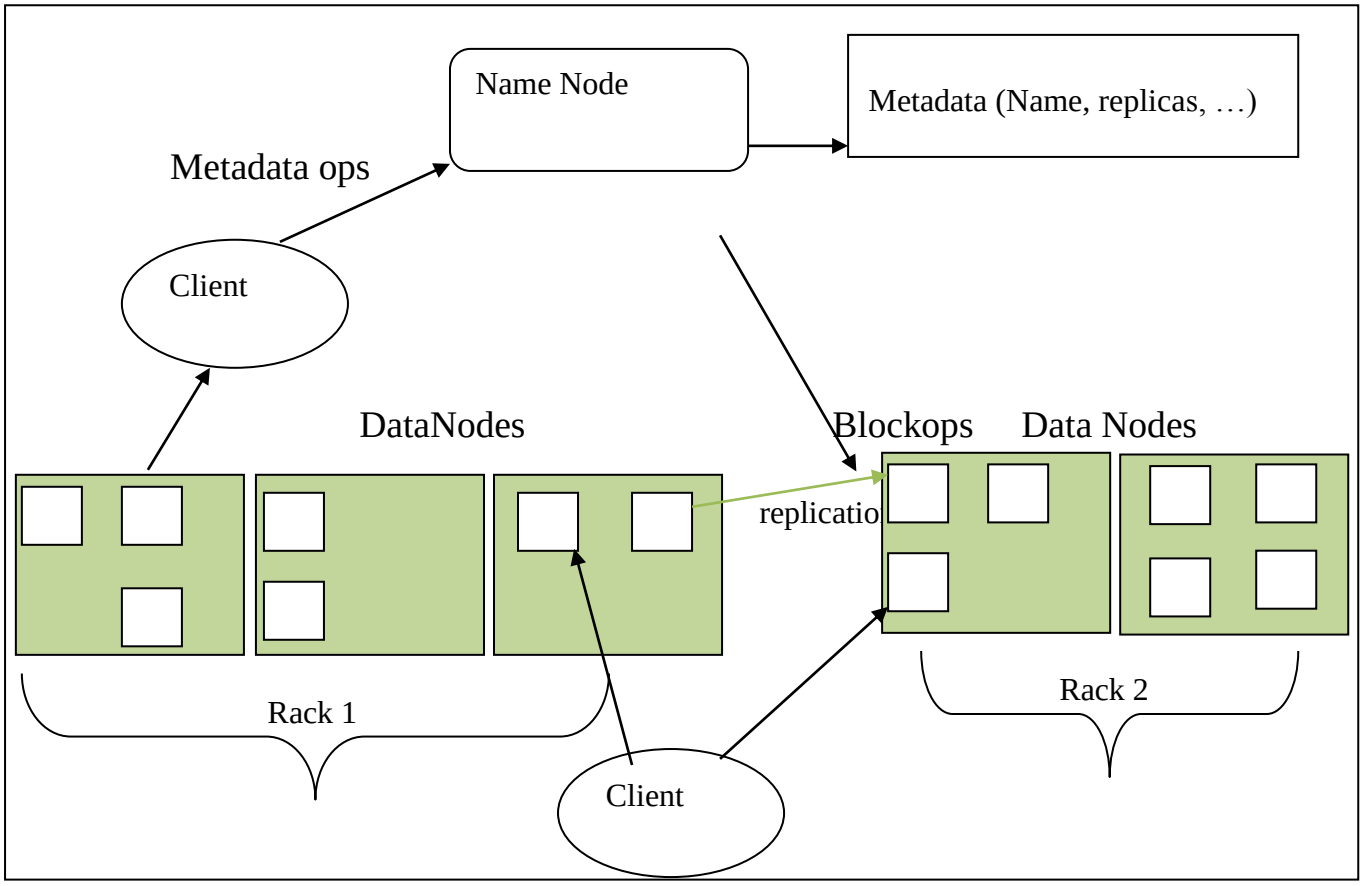


Figure 2: HDFS Architecture

Figure 2 depicts the HDFS architecture. It stores data in redundantly across various nodes in order to increase reliability. Namenode stores the metadata about which files are stored in which nodes and data nodes actually contain the data.

3.1.1 Configuration of the Parallel HDFS:

A processing system for big data is implemented that can handle structured and unstructured data generated by twitter efficiently. It consists of MapReduce and a parallel HDFS. HDFS provided in the Hadoop stack is used to fetch and store the data. To analyse the data efficiently MapReduce is used.

Table 1: HDFS servers

Server	Components	Functions
Master node	Main name node, data node, MapReduce	Main server for parallel distribution, data node, data loading
Slave node (1)	Secondary name node & data node	Backup server for main server
Slave node (2)	Data node	Data loading and analysing
Slave node (3)	Data node	Data loading and analysing

HDFS is made especially for using distributed computation efficiently by redundantly storing local data for processing on local nodes and thereby reducing network bandwidth usage. When parallel configured, as shown in Figure 3, it uses 4 Linux based nodes, 64 megabytes chunks are used in each node for storing data. It maintains a copy of name server in case of any catastrophic failure. Various servers perform the following function.

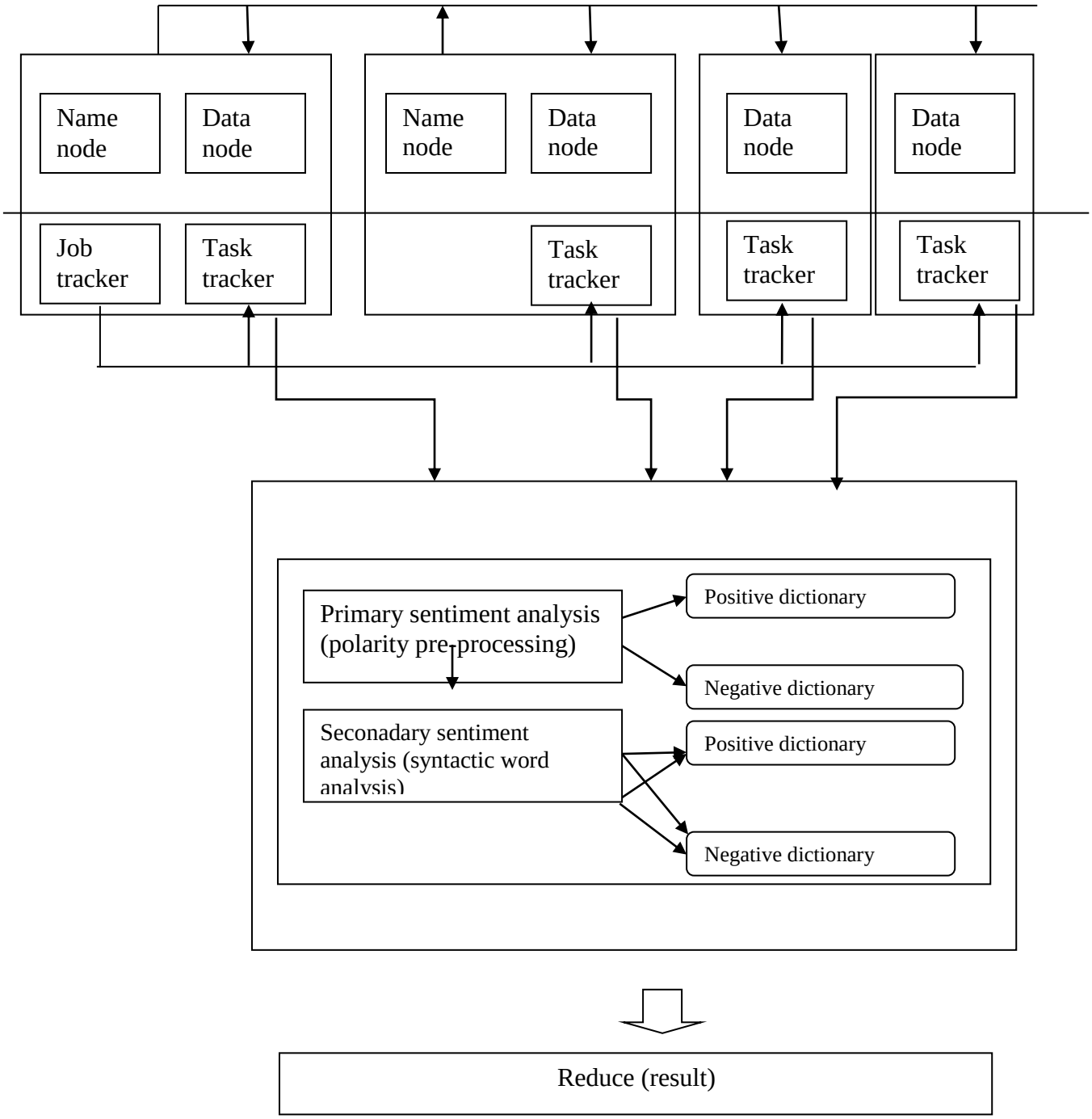


Figure 3: Data extraction

Figure 3 depicts the data extraction method using HDFS & Map Reduce. Each tweet is mapped as a separate job and managed by a JobTracker. Then TaskTracker run it on map reduce framework.

3.1.2 Map Reduce Functions for Sentiment Analysis:

MapReduce is a programming paradigm. It utilises distributed computing by allowing various parallel nodes to work on the data simultaneously using the concept called Map.

It firsts assigns different parts of data to the nodes in which it is locally stored to reduce bandwidth usage and the results of various nodes are combined using "reduce" function.

Here each mapper calculates the sentiment of each tweet and sends the score to reducer for sorting the tweets by sentiments.

Dictionaries for Sentiment Analysis:

Two dictionaries of each positive and negative words respectively is used here. Each dictionary contains English language words classified as positive or negative.

3.2 Sentiment analysis Procedure

This section focuses on the process for computing sentiment in accordance with previously explained concepts. The various steps for computing sentiment using MapReduce and HDFS are explained as following. First, twitter data is fetched from using the Twitter Developer API using Apache Flume. Flume is also a part of Apache's Hadoop distribution. It provides the ability to ingest high volume streaming data. Second, relevant data from the data gathered in previous step is extracted. Third, data then extracted is saved in HDFS. Fourth, the data is then transferred into the parallel HDFS. Fifth, sentiment is computed with the Mapper and Reducer functions.

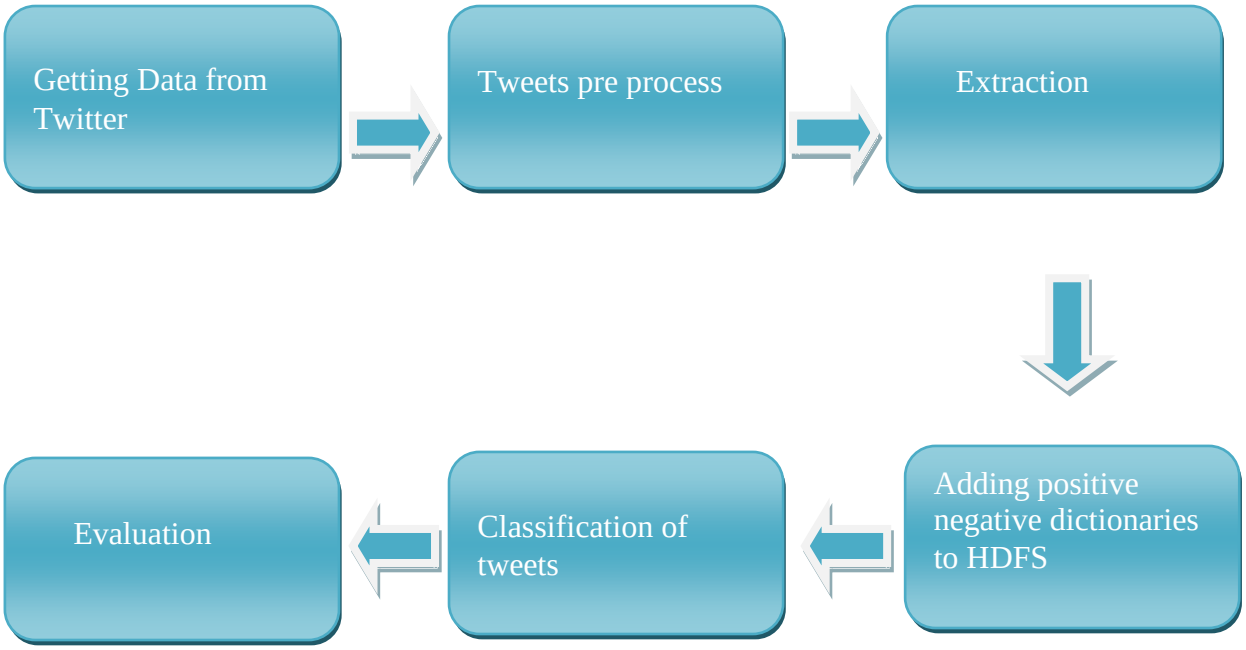


Figure 4: Flow chart for sentiment analysis process.

Figure 4 presents the flow chart for sentiment analysis process. Initial steps focus on fetching and storing data, preparing metadata whereas last three steps focus on computation and visualisation of results.

3.2.1 Phase –I

Download the tweets (data set) from twitter relating to a particular event using Apache Flume.

Apache Flume

- Flume is a part of Hadoop stack that provides reliable and distributed service for fetching, combining, and transferring big data.

- Its architecture is simple and focuses on streaming data for example router logs.

The data collection method of the proposed system was processed through Twitter.

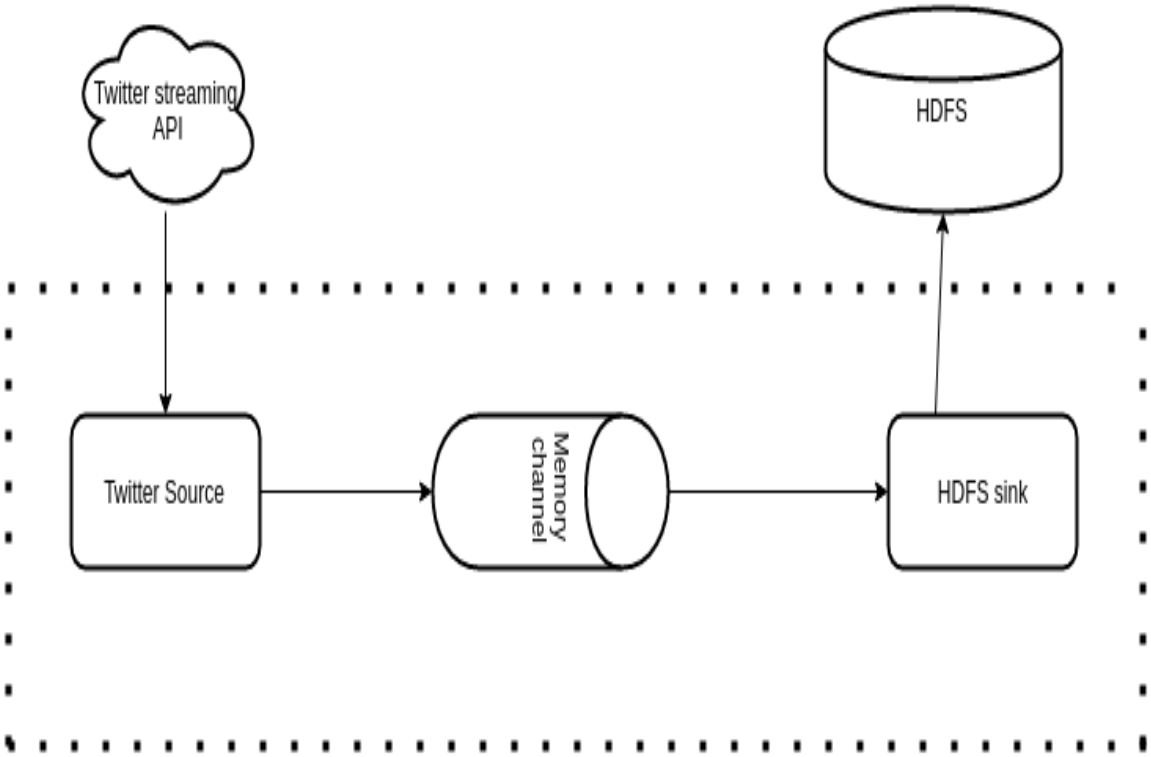


Figure 5: Fetching tweets

Figure 5 discusses the mechanism of fetching tweets from Twitter using Apache Flume. It uses a three tier source channel sink mechanism for storing tweets.

3.2.2 Phase –II

<u>Extraction and Processing of Necessary Data</u>

(Java Script Object Notation)JSON is a lightweight data-interchange format and it is language independent. JSON format is somewhat like that of XML and is similar to the code for creating JavaScript objects.

Some JSON syntax rules:

- All the data is in key value pairs.
- Each key is separated by commas.
- Objects are held in curly braces.

Twitter JSON reply structure (fields explained below):


```
{
  "extended_entities": {
    "media": [
      {
        "pic_url":"pic.twitter.com/rYYcfh8ZHE",
        "type":"photo",
        "source_user_id": 744912907515854848,
        "url":"https://t.co/rYYcfh8ZHE",
        "source_status_id":764216391251746820,
        "media_url":"http://pbs.twimg.com/media/CpsKfqiW8AACJ9E.jpg",
        "indices":[
          139,
          140
        ],
        "sizes":{
          "small":{
            "h":356,
            "w":680,
            "resize":"fit"
```

"large": {
"h":607,
"w":1158,
"resize":"fit"
},
"thumb":{
"resize":"crop"
"h":150,
"w":150,
},
"medium":{

"resize":"fit"
"h":607,
"w":1158,
}
},
"id_str":"764216100880052224",
"expanded_url":
"http://twitter.com/GUCCIFER_2/status/764216391251746820/photo/1",
"source_user_id_str":"744912907515854848"
"source_status_id_str":"764216391251746820",
"media_url_https":"https://pbs.twimg.com/media/CpsKfqiW8AACJ9E.jpg",
"id": 764216100880052224,
},
{
"display_url":"pic.twitter.com/rYYcfh8ZHE",
"type":"photo",
"media_url":"http://pbs.twimg.com/media/CpsKfrUWcAA8wmJ.jpg",
"source_user_id":744912907515854848,
"url":"https://t.co/rYYcfh8ZHE",

"url":"https://t.co/rYYcfh8ZHE",
"source_status_id":764216391251746820,
"indices":[
139,
140
],
"sizes":{
"small":{
"resize":"fit"
"h":626,
"w":628,
},
"large":{

"id_str":"764216101089734656",
"expanded_url":
"http://twitter.com/GUCCIFER_2/status/764216391251746820/photo/1",
"media_url_https":"https://pbs.twimg.com/media/CpsKfrUWcAA8wmJ.jpg",
"source_status_id_str":"764216391251746820",
"source_user_id_str":"744912907515854848"
"id":764216101089734656,
}
]
}

3.2.3 Phase –III

Data Loading and Sentiment Analysis

As described earlier the data is stored in parallel HDFS. Map and Reduce functions then compute the sentiment by fetching data from parallel HDFS. Figure 4 shows the general functionality of Map and Reduce. After mapper phase sentences which are analysed are separated into key value pairs. Then they are sorted using key value as sorting criteria. At last the data are reduced using key to obtain final results. The final scores, however, are attained by using the sentiment analysis function proposed in earlier section. Table 2 depicts a sample result by applying this procedure on some sample sentences.

```
// Determining sentiment
(1)     Inputs:
(2)     H – hashtag
(3)     D – dataset
(4)     S={S_1, S_2, S_3, ..., S_k} – sentences

Initialisation:
(5)     pcount – count of positive tweets, initialized 0
(6)     ncount – count of positive tweets, initialized 0

(7)     Output:
(8)             R – result

(9)     foreach (S_n ∈ S)
(10)            Calculate pcount(K, S_n)  //from positive words dictionary
(11)            Calculate ncount(K, Sn) //from negative words dictionary
(12)    end
(13)    if pcount and ncount are 0 then R=0
(14)    R = pcount – ncount
(15)    If R is 0 then R = pcount
(16)    return R
```

Algorithm 1: Dictionary based Polarity processing algorithm [2]

Table 2: An example of sentiment analysis results

Date	Hashtag	Tweet	Score
12/3/16	IndiaVsPak	Excellent performance by our boys!	+1
11/5/16	DonaldWins	I am very feeling sad after Trump victory	-1
23/6/16	Modi	Modi is best pm ever!	+1

Knowledge base of positive and negative words:

Positive

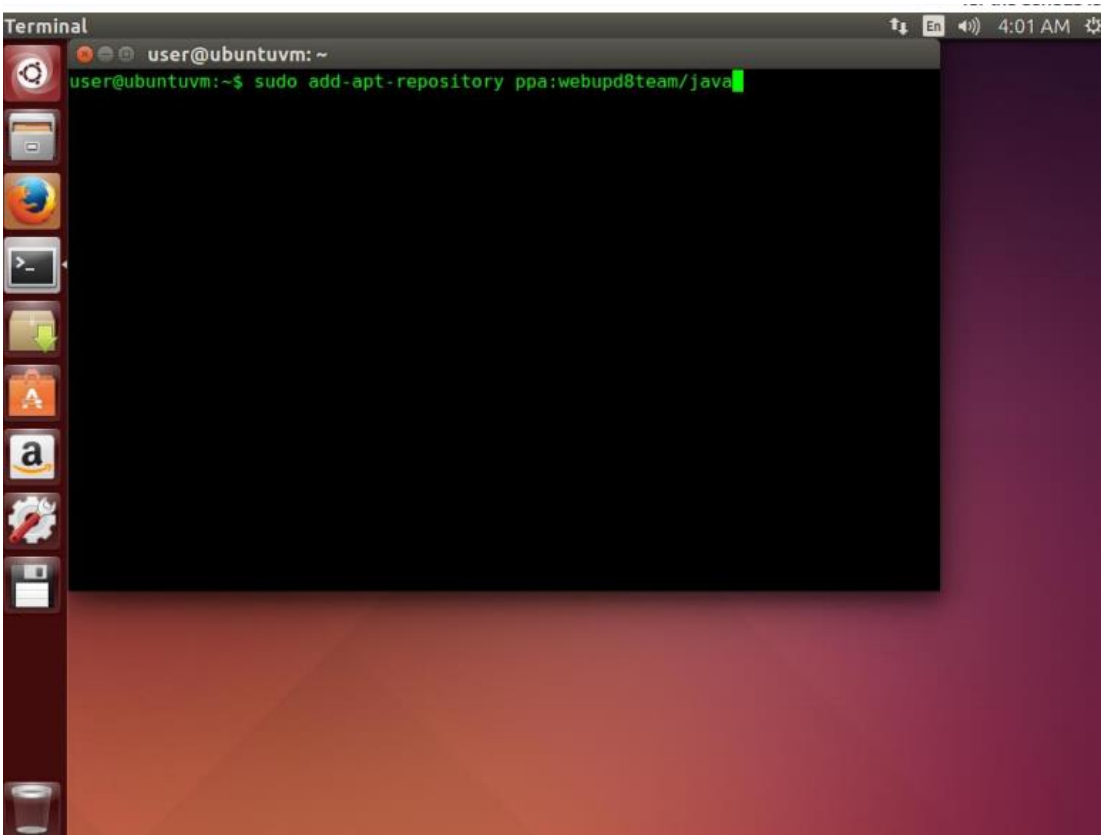
affirmative
affluence
affluent
afford
affordable
affordably
afordable
agile
agilely
agility
agreeable
agreeableness
agreeably
all-around
alluring
alluringly
altruistic
altruistically
amaze
amazed
amazement
amazes
amazing
amazingly
ambitious
ambitiously
ameliorate
amenable
amenity
amiability
amiabily
amiable
amicability
amicable
amicably

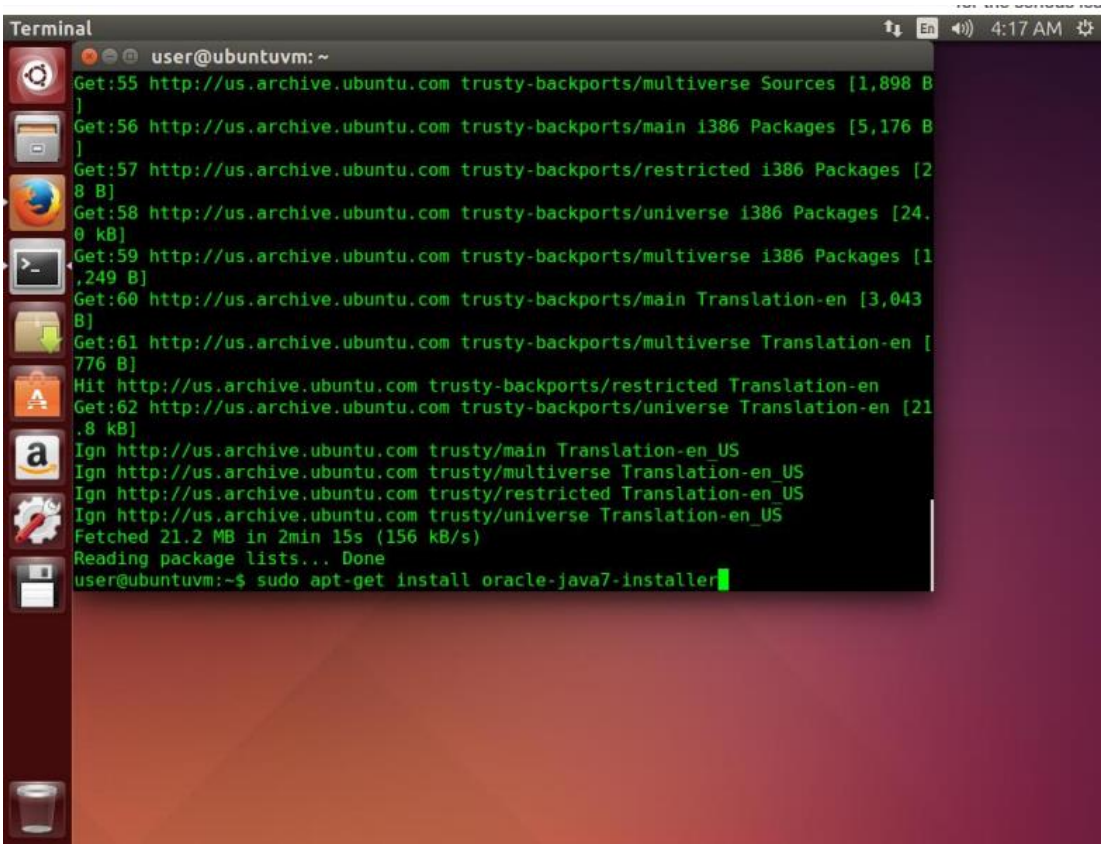
Negative

abuses
abusive
abysmal
abysmally
abyss
accidental
accost
accursed
accusation
accusations
accuse
accuses
accusing
accusingly
acerbate
acerbic
acerbically
ache
ached
aches
achey
aching
acrid
acridly
acridness
acrimonious
acrimoniously
acrimony
adamant
adamantly
addict
addicted
addicting
addicts

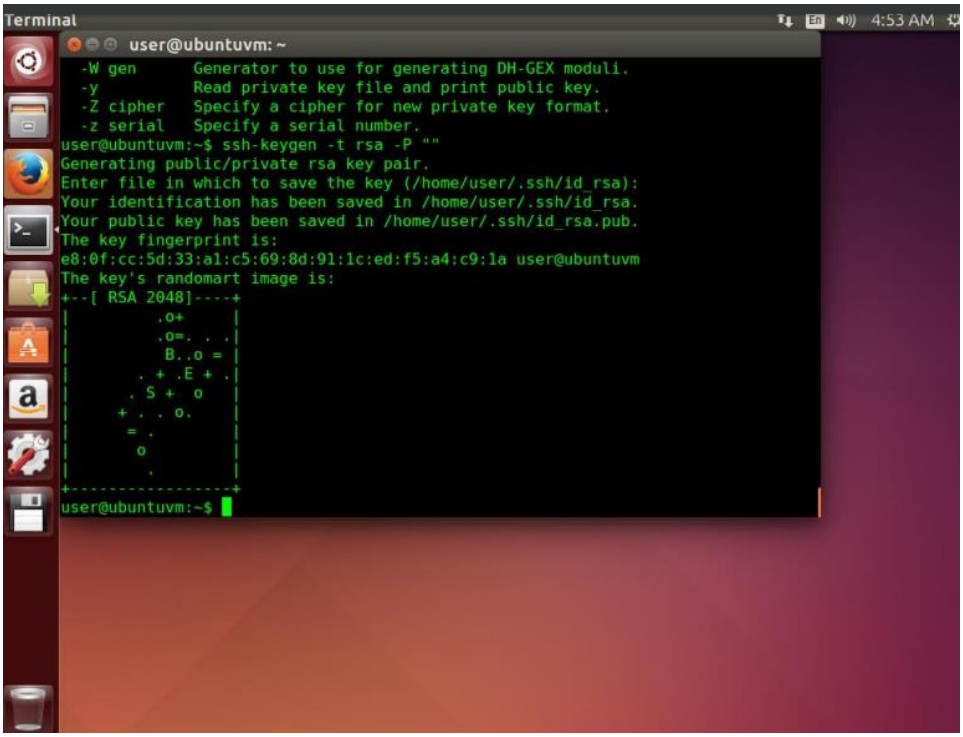
3.3 Environment Setup

Step I: Install Java, add the repository as shown in screen below





Step II: Now install openssh server

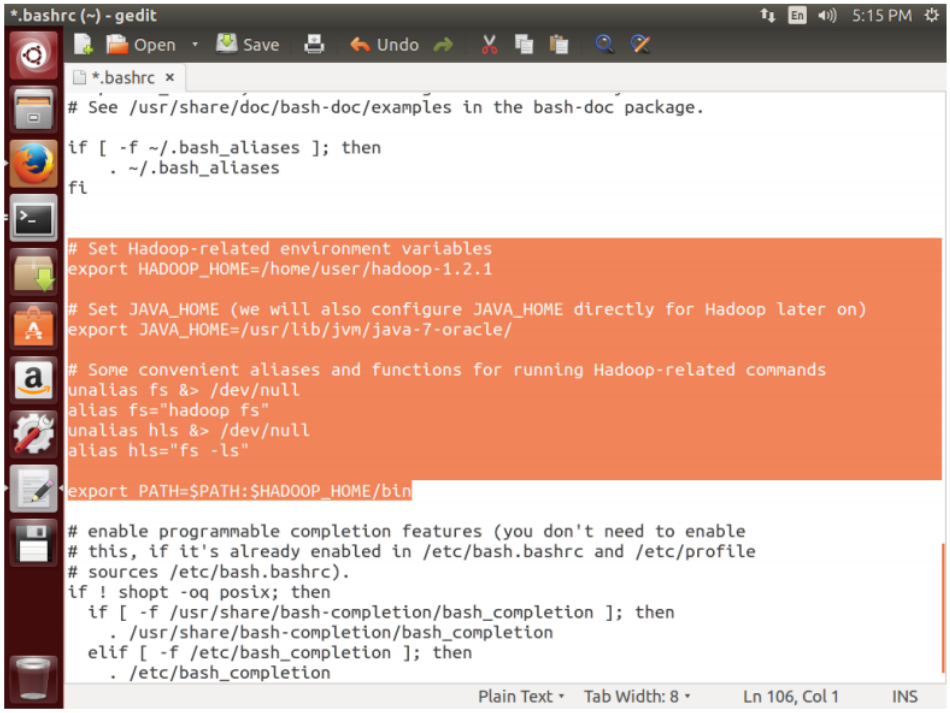


Step III: Download Hadoop release from this URL –

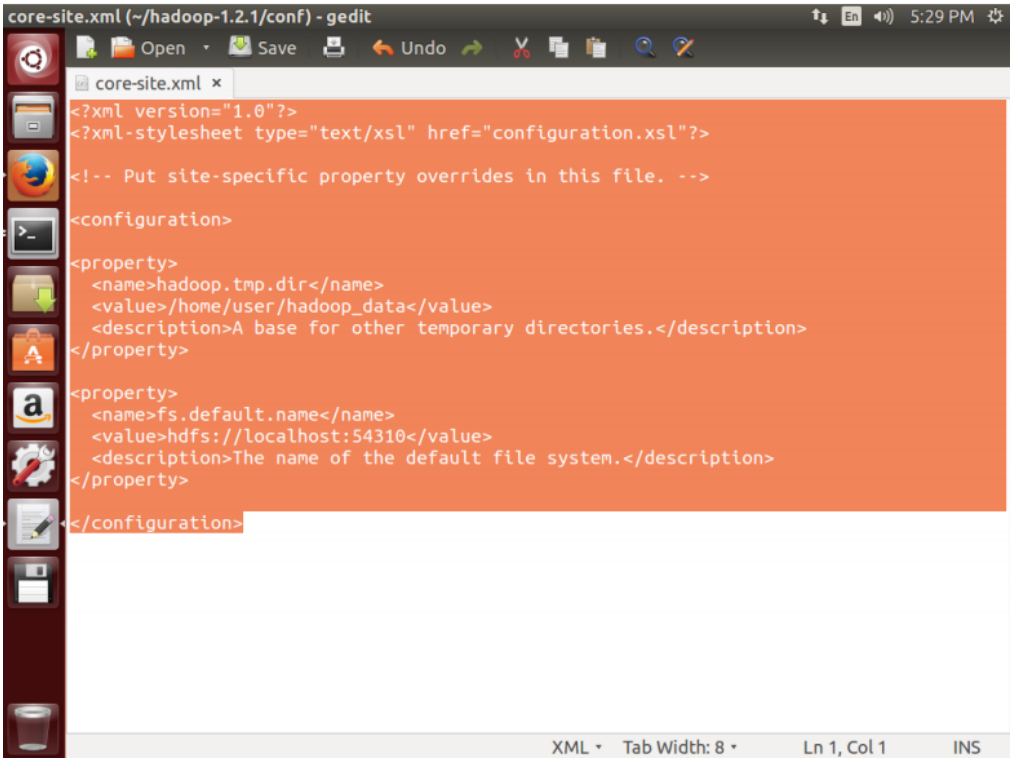
http://apache.bytenet.in/hadoop/common/stable1/

And download Hadoop.tar.gz

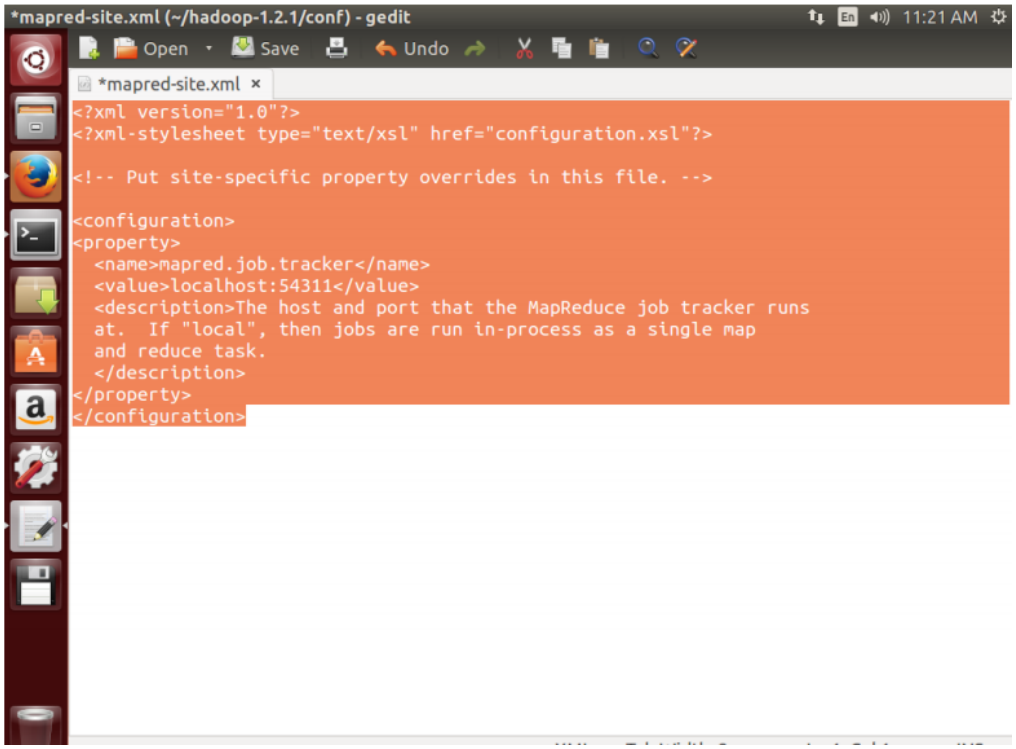
Step IV: Now open your .bashrc as shown below



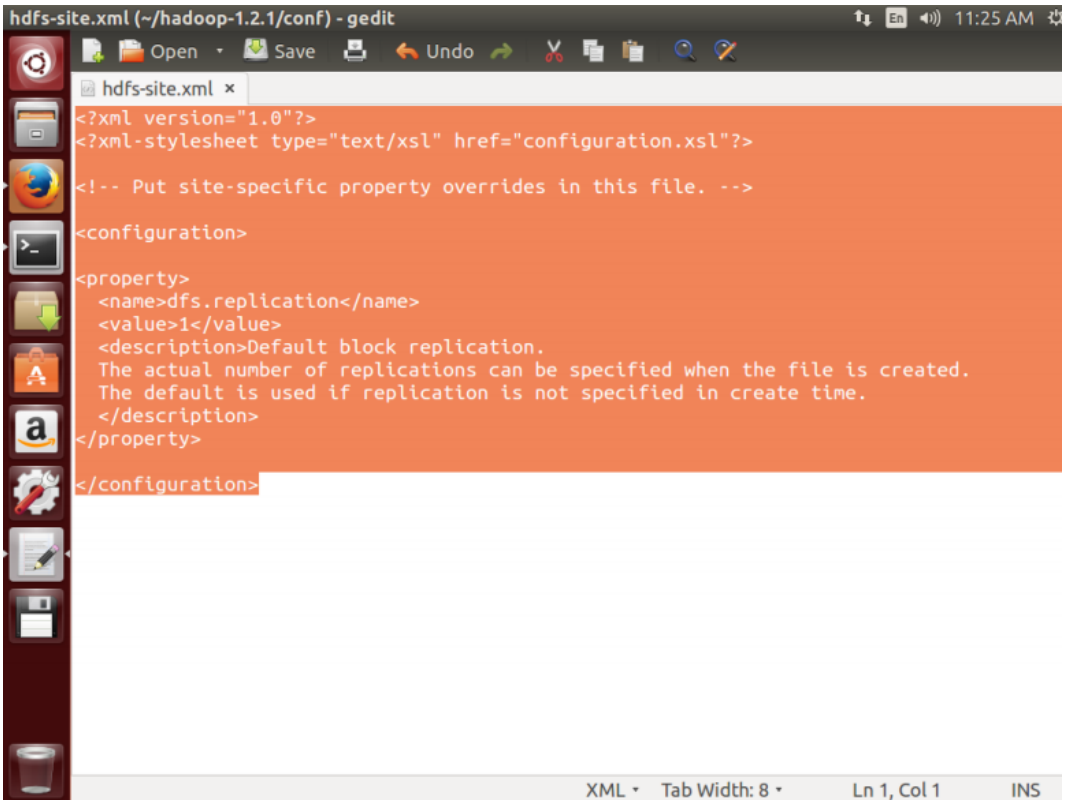
Step V: Now edit the core-site.xml as follows:



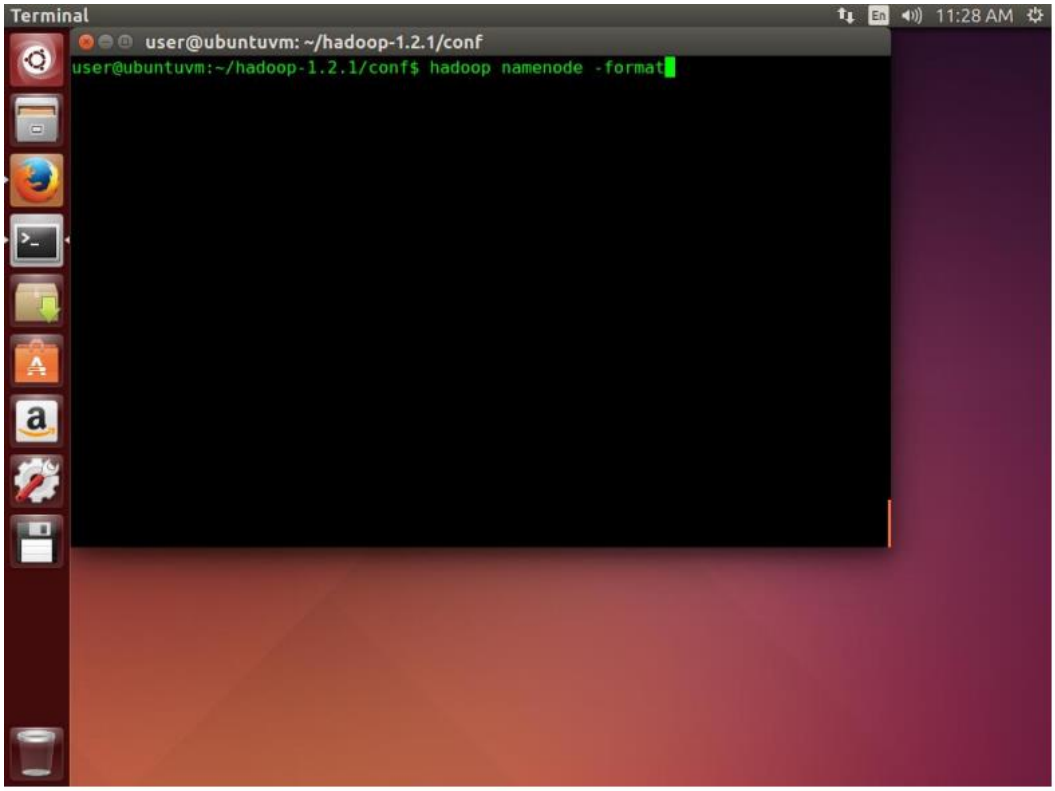
Step VI: Now edit the mapred-site.xml



Step VII: Edit hdfs-site.xml



Step VIII: Format namenode



Now the system has been successfully configured to use Map Reduce application

Chapter-4
Performance Analysis

4.1 Using a multi-node cluster

To make the best use of distributed computation model of Hadoop we have to use multiple nodes in a cluster so that they can work in parallel and reduce the computation time. But setting up a multimode cluster itself requires special hardware and administration skills.

Here cloud computing comes to rescue. It provides on demand delivery of resources and is reliable and secure. In this project we used the cloud services offered by amazon as AWS.

We have basically two needs here. One is storage and other is computation. In AWS tier we can use Amazon S3 for storage and Amazon EMR for computation.

Amazon S3:

Amazon's "Simple Storage Service (Amazon S3) provides a user friendly simple web interface for storing and accessing virtually any amount of data on their servers. It can easily scale past trillions of objects worldwide and deliver 99.999999% durability."

Amazon Elastic Map Reduce:

"Amazon EMR delivers the complete Hadoop stack as a service. One can also make use of other popular distributed frameworks such as HBase, Apache Spark, Flink, Presto.

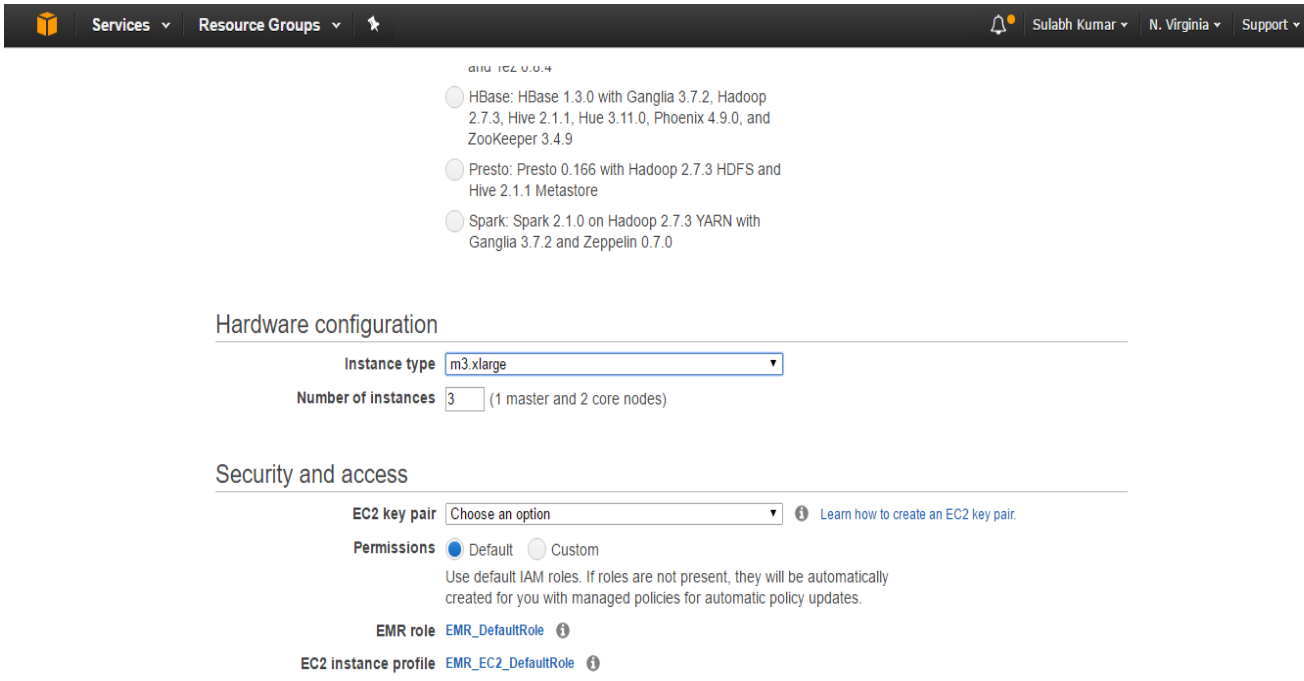


Figure 7 shows the configuration of cluster on AWS for processing the sentiments

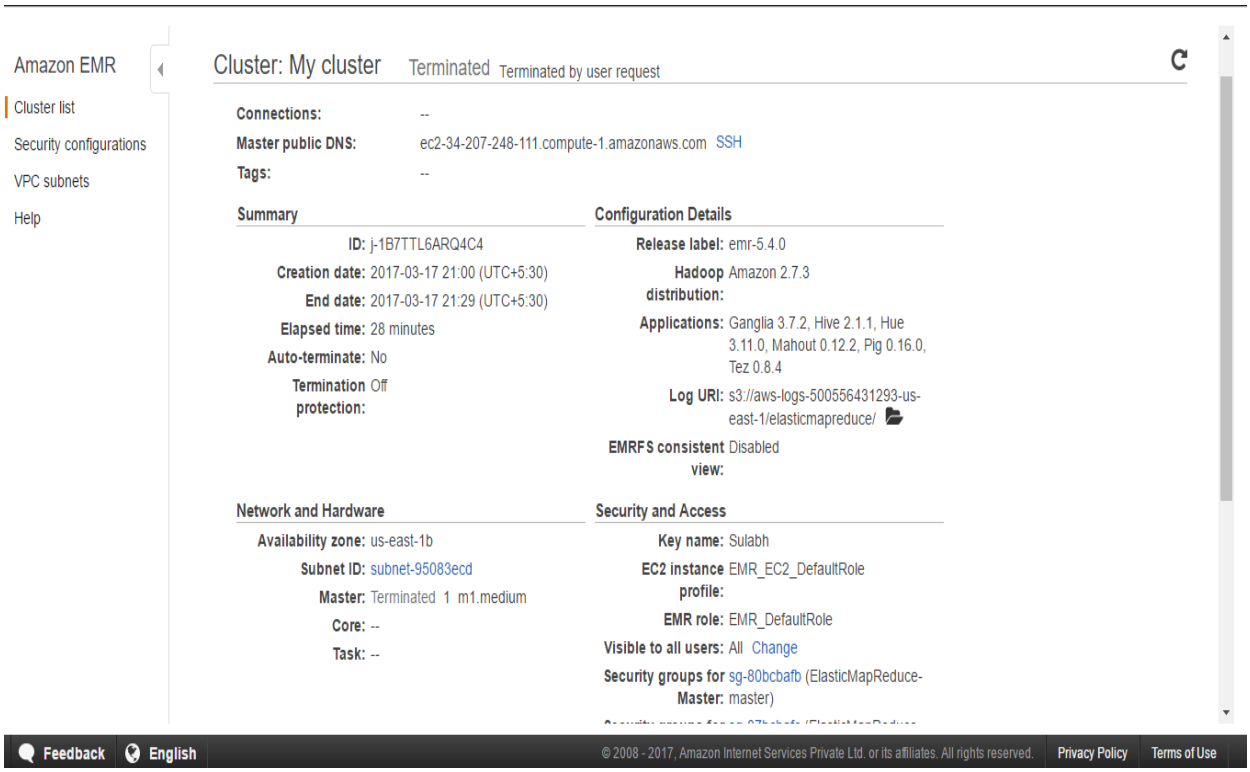


Figure 8: showing the successful termination of a cluster instance

4.2 Output

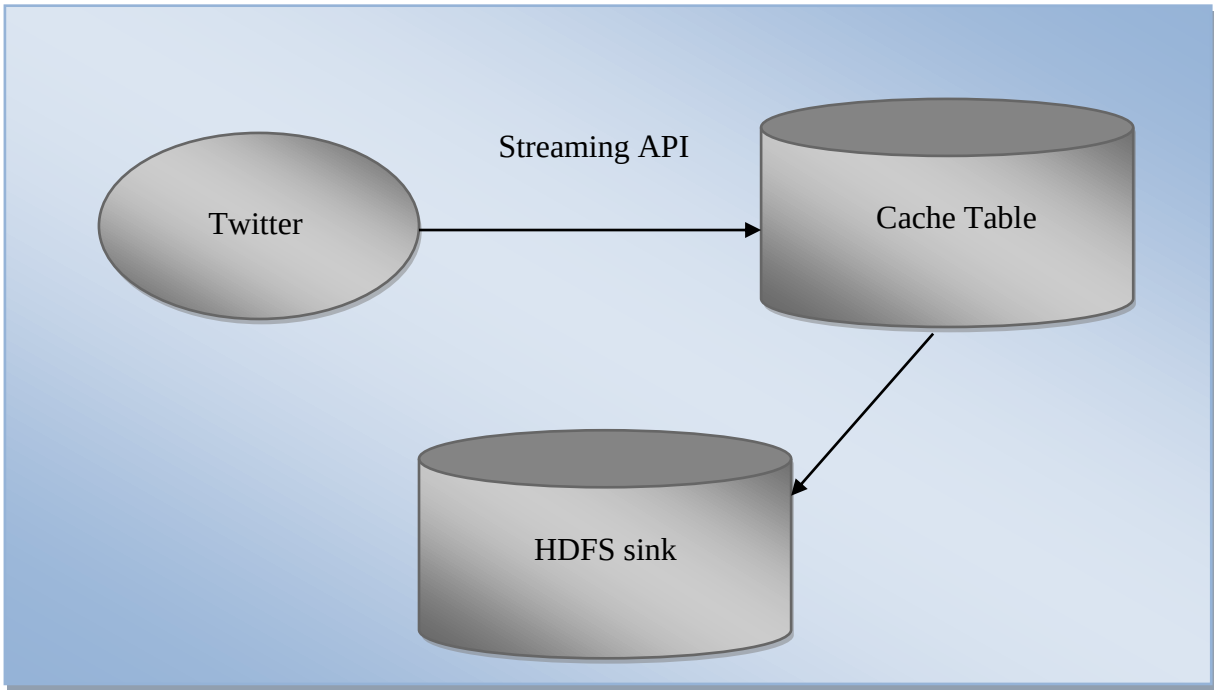
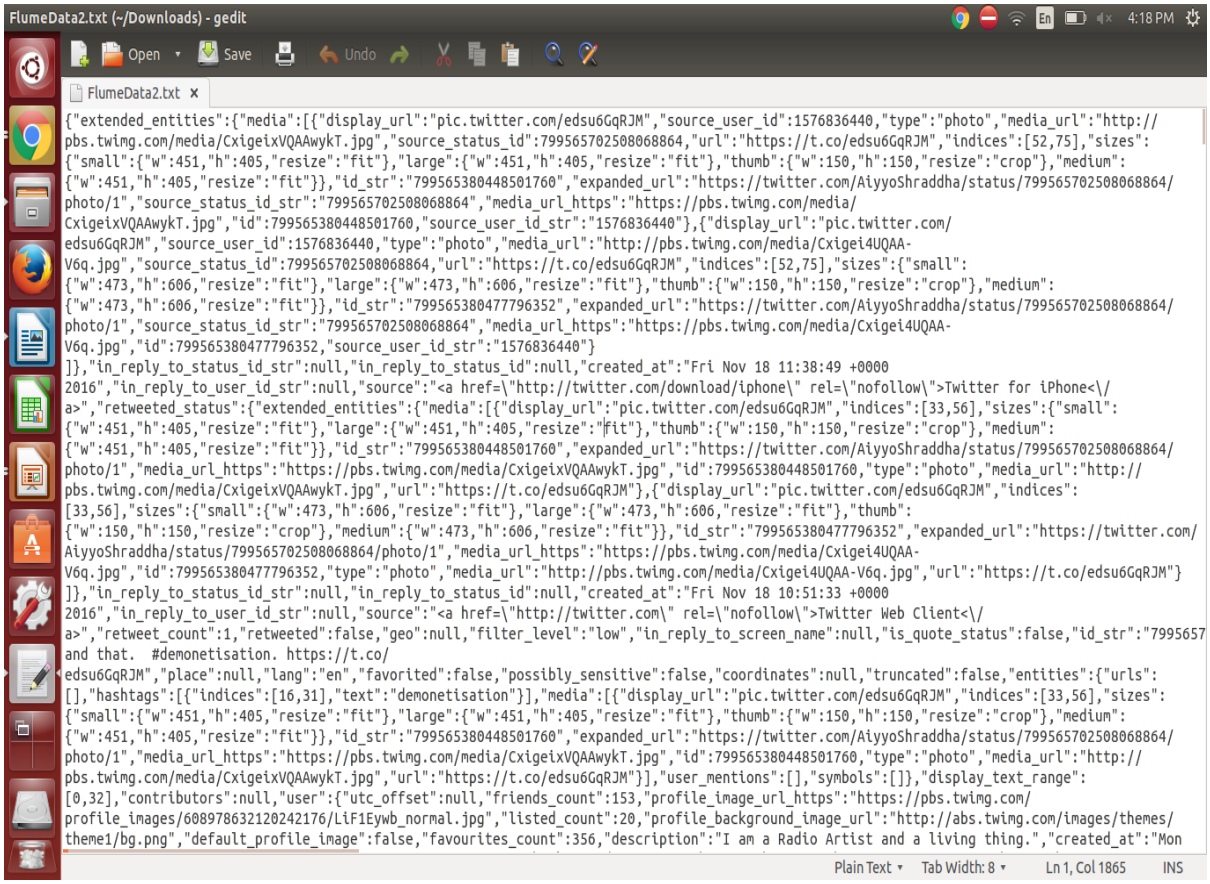


Figure 6: Data flow from Twitter to HDFS.

Figure 6 depicts the data flow from Twitter to HDFS. Twitter Developer's Streaming API is queried with a specific hashtag and it returns dataset about that hashtag

On querying Twitter API, it returns tweets in the JSON format as described before. Flume acts as an agent between Twitter and HDFS. It helps ingesting a high volume of tweets using its source-channel-sink architecture. Here source is the Twitter API, channel is primary memory and sink is HDFS.

4.1.1 Step I (Result): Using Twitter API to fetch data



Raw twitter data in JSON format

4.1.2 Step 2(Intermediate): Extraction of important tags from raw JSON tweets file

4.1.3Step 3 (Result): polarity processing to determine the sentiment

Column A: sentiment (1 -> positive, -1 -> negative & 0 -> neutral)

Column B: Actual tweet

Column C: Location

part-r-00000.csv - LibreOffice Calc

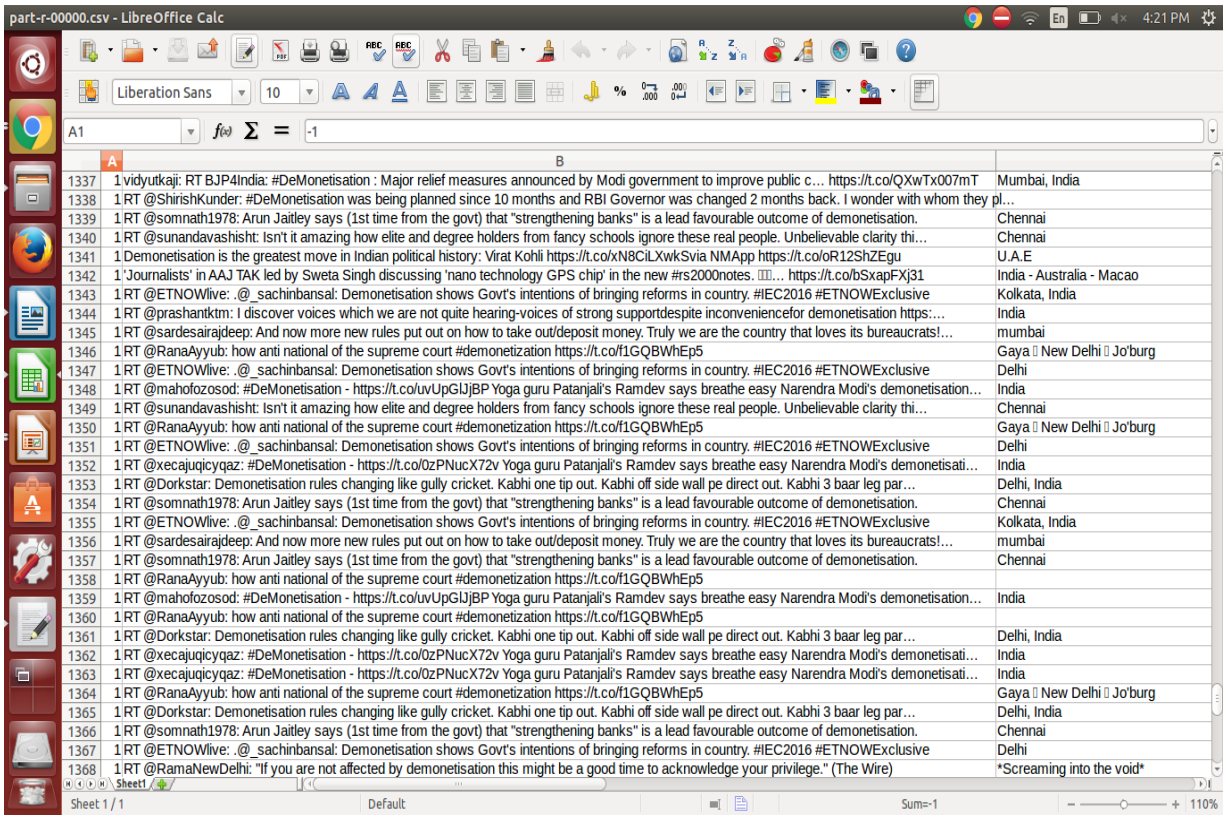
	A	B	
1	-1	#Demonetisation I don't hv BM Y trouble me 2stand in Q at bank/ATM?Soilder-I don't hv personal fight with Pak Y shud I face bullets 4 u?	
2	-1	RT @indiantweeter: Unconfirmed Reports that ruling party of delhi has lost around 3400 crore in cash because of demonetisation may create…	New Delhi, India
3	-1	RT @brumbyOz: Dont panic. This appears to be a rumor. At least in this matter news channels should not quote sources and wait for…	Bermuda Triangle
4	-1	RT @_YogendraYadav: Lousy execution of #DeMonetisation but opposition demand for roll-back not responsible. Fake notes will come back RBI…	Delhi, India
5	-1	Property Prices In Delhi-NCR To Fall Thanks To Demonetisation https://t.co/OZtGXrvBEzvia NMApp https://t.co/QcIdHO7bb5	U.A.E
6	-1	RT @_MiteshPatel: It's your mistake.Instead of attacking the lack of Planning you should have bluntly said that DeMonetisation is…	
7	-1	RT @pbhushan1: Was Modi's unplanned demonetisation decision taken in panic due to imminent exposure of payoffs of Crores to him by…	
8	-1	RT @maqbool_sm: PM @narendramodi should tender apology to 55 families who lost their family members due to the #demonetisation chaos: @rssu	Jhabua, India
9	-1	10 MORE rumours related to demonetisation that you might have believed as true https://t.co/vpMqb0syJf via @opindia_com	Bharat
10	-1	RT @mkvenu1: The Guardian view on India's demonetisation: Modi has brought havoc to India | Editorial https://t.co/FJPd2FSwhk	singapore
11	-1	RT @mkvenu1: The Guardian view on India's demonetisation: Modi has brought havoc to India | Editorial https://t.co/FJPd2FSwhk	singapore
12	-1	RT @maqbool_sm: PM @narendramodi should tender apology to 55 families who lost their family members due to the #demonetisation chaos: @rssu	Jhabua, India
13	-1	RT @pbhushan1: Was Modi's unplanned demonetisation decision taken in panic due to imminent exposure of payoffs of Crores to him by…	
14	-1	RT @abpnewstv: Do not miss 'Meri Awaaz Suno' tonight at 8 PM only on ABP News#demonetisation https://t.co/gLWeYgFkgi	Россия, Барнаул
15	-1	RT @indiantweeter: Unconfirmed Reports that ruling party of delhi has lost around 3400 crore in cash because of demonetisation may create…	
16	-1	RT @IndianExpress: Centre's decision on demonetisation smacks of insensitivity: Arvind Kejriwal | https://t.co/FyYJYLbZLJ https://t.co/kCg8…	Kingdom of Saudi Arabia
17	-1	RT @indiantweeter: Unconfirmed Reports that ruling party of delhi has lost around 3400 crore in cash because of demonetisation may create…	New Delhi, India
18	-1	RT @_YogendraYadav: Roll-back of #DeMonetisation not possible nor needed.Govt can still save national trauma by granting universal exempt…	Delhi, India
19	-1	10 MORE rumours related to demonetisation that you might have believed as true https://t.co/vpMqb0syJf via @opindia_com	Bharat
20	-1	RT @thomasisaaq: The first major protest in the country against the manner in which #demonetisation is being implemented starts before @RBI…	KL 10
21	-1	RT @AAPInNews: #DeMonetisation Sikar: Days before daughters' wedding tea seller dies https://t.co/r6FEdNu0vb	
22	-1	RT @_MiteshPatel: It's your mistake.Instead of attacking the lack of Planning you should have bluntly said that DeMonetisation is…	
23	-1	RT @timesofindia: Supreme Court refuses to restrain lower courts from hearing pleas on demonetisation https://t.co/H0GXZ5dlD6 https://t.co/…	Vizag, INDIA
24	-1	RT @brumbyOz: Dont panic. This appears to be a rumor. At least in this matter news channels should not quote sources and wait for…	Bermuda Triangle
25	-1	RT @timesofindia: Supreme Court refuses to restrain lower courts from hearing pleas on demonetisation https://t.co/H0GXZ5dlD6 https://t.co/…	Vizag, INDIA
26	-1	RT @pbhushan1: Was Modi's unplanned demonetisation decision taken in panic due to imminent exposure of payoffs of Crores to him by…	
27	-1	Property Prices In Delhi-NCR To Fall Thanks To Demonetisation https://t.co/OZtGXrvBEzvia NMApp https://t.co/QcIdHO7bb5	U.A.E
28	-1	#Demonetisation I don't hv BM Y trouble me 2stand in Q at bank/ATM?Soilder-I don't hv personal fight with Pak Y shud I face bullets 4 u?	
29	-1	RT @ChitraSarwara: The govt's demonetisation has devastated farmers landless labourers middle class pensioners &amp	petty traders.#YouthCo…
30	-1	RT @_YogendraYadav: Lousy execution of #DeMonetisation but opposition demand for roll-back not responsible. Fake notes will come back RBI…	Delhi, India
31	-1	RT @IndianExpress: Long queues outside banks a 'serious issue' says SChttps://t.co/YSzUJwMb6U https://t.co/1vKZFNIW0r	
32	-1	RT @IndianExpress: Centre's decision on demonetisation smacks of insensitivity: Arvind Kejriwal | https://t.co/FyYJYLbZLJ https://t.co/kCg8…	Kingdom of Saudi Arabia

Sheet 1 / 1 Default Sum=-1 110%

part-r-00000.csv - LibreOffice Calc

	A	B	
531	0	.@_sachinbansal: E-commerce industry did witness a small dip post demonetisation move. #IEC2016 #ETNOWExclusive	India
532	0	RT @myvotetoday: #DoublePoll विभिन्न पोल में 85-95% लोग #DeMonetisation के पक्ष में हैंकौन सा TV चैनल जानबूझकर विरोधकर रहे लोगों की…	New Delhi, India
533	0	RT @naoyafujiwara: 紙幣交換でインドの金融システムに対する信頼が一瞬で揺らいでしまったhttps://t.co/9zHimLXaXk	
534	0	A Bride Without An Engagement Ring. How Demonetisation Has Affected Indian Weddings https://t.co/gWE6OfzfPV	India
535	0	RT @ShirishKunder: The important question is: Why was #DeMonetisation announced so suddenly without any preparation? What was the EMERGENCY?	
536	0	RT @mediacrooks: Today @ghulamnazad SPAT on our Uri martyrs &amp	whitewashed crimes of Pak terrorists
537	0	RT @vikaskyogi: SC raps Modi Gov. SC To Government On Demonetisation: 'We Will Have Riots On The Streets' https://t.co/5yFbecRH9S	
538	0	RT @aajtak: नोटबंदी से ध्वस्त हो गई कंपनियों की तीसरी तिमाही जनवरी के नतीजे में दिखेगा असर https://t.co/uVyPhtNJy6 https://t.co/k0VOxp85V4	India
539	0	RT @KotianPravin: सुप्रीम कोर्ट की कड़ी टिप्पणी मामला गंभीर सड़कों पर हो सकते हैं दंगे- Amarujala https://t.co/xT4LuD5Xz2... https://t.co…	Kingdom of Saudi Arabia
540	0	RT @MANJULtoons: #DeMonetisation #CashCrunch My #cartoon https://t.co/ygG5zL1RPW	
541	0	RT @afpohindia: #SupremeCourt Should Not Stay Proceedings Against #Demonetisation In Various High Courts In Present Circumstances-…	New Delhi, India
542	0	RT @AAPlogical: Modi should learn from Bill Gates how to rollback Day 1: Lauds Demonetisation Day 2: No opinion on it https://t.co/Ga6yPC…	
543	0	RT @Vidyut: Only U-Turns. https://t.co/gHnVwA8Mws	Pune, India
544	0	मनसे नेता अखिल चित्रे ने सुप्रीम कोर्ट में सरकार पर नोटबंदी पर पहले से व्यवस्था ना करने के खिलाफ याचिका दायर की#DeMonetisation	Noida, India
545	0	RT @YRDeshmukh: If you are Modi's opponent and you think #DeMonetisation will ensure his defeat then don't ask for a rollback. Sit back an…	Ahmedabad
546	0	RT @RanaAyyub: the jumla for UP elections https://t.co/vMIAPwWSJ1	Hyderabad
547	0	RT @devyanidilli: Moradabad trader AvinashGupta deposit Rs 1.55 L in 10s 50s 100s fr ppl in QRespect#DeMonetisation #IAmWithModi https…	
548	0	RT @ShirishKunder: The important question is: Why was #DeMonetisation announced so suddenly without any preparation? What was the EMERGE	New Delhi, India
549	0	RT @ndtv: Exchange of old notes for new may be stopped: government sources#demonetisation	New Delhi
550	0	RT @ShirishKunder: The important question is: Why was #DeMonetisation announced so suddenly without any preparation? What was the EMERGE	Mumbai
551	0	RT @AiyyoShraddha: This and that. #demonetisation. https://t.co/edsu6GqRJM	Mysuru, India
552	0	Taxman sends notice to Tata Trusts following demonetisation:Srcs https://t.co/72QsUkgnta https://t.co/gxJJ96ONlt	
553	0	https://t.co/7rRDSXLePAराष्ट्रवाद के हर मुद्दे पर मोदी तुम सब पर भारी है। तुम सब को कुर्सी प्यारी उसे माँ भारती... https://t.co/59w0m6ECg4	Bangalore
554	0	RT @sharmanagendar: SC To Government On Demonetisation: 'We Will Have Riots On The Streets': https://t.co/yUvXGcCktj	
555	0	RT @ShirishKunder: This is 'Gujarat Riots' without planning. https://t.co/MLSpG6gwpg	Surat,Gujarat
556	0	RT @CNNnews18: The dreaded dacoit of 70s &amp	80s Malkhan Singh too was seen lini
557	0	RT @MANJULtoons: #DeMonetisation #CashCrunch My #cartoon https://t.co/ygG5zL1RPW	Pune
558	0	RT @EconomicTimes: PM @narendramodi should apologise to the 125 crore people of the country: @rssurjewala on #demonetisation https://t.co/u…	New Delhi
559	0	RT @ETNOWlive: .@_sachinbansal: E-commerce industry did witness a small dip post demonetisation move. #IEC2016 #ETNOWExclusive	Kolkata, India
560	0	RT @sharmanagendar: SC To Government On Demonetisation: 'We Will Have Riots On The Streets': https://t.co/yUvXGcCktj	Maryland, USA
561	0	RT @TimesNow: No action only announcements	action-less whimsical announcemen
562	0	RT @ShirishKunder: The important question is: Why was #DeMonetisation announced so suddenly without any preparation? What was the EMERGE	your heart

Sheet 1 / 1 Default Sum=-1 110%



	A	B	
1337	1	vidyutkaji: RT BJP4India: #DeMonetisation : Major relief measures announced by Modi government to improve public c… https://t.co/QXwTx007mT	Mumbai, India
1338	1	RT @ShirishKunder: #DeMonetisation was being planned since 10 months and RBI Governor was changed 2 months back. I wonder with whom they pl…	
1339	1	RT @somnath1978: Arun Jaitley says (1st time from the govt) that "strengthening banks" is a lead favourable outcome of demonetisation.	Chennai
1340	1	RT @sunandavashisht: Isn't it amazing how elite and degree holders from fancy schools ignore these real people. Unbelievable clarity thi…	Chennai
1341	1	Demonetisation is the greatest move in Indian political history: Virat Kohli https://t.co/xN8CiLXwkSvia NMApp https://t.co/oR12ShZEgu	U.A.E
1342	1	'‘Journalists' in AAJ TAK led by Sweta Singh discussing 'nano technology GPS chip' in the new #rs2000notes. … https://t.co/bSxapFXj31	India - Australia - Macao
1343	1	RT @ETNOWlive: .@_sachinbansal: Demonetisation shows Govt's intentions of bringing reforms in country. #IEC2016 #ETNOWExclusive	Kolkata, India
1344	1	RT @prashantktm: I discover voices which we are not quite hearing-voices of strong supportdespite inconveniencefor demonetisation https:…	India
1345	1	RT @sardesairajdeep: And now more new rules put out on how to take out/deposit money. Truly we are the country that loves its bureaucrats!…	mumbai
1346	1	RT @RanaAyyub: how anti national of the supreme court #demonetization https://t.co/f1GQBWhEp5	Gaya New Delhi Jo'burg
1347	1	RT @ETNOWlive: .@_sachinbansal: Demonetisation shows Govt's intentions of bringing reforms in country. #IEC2016 #ETNOWExclusive	Delhi
1348	1	RT @mahofozosod: #DeMonetisation - https://t.co/uvUpGlJjBP Yoga guru Patanjali's Ramdev says breathe easy Narendra Modi's demonetisation…	India
1349	1	RT @sunandavashisht: Isn't it amazing how elite and degree holders from fancy schools ignore these real people. Unbelievable clarity thi…	Chennai
1350	1	RT @RanaAyyub: how anti national of the supreme court #demonetization https://t.co/f1GQBWhEp5	Gaya New Delhi Jo'burg
1351	1	RT @ETNOWlive: .@_sachinbansal: Demonetisation shows Govt's intentions of bringing reforms in country. #IEC2016 #ETNOWExclusive	Delhi
1352	1	RT @xecajuqicyqaz: #DeMonetisation - https://t.co/0zPNucX72v Yoga guru Patanjali's Ramdev says breathe easy Narendra Modi's demonetisati…	India
1353	1	RT @Dorkstar: Demonetisation rules changing like gully cricket. Kabhi one tip out. Kabhi off side wall pe direct out. Kabhi 3 baar leg par…	Delhi, India
1354	1	RT @somnath1978: Arun Jaitley says (1st time from the govt) that "strengthening banks" is a lead favourable outcome of demonetisation.	Chennai
1355	1	RT @ETNOWlive: .@_sachinbansal: Demonetisation shows Govt's intentions of bringing reforms in country. #IEC2016 #ETNOWExclusive	Kolkata, India
1356	1	RT @sardesairajdeep: And now more new rules put out on how to take out/deposit money. Truly we are the country that loves its bureaucrats!…	mumbai
1357	1	RT @somnath1978: Arun Jaitley says (1st time from the govt) that "strengthening banks" is a lead favourable outcome of demonetisation.	Chennai
1358	1	RT @RanaAyyub: how anti national of the supreme court #demonetization https://t.co/f1GQBWhEp5	
1359	1	RT @mahofozosod: #DeMonetisation - https://t.co/uvUpGlJjBP Yoga guru Patanjali's Ramdev says breathe easy Narendra Modi's demonetisation…	India
1360	1	RT @RanaAyyub: how anti national of the supreme court #demonetization https://t.co/f1GQBWhEp5	
1361	1	RT @Dorkstar: Demonetisation rules changing like gully cricket. Kabhi one tip out. Kabhi off side wall pe direct out. Kabhi 3 baar leg par…	Delhi, India
1362	1	RT @xecajuqicyqaz: #DeMonetisation - https://t.co/0zPNucX72v Yoga guru Patanjali's Ramdev says breathe easy Narendra Modi's demonetisati…	India
1363	1	RT @xecajuqicyqaz: #DeMonetisation - https://t.co/0zPNucX72v Yoga guru Patanjali's Ramdev says breathe easy Narendra Modi's demonetisati…	India
1364	1	RT @RanaAyyub: how anti national of the supreme court #demonetization https://t.co/f1GQBWhEp5	Gaya New Delhi Jo'burg
1365	1	RT @Dorkstar: Demonetisation rules changing like gully cricket. Kabhi one tip out. Kabhi off side wall pe direct out. Kabhi 3 baar leg par…	Delhi, India
1366	1	RT @somnath1978: Arun Jaitley says (1st time from the govt) that "strengthening banks" is a lead favourable outcome of demonetisation.	Chennai
1367	1	RT @ETNOWlive: .@_sachinbansal: Demonetisation shows Govt's intentions of bringing reforms in country. #IEC2016 #ETNOWExclusive	Delhi
1368	1	RT @RamaNewDelhi: "If you are not affected by demonetisation this might be a good time to acknowledge your privilege." (The Wire)	*Screaming into the void*

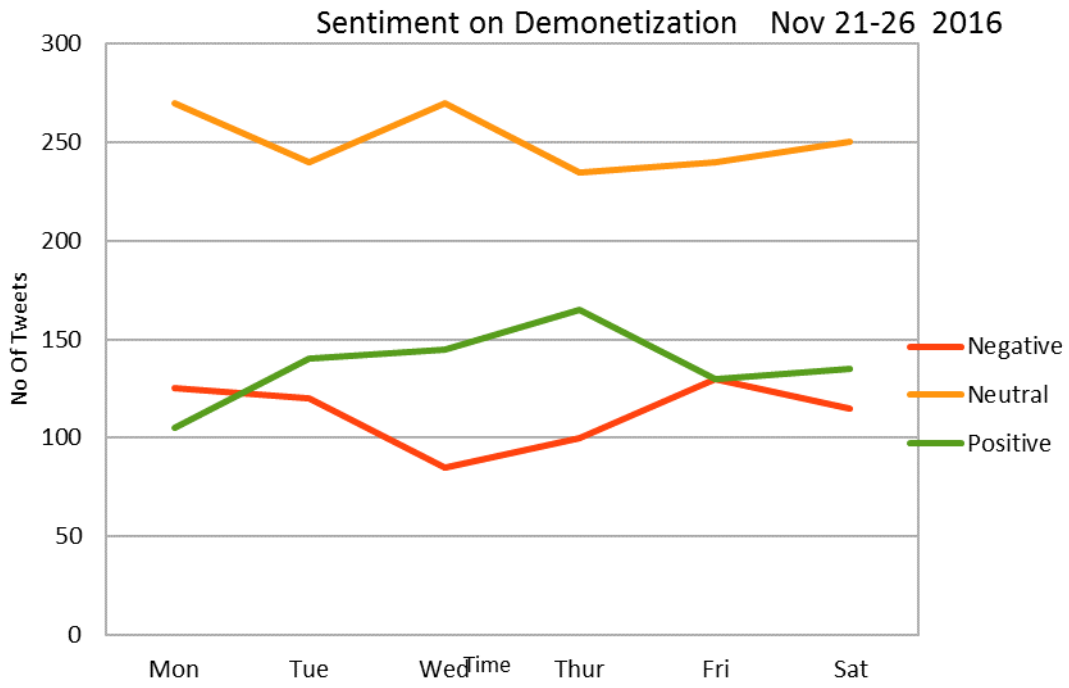
4.2 Observations

4.2.1 Result set for November, 2016

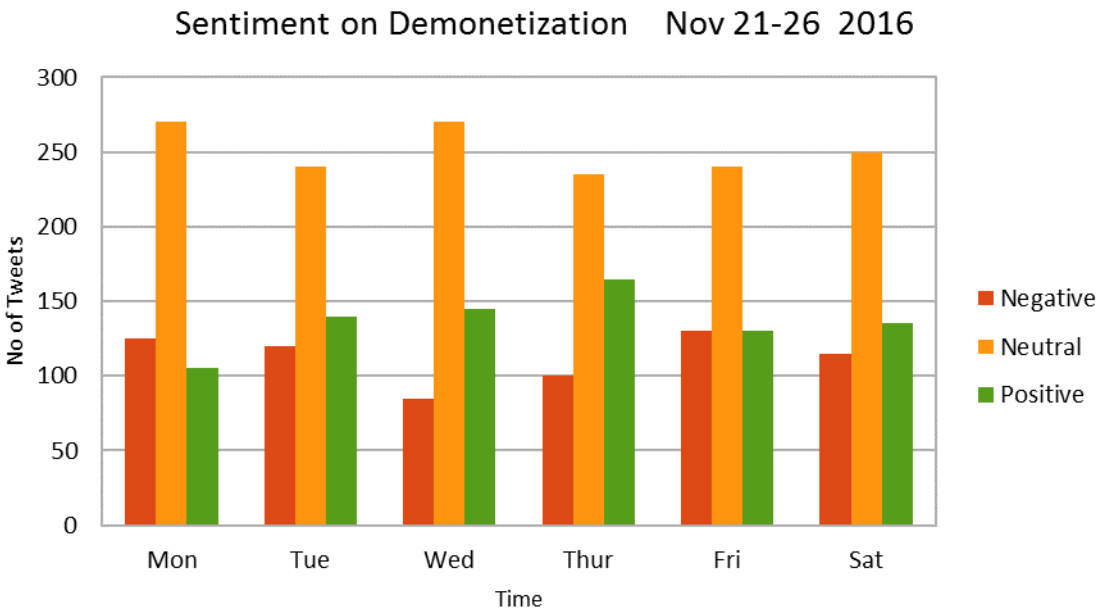
Result from data gathered from Nov 21-26 2016

	Negative	Neutral	Positive
Mon	125	270	105
Tue	120	240	140
Wed	85	270	145
Thur	100	235	165
Fri	130	240	130
Sat	115	250	135
Average	112.5	250.83	136.67
Variance	239.58	203.47	322.22
Deviation	16.96	15.63	19.66

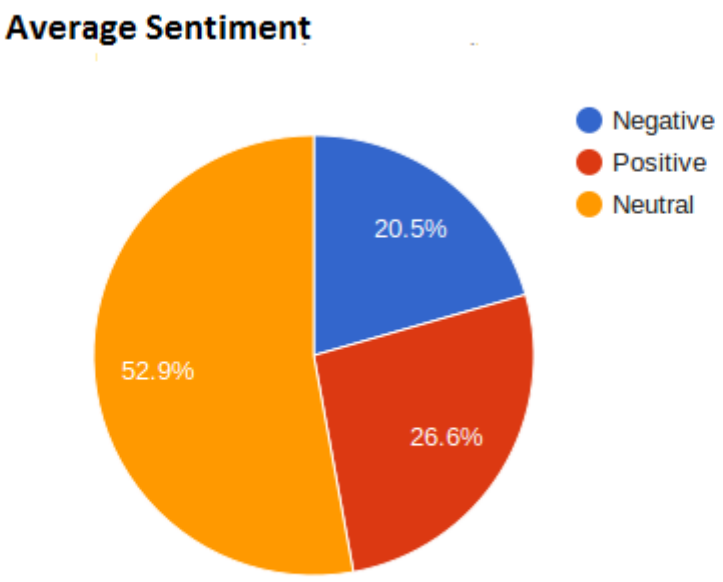
Table 3: Average, variance and deviation of negative, positive and neutral tweets.



Graph depicting the trend of negative, positive and neutral tweets. It is evident that number of positive tweets is approximately always greater than that of negative.



Graph depicting the amount of negative, positive and neutral tweets. Number of positive tweets is approximately always greater than that of negative.

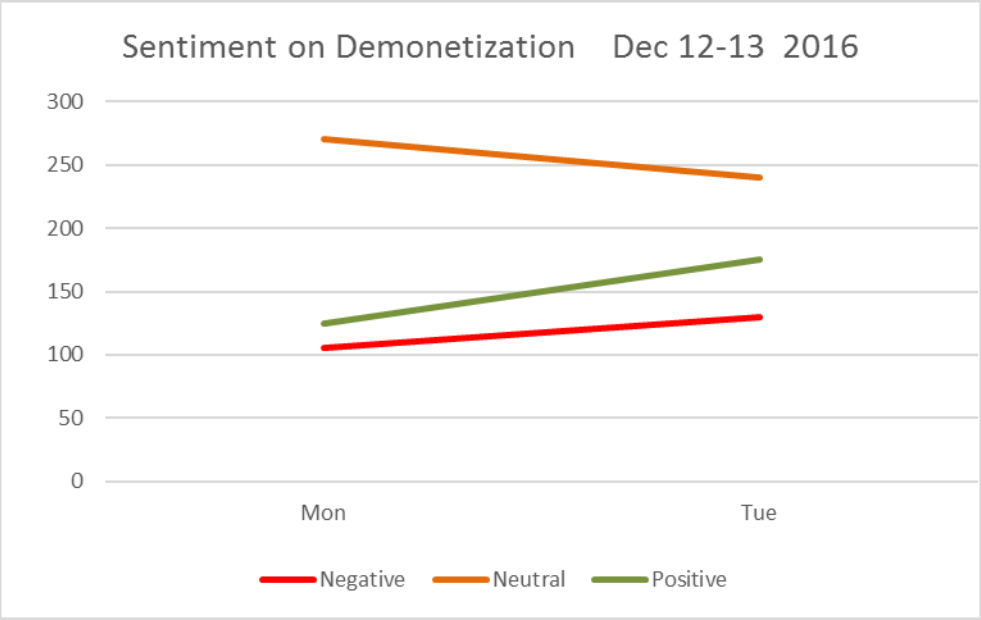


Pie Chart depicting the percentage of positive, negative and neutral tweets. It can be inferred that approximately half of the people have neutral sentiment regarding demonetisation and only 20% expressing negative opinion in tweets.

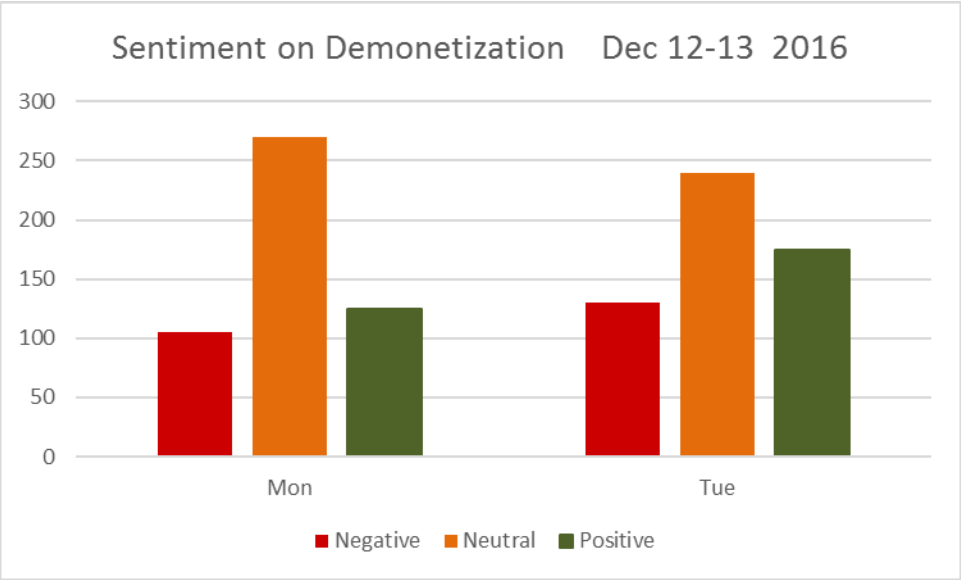
4.2.2 Result Set for December, 2016

Result from data gathered from Dec 12-13 2016

	Negative	Neutral	Positive
Mon	105	270	125
Tue	130	240	175
Average	117.5	255	150
Variance	312.50	450.00	1250.00
Deviation	17.68	21.21	35.36



Graph depicting the trend of negative, positive and neutral tweets. It is evident that number of positive tweets is approximately always greater than that of negative.



Graph depicting the trend of negative, positive and neutral tweets. It is evident that number of positive tweets is approximately always greater than that of negative.

Chapter-5
Conclusions

An algorithm for analysing text and data processing system are implemented to analyse the opinions of users from huge volume of data (text) generated by Twitter. The proposed method is composed of a parallel HDFS system based on the Hadoop ecosystem and on MapReduce functions. The method proposed successfully processed the data and scaled well as the number of data items increased. The whole load of the system is distributed across nodes by parallel processing. The results of sentiment analysis with the system are quite close to the results of manually classifying the sentiments. The system reported 9.2% increase in the number of positive tweets in the month of December as compared to that of November which is explainable as the situation now has become quite stable because people are facing less difficulty.

5.2 Future Scope:

The proposed system can also be used in fetching emotions as Twitter posed 140 characters limit for a person to expresses his/her emotions thus people tend to use emotions to express their feeling thus in order to achieve better result this can pose a challenge to achieve that. Therefore using emotion classifier can help in achieving higher accuracy in results.

5.3 Applications:

The following can be used as tool in various fields such as,

1. Opinion polls for elections.
2. Product reviews and feedback by companies and businesses.

   In order to understand customer urge and need and creating business solution accordingly using the analysis result.

References

[1] Khan, Nawsher, et al. "Big data: survey, technologies, opportunities, and challenges." *The Scientific World Journal* 2014 (2014).

[2] Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann. "Recognizing contextual polarity in phrase-level sentiment analysis." *Proceedings of the conference on human language technology and empirical methods in natural language processing*. Association for Computational Linguistics, 2005.

[3] Ha, Ilkyu, Bonghyun Back, and Byoungchul Ahn. "MapReduce functions to analyze sentiment information from social big data." *International Journal of Distributed Sensor Networks* 2015 (2015): 5.

[4] Go, Alec, Richa Bhayani, and Lei Huang. "Twitter sentiment classification using distant supervision." *CS224N Project Report, Stanford* 1 (2009): 12.

[5] Pak, Alexander, and Patrick Paroubek. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining." *LREc*. Vol. 10. 2010.

[6] Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann. "Recognizing contextual polarity in phrase-level sentiment analysis." *Proceedings of the conference on human language technology and empirical methods in natural language processing*. Association for Computational Linguistics, 2005.

[7] Baccianella, Stefano, Andrea Esuli, and Fabrizio Sebastiani. "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining." *LREC*. Vol. 10. 2010..

[8] Turney, Peter D. "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews." *Proceedings of the 40th annual meeting on association for computational linguistics*. Association for Computational Linguistics, 2002.

[9] "Tweets — Twitter Developers." Twitter. Twitter, n.d. Web. 25 Apr. 2017. <https://dev.twitter.com/overview/api/tweets>.

[10] P. D. Turney, “Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews,” in Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL '02), pp. 417–424, Association for Computational Linguistics, Stroudsburg, Pa, USA, July 2002.

[11] P. D. Turney and M. L. Littman, “Measuring praise and criticism: Inference of semantic orientation from association,” ACM Transactions on Information Systems, vol. 21, no. 4, pp. 315–346, 2003.

[12] A.Pakand P.Paroubek,“Twitter a sa corpus for sentiment analysis and opinion mining,” Proceedings of the 7th Conference on International Language Resources and Evaluation (LREC ’10),2010.

[13] S. Mukherjee and P. Bhattacharyya, “Sentiment analysis in twitter with lightweight discourse analysis,” in Proceedings of the 24th International Conference on Computational Linguistics (COLING ’12), pp. 1847–1864, December 2012.

[14] A. Go, R. Bhayani, and L. Huang, “Twitter sentiment classification using distant supervision,” Project Report, Stanford University, 2009.

[15] L. Barbosa and J. Feng, “Robust sentiment detection on twitter from biased and noisy data,” in Proceedings of the 23rd International Conference on Computational Linguistics (Coling ’10), pp.36–44, August 2010.

[16] V. N. Khuc, C. Shivade, R. Ramnath, and J. Ramanathan,“Towards building large-scale distributed systems for Twitter sentiment analysis,” in Proceedings of the 27th Annual ACM Symposium on Applied Computing (SAC ’12), pp. 459–464, March 2012.

[17] Yi, J., T. Nasukawa, R. Bunescu, and W. Niblack: 2003, “Sentiment Analyser: Extracting Sentiments about a Given Topic using Natural Language Processing Techniques”, In:Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM-2003).MelbourneFlorida.

Appendices

Source Code
SentimentsMapper.java

```
public class SentimentsMapper extends
                Mapper<LongWritable, Text, IntWritable,Text>{

        static ArrayList<String> positiveWords = new ArrayList<String>();
        static ArrayList<String> negativeWords = new ArrayList<String>();

        @Override
        protected void map(LongWritable key, Text value, Mapper<LongWritable, Text,
IntWritable,Text>.Context context)
                        throws IOException, InterruptedException {


                //String uriStr =  "s3n://sentimentproj/input/";
                //URI uri = URI.create(uriStr);
                //FileSystem fs = FileSystem.get(uri, context.getConfiguration());
                //Path pt1 = new Path(context.getConfiguration().get("s3n://sentimentproj/input/positive-
words.txt"));
                //Path pt2 = new Path(context.getConfiguration().get("s3n://sentimentproj/input/negative-
words.txt"));


                Path pt1=new Path("hdfs://localhost:54310/dictionary/op/positive-words.txt");
                Path pt2=new Path("hdfs://localhost:54310/dictionary/op/negative-words.txt");

                Configuration conf = new Configuration();

                conf.addResource(new Path("/usr/local/hadoop/conf/core-site.xml"));
                conf.set("fs.defaultFS", "hdfs://localhost:54310");
                conf.set("mapreduce.jobtracker.address", "hdfs://localhost:54310");

                FileSystem fs = FileSystem.get(conf);

    BufferedReader positiveReader = new BufferedReader(new InputStreamReader(fs.open(pt1)));
    BufferedReader negativeReader = new BufferedReader(new InputStreamReader(fs.open(pt2)));
    String word;


                while ((word = negativeReader.readLine()) != null) {
                        negativeWords.add(word);
                }
                while ((word = positiveReader.readLine()) != null) {
                        positiveWords.add(word);
                }

                negativeReader.close();
                positiveReader.close();

                String tweet=null;
                String location = "";
                int month = -1;
```

```
		int year = -1;
		int score = -1;
		JSONParser parser = new JSONParser();
		JSONObject jsonObject = null;
		try {
			jsonObject = (JSONObject) parser.parse(value.toString());
			if(jsonObject != null){
				tweet = jsonObject.get("text").toString();
				score = getSentimentScore(tweet);
				String date = jsonObject.get("created_at").toString();
				JSONObject user = (JSONObject) jsonObject.get("user");
				if(user.get("location")!=null)
					location = user.get("location").toString();
				Date date1 = getDate(date);
				if(date1 != null){
					Calendar cal = Calendar.getInstance();
					cal.setTime(date1);
					month = cal.get(Calendar.MONTH);
					year = cal.get(Calendar.YEAR);
				}
				//jsonObject.
		}
		} catch (ParseException e) {
			System.err.println("ERROR PARSING JSON OBJ");
			e.printStackTrace();
		}
		tweet = tweet.replaceAll(",", "");
		tweet = tweet.replaceAll("\n", "");
		String opVal = tweet + "," + "\"" + location + "\"";
		context.write(new IntWritable(score), new Text(opVal));
	}


	private  int getSentimentScore(String ip) {
		// normalize!
		ip = ip.toLowerCase();
		ip = ip.trim();
		// remove all non alpha-numeric non whitespace chars
		ip = ip.replaceAll("[^a-zA-Z0-9\\s]", "");

		int ncount = 0;
		int pcount = 0;

		// so what we got?
		String[] words = ip.split(" ");

		// check if the current word appears in our reference lists...
		for (int i = 0; i < words.length; i++) {
			if (positiveWords.contains(words[i])) {
				pcount++;
			}
			if (negativeWords.contains(words[i])) {
				ncount++;
			}
		}

		// positive matches MINUS negative matches
		int result = (pcount - ncount);

		// negative?
```

```
                if (result < 0) {
                        return -1;
                        // or positive?
                } else if (result > 0) {
                        return 1;
                }

                // neutral to the rescue!
                return 0;

        }


        public  Date getDate(String date)  {

                final String TWIT="EEE MMM dd HH:mm:ss ZZZZZ yyyy";
                SimpleDateFormat simpleformat = new SimpleDateFormat(TWIT);
                simpleformat.setLenient(true);
                Date tmp1=null;
                try {
                        tmp1 = simpleformat.parse(date);
                } catch (java.text.ParseException e1) {
                        // TODO Auto-generated catch block
                        e1.printStackTrace();
                }
                return tmp1;
        }
}       }
}
```

SentimentsReducer.java

```
public class SentimentsReducer
extends Reducer<IntWritable, Text, IntWritable, Text> {
@Override
public void reduce(IntWritable key, Iterable<Text> values,
 Context context)
 throws IOException, InterruptedException {
        for (Text txt : values) {
                context.write(key, txt);
        }

}
}
```

SentimentsDriver.java

```
public class SentimentsDriver extends Configured implements Tool{
public int run(String[] args) throws Exception{
        Configuration conf = new Configuration();
   String[] otherArgs = new GenericOptionsParser(conf, args).getRemainingArgs();
   if (otherArgs.length != 2) {
```

```
        System.err.println("Usage: Driver <in> <out>");
        System.exit(2);
    }
    conf.set("mapred.textoutputformat.separator", ",");
    Path in = new Path(otherArgs[0]);
    Path out = new Path(otherArgs[1]);
    Job job=  Job.getInstance(conf);
  job.setJarByClass(SentimentsDriver.class);
  job.setJobName("SentiAnalysis");
  FileInputFormat.addInputPath(job, in);
  FileOutputFormat.setOutputPath(job,out);
  job.setMapperClass(SentimentsMapper.class);
  job.setReducerClass(SentimentsReducer.class);
  job.setOutputKeyClass(IntWritable.class);
  job.setOutputValueClass(Text.class);
  System.exit(job.waitForCompletion(true) ? 0:1);
  boolean success = job.waitForCompletion(true);
  return success ? 0 : 1;
  }
 public static void main(String[] args) throws Exception {
  SentimentsDriver driver = new SentimentsDriver();
  int exitCode = ToolRunner.run(driver, args);
  System.exit(exitCode);
  }
 }
```