

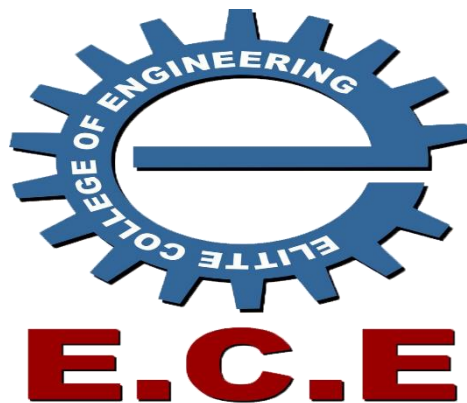
PROJECT REPORT

ON

Sentiment Analyzing of Buzzword B.1.1.529 using Twitter Data

Submitted in the partial fulfillment of the requirement for the award of degree of

BACHELOR OF TECHNOLOGY
IN
ELECTRICAL ENGINEERING



UNDER THE GUIDANCE OF

Dr. Sandip Roy
HOD and Associate Professor,
Department of Computational Science
Brainware University

Submitted by

Name: SRIJAN CHAKRABORTY
University Roll Number: 34601619046
Department of Electrical Engineering
ELITTE COLLEGE OF ENGINEERING



Maulana Abul Kalam Azad University of Technology
Kolkata, West Bengal, India

DECLARATION

I hereby declare that this project report is based on my original work except for citations and quotations which have been duly acknowledged. I also declare that it has not been previously and concurrently submitted for any other degree or award at any University or any other Institution.

SRIJAN CHAKRABORTY
University Roll Number: 34601619046
ELITTE COLLEGE OF ENGINEERING
Maulana Abul Kalam Azad University of Technology

APTECH COMPUTER EDUCATION
121, FEEDER ROAD PIONEER PLAZA, BESIDE SAKAL SANDHYA SHOP, BELGHORIA
KOLKATA WESTBENGAL 700056



CERTIFICATE

This is to certify that this project entitled “ANALYZING OMICRON SENTIMENT USING TWITTER DATA” submitted by SRIJAN CHAKRABORTY (34601619046), student of ELECTRICAL ENGINEERING Department, ELITTE COLLEGE OF ENGINEERING, SODEPUR, West Bengal in the partial fulfilment of the requirement for the award of BACHELOR OF TECHNOLOGY of Maulana Abul Kalam Azad University of Technology, West Bengal, is a record of students’ own study carried under my supervision and guidance. This report has not been submitted to any other University or Institution for the award of any degree.

Dr. Sandip Roy

HOD and Associate professor
Department of Computer Science
Brainware University

Prasun Krishna Dutta

Director of Aptech Belgharia

ACKNOWLEDGEMENT

To the Almighty, I genuflect before thee.

This thesis is dedicated to my mother, Mrs. Sumita Chakraborty, my father, Mr. Swapan Chakraborty whose influences have been at the core and heart of my efforts at bringing my research to fruition. I would like to express my profound sense of acknowledgement and gratitude to my supervisor, for his unflinching support and faith in my ability to conduct research in this field. Without his aid and support this work would not have been possible at all.

I am indebted to Dr. Sandip Roy Department of Computational Science, Brainware Group of Institutions - Sabita Devi Education Trust, Brainware University, Barasat, West Bengal. I thank Department of Electrical Engineering, Elite College of Engineering, Sodpur, West Bengal for extending invaluable support to help conduct my research activities.

However, it would be incomplete without pausing and reflecting on support that I have had from every possible quarter in my professional and private life. I remain especially and forever indebted to Dr. Sandip Roy. It is with his influence and technical guidance that I found my source of sustenance and inspiration to forge ahead with my work. None of this would perhaps been possible in the manner that I have managed to do, had it not been for his presence and for his ability to lift my spirits and hopes when all seemed so distant and insurmountable.

And finally, a few words for Mr. Sandip Roy - a person who has, in a short span of time, wavered not for once to extend his help, support and cooperation when the time came to translate my thoughts and ideas into words.

Srijan Chakraborty
University Roll Number: 3460161904

ABSTRACT

Today microblogging has become a very common platform for exchanging opinion among us. Many users exchange their thoughts on a various aspect of their activity. Consequently, microblogging websites are the substantial origin of information for sentiment analysis and opinion mining. Twitter is a famous microblogging website where 500 million tweets are posted every day. Corona Virus or COVID-19 first appeared in December, 2019 in Wuhan, China. People tweeted aggressively on twitter at that time. All tweets are categorized into 3 categories (Positive, NegativeandNeutral). On 26 November 2021, WHO designated the variant B.1.1.5 29 a variant of concern (VOC), following advice fromthe WHO's TechnicalAdvisory Group on Virus Evolution. This thesis, we summarize the data set of Twitter messages related to recent Omicron 2021 for predicting the chances of opinions about the Omicron variant to better understand many aspects. We use NRC Emotion Lexicon to determine the overall tone of the event by eight emotions. Furthermore, we use a Deep Learning tool named, ParallelDots AI APIs by ParallelDots Inc that can analyze the sentiment into positive, negative and neutral. This tool can be helped to extract various peoples' sentiment and summarize the results for further decision making.

Table of Contents

1. Introduction	7
2. Literature Review	7
2.1 Traditional Sentimental Analysis	8
2.2 Buzzword Sentimental Analysis	9
3. Data Preparation	9
3.1 Twitter Data Analysis	10
3..2 Data Processing	10
4. Experimental Setup and Result Analysis	11
5. Conclusion and Future Scope	11
6. Declaration	11
Appendices	12
Appendix 1: How to extract tweets using R language	14
Appendix 2: How to extract tweets using Twitter Archiver Tool	16
Appendix 3: How to preprocess tweets using R language	17
Appendix 4: How to analyze emotion using syuzhet in CRAN	
Package	18

1. Introduction

As because of the massive growth of user-created data in the recent WWW websites, people from various backgrounds tweet massive amount of textual remarks deliberating their thoughts in a different perspective of their emotions and public to everyone. Natural Language Processing (NLP) can be categorized into opinion and text mining. This technique is helped for isolating the opinions of posting on various social media platforms like Twitter, Reddit, and Facebook etc. In today's world text or opinion, mining is helpful for judging public views regarding a newly released item, movie, song, book, etc. It also differentiated among positive, negative and neutral opinion and recommendations. It also becomes a common practice for the common people to pose their expressions towards the Omicron on the social media. Different reporters have been taking an interview with the WHO and Doctors to know their views and communicate with the people through TV program, YouTube, etc. People express their opinions with each other regarding the virus which ran on the TV show. It is very expensive and time-consuming task to search people's opinions via surveys and polls. Now various social media sites (like Twitter, Reddit, Facebook, etc.) are extensively used by the public when they can share their opinions publicly. One of important microblogging site receiving around 500 million tweets every day where the daily limit of each user is 2,400 tweets and 140 characters every tweet. Hence, Twitter is one of the relevant platforms where each people can connect with different communities and express their opinions loud and clear. Sentiment analysis is a type of data mining process that determines the public opinion through NLP. It is the process of classifying opinions into three categories like "positive", or "negative" or "neutral". This data quantifies the public's reactions toward certain people, communities which divulge the contextual polarity of the information.

Many research already organized on Omicron sentiment analysis on Twitter data. The main aim of the research is to predict the cases based on people's opinions. Popular debate TV show between Dr. Kunal Sarkar, Pradip Mitra and Atria Biswas called "Omicron Concern". In this TV show, the audience freely posted comments on Twitter. It has become an admissible communication platform in the health area on basis of Omicron. Hence, health sentiment analysis is a common practice for analyzing the people's opinion towards the health campaign. Due to the second-most populous country and the cases in the world, India health situation is most fluctuating. Every step of the Doctors and WHO would have several views of the oppositions. But today common people post their opinions on the social media regarding every health step (like case of all India as of 21 December 2021 reported 200 cases of Omicron). Therefore our aim is to analyze the emotion of web users concerning every health conversion of omicron, their doctors and their steps based on tweets on the social media. The detailed sections of the thesis are as follows. Section 2 is described literature Survey. Section 3 deals with on data preparation along with our corpus analysis. Experimental setup and result analysis part are discussed in section 4 and the section 5 draws conclusions and examines the possibilities of next future.

2.Literature Review

Literature survey can be categorized into two different subsections, Traditional sentiment analysis and examine related to the people's political judgment.

2.1 Traditional Sentiment Analysis

There are various research papers published on sentiment analysis.

Tumey's unsupervised learning algorithm examined Pointwise Mutual Information (PMI) for estimating the opinion inclination of phrases. In opposed to it, Pang, Lee, and Vaithyanathan suggested that three machine learning methods such that, Support Vector Machines, Naive Bayes, and Maximum Entropy Classification don't function as well on the classification of sentiment or on traditional topic-based segregation.

Chesley et al. presented that how a particular blog expressed subjectivity vs. objectivity and also how a post expressed positive or negative polarity. Godbole et al. researched on recent events reported in newspapers and blogs which expressed opinions of news entities (people, places and things). Their assigning scores representing sentiment to each separate entity in the lexicon. Early research contributed to a longer document such as movie review and blogs.

Birmingham et al. presented in his research on microblogs where they examined the hypothesis that easily classifies the opinion in small-scale documents than in larger scale documents. They experimented with the microblogs from Twitter, blog posts, micro-reviews from blippr and movie reviews and culminated that it is easier to identify sentiment from microblogs. There has been some research on sentiment analysis on Twitter. Pak and Paoubek suggested that microblogging has become a popular tool for expressing opinions of people. Their sentiment classifier can determine the sentiments of a given text. Davidov also examined on data from Twitter. Their supervised sentiment classification framework performed on fifteen smileys as sentiment labels and fifty Twitter tags which identified and classified of various sentiment types of short texts.

Bakliwal et al. focused on the task of tweet sentiment identification using a corpus of pre-annotated tweets. They have used unigram and bigram features and presented a sentiment scoring function which boosts the classification accuracy.

2.2 Buzzword Sentiment Analysis:

On-demand interest in the online Omicron sentiment analysis is to predict the result of the country.

Few days ago, the WHO designated a new variant of the coronavirus, B.1.1.529, as a variant of concern which has been named Omicron. Right after that, we saw an outbreak of tweets about the Omicron variant on Twitter. we can analyze the sentiments of the tweets about the Omicron variant. As of 22 December 2021, the Omicron variant had been identified in 110 countries across all six WHO Regions. Current understanding of the Omicron variant continues to evolve as more data become available. This technical brief provides updated early evidence on key aspects of Omicron related to transmission, severity and impact on interventions such as diagnostics, therapeutics and vaccines, and outlines a set of priority actions for Member States. How populations understand these dynamics, perceive risk and follow control measures, including public health and social measures (PHSM). Public health advice is based on current information and will be tailored as more evidence emerges around this. Data on clinical severity of patients infected with Omicron is growing but is still limited. Early data from South Africa, the United Kingdom (UK) and Denmark suggest a reduced risk of hospitalization for Omicron compared to Delta. However, the risk of hospitalization is only

one aspect of severity, which may be altered by admission practices. More data across different countries are needed to understand how clinical markers of severity—such as the use of oxygen, mechanical ventilation and deaths—are associated with Omicron.

In general state condition in 2021, we researched on Twitter for predicting of Indian state Omicron case condition. we have used some tweets using Twitter Archiver tool to extract tweets in English language and examined by Support Vector Machine (SVM), Dictionary Based and Naive Bayes algorithm which classified into three categories like “positive”, or “negative” or “neutral”. The objective of our research is to analyze peoples’ sentiment for determining the results of Thinking in Omicron cases using deep learning method tools available as web service by ParallelDots Inc.

3. Data Preparation

The data has been collected during December 19, 2021 using Twitter’s streaming API (See Appedix). We have collected around 100 tweets mentioning two verified Twitter accounts named @ptcnews and @ndtv respectively may other web side and other gerenal.

3.1 Twitter Data Analysis

We performed the Emotion Analysis using syuzhet in CRAN package which is based on NRC Emotion Lexicon. We explore the time series analysis of tweets over given time period and we visualize that most of the tweets are from Twitter for Android followed by Twitter for iPhone and Twitter Web Client source which is depicted on figure 1.

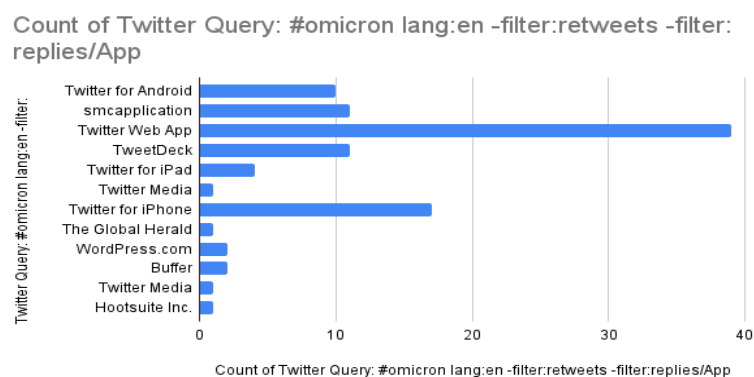


Figure 1. Number of Tweets extracted from different sources

We also explore that on which time of a day maximum retweets is contributed from which Twitter account in our corpus is shown in figure 2.

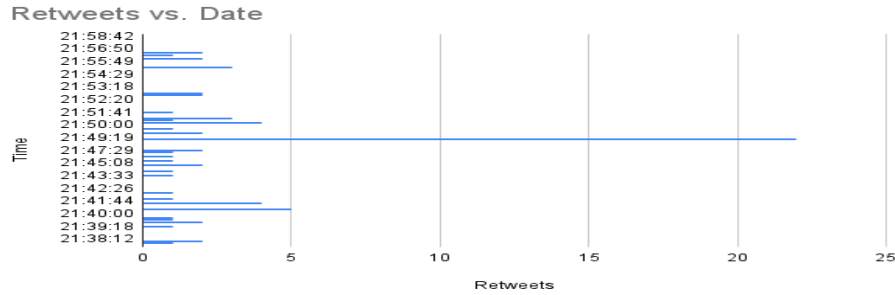


Figure 2. Number of Retweets vs. Time

3.2 Data Processing

Before applying any of the sentiment or emotion analyzing method, we perform data preprocessing. Data preprocessing is a major step in text mining that produces a higher quality of text classification and reducing computational complexity. The following points have mentioned as data preprocessing steps of the tweet.

- Remove all hashtags (e.g. #topic), screen name (e.g. @username) and all URLs (e.g. www.xyz.com)
- Remove all punctuations, symbols, and numbers
- Replace any non UTF-8 by space
- Remove Stopwords
- Substitute all the emoticons with their sentiment
- Change text to lowercase
- Replace words with their stems or roots

4.Experimental Setup and Result Analysis

We performed the Emotion Analysis using syuzhet in CRAN package which is based on NRC Emotion Lexicon. In Figure 3 show the distribution of emotion of Omicron during the pandemic situation in December19,2021.

In Figure 3 more than 60 tweets express positive sentiment and around 60 tweets indicate negative sentiment whereas in. In figure 3, more than 40 tweets express trust as an emotion. As can be seen, the disgust emotion in figure 3 is least emotion score. Similarly, joy emotion is placed better in figure 3.

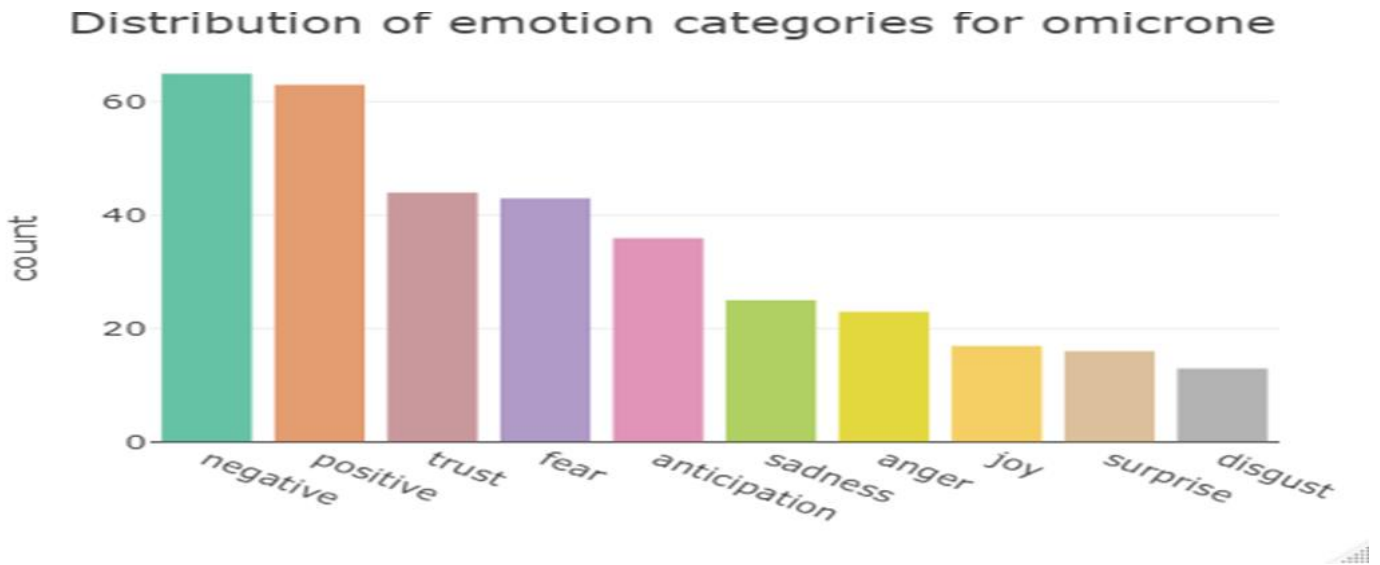


Figure 4. Emotion distribution of Omicron December19,2021

5. Conclusion and Future Scope

Only 12.18% of people of the total population in India have communicated using English language and most of the common people express their opinions using native language. Therefore, analyzing sentiment using the only English language isn't worthy of every time. Our research will be more robust when we will implement the lexicon of different native languages used in India. Furthermore, we will also extend our research using Latent Dirichlet Allocation (LDA) based topic model and also identifies the correlated topic across a different category for the event.

6. Declaration

The author declares and solemnly affirm that this research has neither been funded by any political or religious groups nor are the authors in any way affiliated to any institutions with direct or indirect access to groups with biased interests. This research work has been carried out exclusively and independently by the authors in the interests of technology and progress of science in sentiment analysis and related fields.

Appendices

Appendix 1: How to extract tweets using R language

Prerequisites

1. You have already installed R and are using RStudio or install R using <https://cran.r-project.org/bin/windows/base/>.
2. In order to extract tweets, you will need a Twitter application and hence a Twitter account. If you don't have a Twitter account, sign up using <https://twitter.com/signup>.
3. Use your Twitter login ID and password to sign in at Twitter Developers using <https://apps.twitter.com/>.

Steps for creating Twitter Apps

1. Create a new application

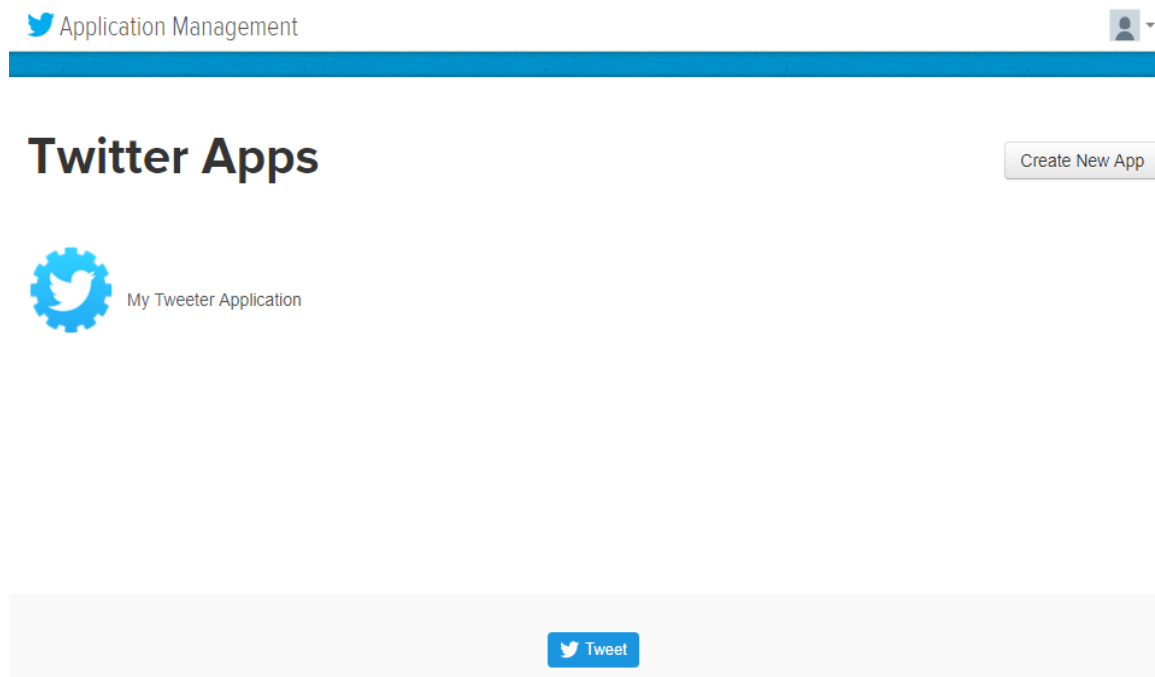


Figure A.1.1 Application Management Menu of Twitter

2. Fill out the new app form. Names should be unique, i.e., no one else should have used this name for their Twitter app. Give a brief description of the app. then create the new app then you got api key and api secret and bearer token and access token and secret. then you go to setting you got **authentication settings**. in settings you got **enable 3-legged oauth**. on all of this. fill out the new app form. names should be unique, i.e., no one else should have used this name for their twitter app. give a brief description of the app. you can change this later on if needed. enter your website or blog address. in callback urls you fill up in the google spread sheet url or google url and the website url is your twitter home page and terms of service and privacy policy is twitter terms of service and privacy policy. fill up the box in this order you got save. go to the access level choose the read, write and direct messages.

Enable 3-legged OAuth ☒

Use 3-legged OAuth for Sign in with Twitter, posting Tweets on behalf of other accounts and more. Get more information in the docs.

Request email address from users ☒

These additional permissions require that you provide URLs to your application or service's privacy policy and terms of service.

Callback URLs (required)

<https://docs.google.com/spreadsheets/d/11LG2L12NfTg6z7p5ClaccaRp5JscMuzRzAyA1OWX5g/edit#gid=0>

[+ Add another](#)

Website URL (required)

<https://twitter.com/home>

Organization name (optional)

Organization URL (optional)

<https://>

Terms of service (required)

<https://twitter.com/en/tos>

Privacy policy (required)

<https://twitter.com/en/privacy>

[Cancel](#) [Save](#)

[Privacy](#) [Cookies](#) [Twitter Terms & Conditions](#) [Developer Policy & Terms](#) [© 2021 Twitter](#)

Figure A.1.2 CREATE ENABLE 3-LEGGED OAUTH

3. Scroll down and click on “Create my access token” button. Note the values of consumer key and consumer secret and keep them secure for future use. If anyone was to get these keys, they could effectively access your Twitter account.

Application Settings

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

Consumer Key (API Key)	My API Key
Consumer Secret (API Secret)	My API Secret
Access Level	Read, write, and direct messages (modify app permissions)
Owner	My Username
Owner ID	My UserID

Figure A.1.3 Application Settings Menu of Twitter

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	My Access Token
Access Token Secret	My Access Token Secret
Access Level	Read, write, and direct messages
Owner	My Username
Owner ID	My UserID

Figure A.1.4 Access Token Menu of Twitter

4. Run the below script on RGUI.

```
install.packages("twitterR")
install.packages("RCurl")
require(twitterR)
require(RCurl)
# User API Key
api_key □ "XX"
# User API Secret
api_secret □ "XX"
"# User Access Token
access_token □ "XX"
# User Access Token Secret
access_token_secret □ "XX"
setup_twitter_oauth(api_key, api_secret, access_token, access_token_secret)
#Search BJP tweets for an example
mytweet □ searchTwitter("#hashtag", n = 10, lang = "en")
```

Appendix 2: How to extract tweets using Twitter Archiver Tool

Note: It was collected using Google Spreadsheet which built the connection to Twitter using Google script by finding the key details from a Twitter account and authorize the applications. After that it can automatically import all the users' recent tweets into the Spreadsheet. The below steps are required to connect for extracting tweets using Twitter Archiver Tool.

Step 1: Go to Google drive and click on the “New” tab to open the Blank spreadsheet.

Step 2: Go to the Add-ons menu, choose Twitter Archiver Tool and select the Authorize menu.

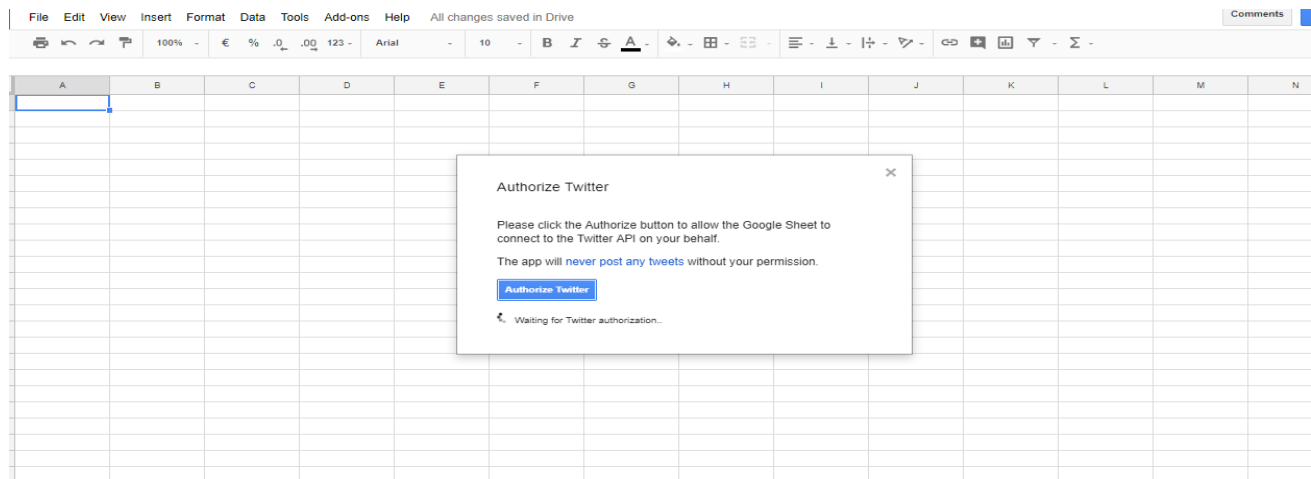


Figure A.2.1 Authorize Twitter menu to allow Google Sheet to connect to Twitter API

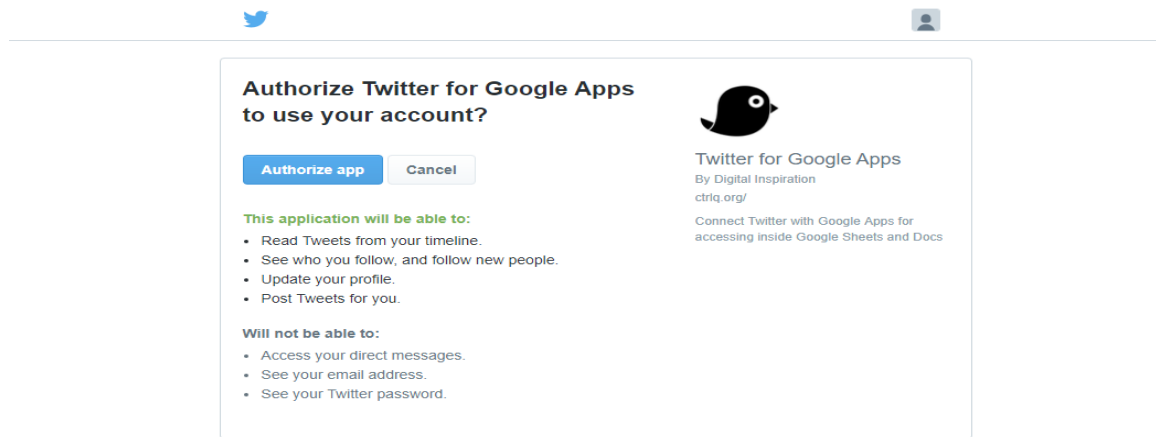


Figure A.2.2 Authorize Twitter for Google Apps to use your Twitter account

Step 3: Allow the Google Sheets to authorize Twitter on your behalf.

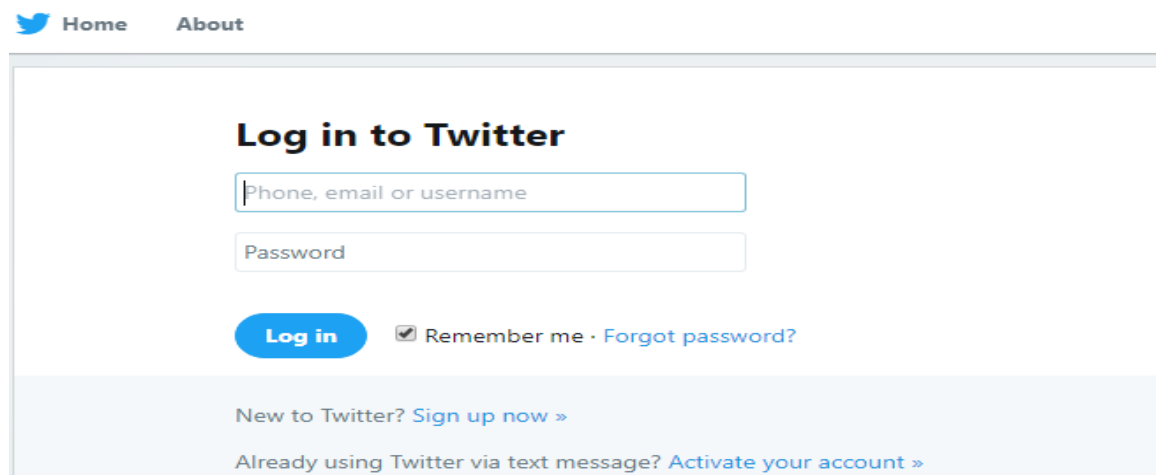


Figure A.2.3 Log-in to twitter account using username and password

Your Google account is now authorized to use Twitter. Please close this tab and create a new Twitter Rule inside the Google Sheet.

Figure A.2.4 Screen display after successful authentication

Step 4: After successful authentication, go to the Twitter Archiver menu again and create a new Search Rule.

Note: Steps for extract tweets using R language (See Appendix 1)

✕

Create a Twitter Search Rule

All of these words

This exact phrase

Any of these words

None of these words

These #hashtags

Written in

Any Language ▾

Near This Place

Advanced Rules

People

To these accounts

Mentioning accounts

From these accounts

Enter the [search criteria](#) above and matching tweets will be saved in the Google sheet.

Create Search Rule

Upgrade to Premium

Cancel

Figure A.2.5 Twitter Search Rule window

2	Date	Screen Name	Full Name	Tweet Text	Tweet ID	App	Followers	Follows
3	11/9/2017 9:50:29	@sanjayezhava	Sanjay Ezhava IN	@PiyushGoyal @RailMinIndia @drmmumbaicr Seems Liq	928477363783577088	Twitter for iPhone	11293	1837
4	11/9/2017 10:33:11	@ramanmann1974	RamanS Mann	#SEBI charges Gujarat Chief Minister @vijayrupanibjp's H	928488115957743616	Twitter Web Client	2768	72
5	11/9/2017 11:07:41	@RuchiraC	Ruchira Chaturvedi	SEBI penalises Gujarat CM @vijayrupanibjp owned comp	928496796271239168	Twitter for Android	3974	433
6	11/9/2017 12:06:03	@yogen198	Yogen Bhawar	@BJP4India @AmitShah @DrGPradhan @INCIndia @INC	928511485168443392	Twitter for Android	3025	1883
7	11/9/2017 12:42:51	@dangarbhara	Bharat Dangar	#Vadodara welcoming Honourable Chief Minister @vijay	928520746661326850	Twitter for Android	33115	494
8	11/9/2017 13:18:53	@das_prashant777	Dr Prashant Das	@DhartiMajithiya @PoonambenMaadam @narendramod	928529814230675456	Twitter for Android	2688	1095
9	11/9/2017 14:26:49	@sriramibjp	Sriram	Muslims turn up in large numbers at Rupani rally. @vijay	928546911530840064	Twitter for iPhone	24645	1516
10	11/9/2017 15:30:35	@Barmer Harish	Harish Chaudhary	Another case of corruption, another malicious BJP man, a	928562955641274371	TwitterDeck	4544	78
11	11/9/2017 16:02:57	@GauravPandhi	Gaurav Pandhi	What a fraud !! Gujarat CM @vijayrupanibjp has been per	928571101529104384	Twitter for Android	49700	658
12	11/9/2017 16:20:24	@rachitseth	Rachit Seth	Congress party demands the resignation of Gujarat CM @	928575491908321281	Twitter for iPhone	41059	2690
13	11/9/2017 16:45:41	@dangarbhara	Bharat Dangar	Glimpses of today's #Gujarat_Gaurav_Mahasampark_Abh	928581856630415360	Twitter for Android	33115	494
14	11/9/2017 16:49:42	@Trolmodji	Narendra Godi	@vijayrupanibjp Vijuuu now u r finally my boy.... Good jo	928582866014842880	Twitter for Android	571	43
15	11/9/2017 17:47:43	@VPras52	Vinay Prasad	SEBI has imposed total penalties worth Rs 6.9 crores on	928597465829335040	Twitter for Android	5124	3062
16	11/9/2017 18:14:22	@pvssarma	pvs sarma	There's no leader in @BJP4Gujarat against who allegation	928604176027807745	Twitter for Android	4565	239
17	11/9/2017 18:32:04	@sureshnakhua	Suresh Nakhua	So @bsindia does a hit job on @vijayrupanibjp . Does a st	928608627341934597	Twitter for Android	70651	2770

Figure A.2.6 Import tweets to the Google Spreadsheet

Appendix 3: How to preprocess tweets using R language

1. Download the search tweets from Google drive which are extracted by Twitter Archiver Tool.
2. Run the below script on RGUI

```
s <- read.csv("file name")
x <- list(s[,4])
# Remove at people mentioned preceding by '#' or '@'
x = gsub("#\\w+", "", s[,4])
x = gsub("@\\w+", "", x)
# Replace any non UTF-8 by space
x=iconv(x, "UTF-8", "UTF-8",sub="")
# Remove punctuation
x = gsub("[[:punct:]]", "", x)
#Remove alphanumeric character
x = gsub("[^[:alnum:]]", "", x)
# Remove numbers
```



```

x = gsub("[[:digit:]]", "", x)
# Remove html links
x = gsub("http\\w+", "", x)
# Remove unnecessary spaces
x = gsub("[ \\t]{2,}", "", x)
x = gsub("^\\s+|\\s+$", "", x)
#Convert text to tolower
x=tolower(x)
# Write x to csv file
write.csv(x, "output file name")

```

Appendix 4: How to analyze emotion using syuzhet in CRAN package

1. Download the search tweets from Google drive which are extracted by Twitter Archiver Tool.
2. Run the below script on RGUI

```

library(syuzhet)
library(plotly)
library(tm)
library(wordcloud)
# Read the file name
s <- read.csv("file name")
Timestamp <- list(s[,1])
Account <- list(s[,2])
Tweets <- list(s[,4])
Retweet <- list(s[,9])
Tweets = gsub("#\\w+", "", s[,4])
Tweets = gsub("@\\w+", "", Tweets)
# Replace any non UTF-8 by space
Tweets = iconv(Tweets, "UTF-8", "UTF-8",sub="")
# Remove punctuation
Tweets = gsub("[[:punct:]]", "", Tweets)
#Remove alphanumeric character
Tweets = gsub("[^[:alnum:]]", " ", Tweets)
# Remove numbers
Tweets = gsub("[[:digit:]]", "", Tweets)
# Remove html links
Tweets = gsub("http\\w+", "", Tweets)
# Remove unnecessary spaces
Tweets = gsub("[ \\t]{2,}", "", Tweets)
Tweets = gsub("^\\s+|\\s+$", "", Tweets)
#Convert text to tolower
Tweets = tolower(Tweets)
Outfile <- list(Timestamp, Account, Tweets, Retweet)
write.csv(Outfile, "Output preprocess file name")
l <- as.Date(as.character(s[,1]), format = "%m/%d/%Y")
m <- l[!duplicated(l)]
syuzhet <- get_sentiment(Tweets, method="syuzhet")
bing <- get_sentiment(Tweets, method="bing")
afinn <- get_sentiment(Tweets, method="afinn")
nrc <- get_sentiment(Tweets, method="nrc")
sentiments <- data.frame(syuzhet, bing, afinn, nrc, l)
write.csv(sentiments, "Output sentiment file name")
#get the emotions using the NRC dictionary
emotions <- get_nrc_sentiment(Tweets)

```

```

emo_bar = colSums(emotions)
emo_sum = data.frame(count=emo_bar, emotion=names(emo_bar))
emo_sum$emotion = factor(emo_sum$emotion, levels=emo_sum$emotion[order(emo_sum$count, decreasing
= TRUE)])
# plot the different sentiments from different methods
plot_ly(sentiments, x=~l, y=~syuzhet, type="scatter", mode="jitter", name="syuzhet") %>%
add_trace(y=~bing, mode="lines", name="bing") %>%
add_trace(y=~afinn, mode="lines", name="afinn") %>%
add_trace(y=~nrc, mode="lines", name="nrc") %>%
layout(title="Recent sentiments of Vijay Rupani in India",
  yaxis=list(title="score"), xaxis=list(title="date"))
# Visualize the emotions from NRC sentiments
plot_ly(emo_sum, x=~emotion, y=~count, type="bar", color=~emotion) %>%
layout(xaxis=list(title=""), showlegend=FALSE,
  title="Distribution of emotion categories for Vijay Rupani (21 Nov - 1st Dec 2017)")
# Comparison word cloud
all = c(
paste(Tweets[emotions$anger > 0], collapse=" "),
paste(Tweets[emotions$anticipation > 0], collapse=" "),
paste(Tweets[emotions$disgust > 0], collapse=" "),
paste(Tweets[emotions$fear > 0], collapse=" "),
paste(Tweets[emotions$joy > 0], collapse=" "),
paste(Tweets[emotions$sadness > 0], collapse=" "),
paste(Tweets[emotions$surprise > 0], collapse=" "),
paste(Tweets[emotions$trust > 0], collapse=" ")
)
all <- removeWords(all, stopwords("english"))
# create corpus
corpus = Corpus(VectorSource(all))
# create term-document matrix
tdm = TermDocumentMatrix(corpus)
# convert as matrix
tdm = as.matrix(tdm)
tdm1 <- tdm[nchar(rownames(tdm)) < 11,]
# add column names
colnames(tdm) = c('anger', 'anticipation', 'disgust', 'fear', 'joy', 'sadness', 'surprise', 'trust')
colnames(tdm1) <- colnames(tdm)
comparison.cloud(tdm1, random.order=FALSE,
  colors = c("#00B2FF", "red", "#FF0099", "#6600CC", "green", "orange", "blue", "brown"),
  title.size=1, max.words=250, scale=c(2.5, 0.4),rot.per=0.4)

```