21B05

by B 05

A Project report on

COVID-19 VACCINE SENTIMENT ANALYSIS

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

Submitted by

P. TARUN PRASAD (17B91A05F1)
K. GOPAL (17B91A0581)
K. VIDYA REKHA (17B91A0589)
K.HARAVIND REDDY (17B91A0579)

Under the Guidance of

Sri.J.RAJANI KANTH,
Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SRKR ENGINEERING COLLEGE (A)

Chinna Amiram, Bhimavaram, West Godavari Dist., A.P.

[2020 –2021]

SRKR ENGINEERING COLLEGE (A)

BONAFIDE CERTIFICATE

This is to certify that the project work entitled **"COVID-19 VACCINE SENTIMENT ANALYSIS"** is the bonafide work of "**P.TARUN PRASAD (17B91A05F1),K.GOPAL(17B91A0581),K.VIDYAREKHA(17B91A0589) ,K.HARAVIND REDDY(17B91A0579)"** who carried out the project work under my supervision in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering.

SUPERVISOR

Sri.J.RAJANI KANTH
Assistant Professor

SELF DECLARATION

We here by declare that the project work entitled "**COVID-19 VACCINE SENTIMENT ANALYSIS"** is a genuine work carried out by us in B.Tech., (Computer Science and Engineering) at SRKR Engineering College(A), Bhimavaram and has not been submitted either in part or full for the award of any other degree or diploma in any other institute or University.

1. P. TARUN PRASAD (17B91A05F1)
2.K. GOPAL (17B91A0581)
3.K. VIDYA REKHA (17B91A0589)
4.K. HARAVIND REDDY (17B91A0579)

TABLE OF CONTENTS

ABSTRACT

The covid virus has changed the way people live but the opinion about vaccines has led to some positive and negative emotions. The information given by media sometimes may be correct and sometimes may be wrong. The twitter where users post their messages and interact with messages called tweets. In this project we aim to analyze the keywords present in the tweets of the public all over the world. Depending on the keywords done by the public, they can decide whether the virus vaccine should be taken or not and the dataset for the sentiment analysis task of the covid virus vaccine was collected from twitter.

LIST OF FIGURES

LIST OF TABLES

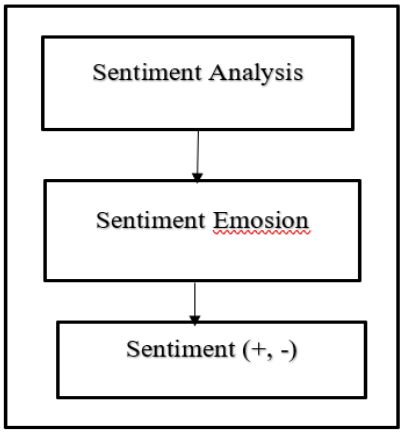
CHAPTER 1

INTRODUCTION

The covid virus is an epidemic which is happened firstly in an country called China, mainly in the city of wuhan at the end of the year 2019. This epidemic has destroyed the world financially and socially. So many people in this world are affected by this virus. To end the destruction done by the covid virus, vaccine plays a key role in saving lives of humans.

Though vaccine came into the market, due to the rumors spread by the media houses, some people are not willing to take the vaccine thinking that they will get side effects. To know the thoughts of people on vaccine, sentiment analysis needs to be done. To do sentiment analysis, we need to have the text of the tweets from the people. We can get the text from tweets by the application called Twitter. Twitter is a social media application like Facebook which is useful to connect to the people all over the world, the users of the application called Twitter had greater than 160 million who use every day. In Twitter, so many international and national interests are being discussed in it.

SENTIMENT ANALYSIS:

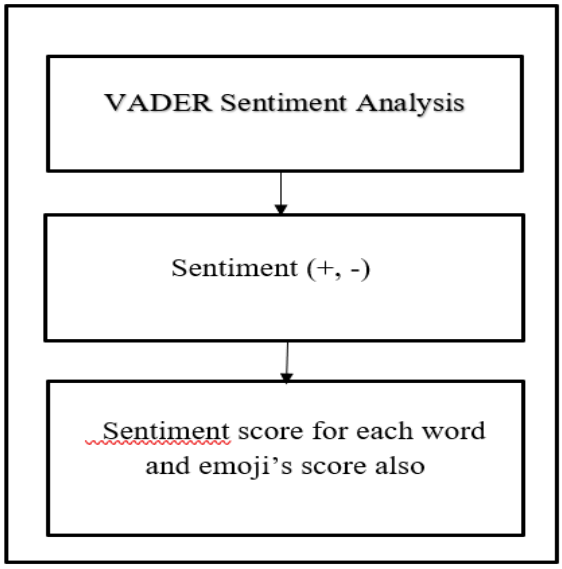


Sentiment Analysis

Analysis of sentiment is the thing which will tell whether the emotion of a person is favourable, unfavourable and neutral. To do analysis on sentiments, we need not only ML but also NLP. The combination of ML and NLP will give accurate score for emotion of a person. Analysis of

sentiment will useful for the following things: Comprehend the experience of customer, brand monitoring, monitoring of the social networking sites. To do analysis of the sentiments, first we divide the text into different parts we require. Then, we recognise the phrase and part which represents the emotion of a person. Give score to the phrase and part which is real number**.**

VADER:



Vader Sentiment Analysis

VADER is a tool which is useful in analysis of sentiments. Sentiments are also known as emotions which are important in analysing them. VADER has the capability to tell regarding positivity of the emotion and negativity of the emotion but also how intense is the emotion**.** VADER is successful in the aspects of the texts on social networking, reviews for films and products**.**

VADER will not require trained data while doing analysis on sentiments. Vader can also consider importance of uppercase letters and also punctuations like "JOY". VADER is good enough to comprehend some words like "optimist", "excellent", "kindness" as positive emotion.

CHAPTER 2

PROBLEM STATEMENT

Generally, people are facing lot of problems during this covid virus pandemic situation and when people lost their hope the global announced about vaccine. The people got some hope after announcing about vaccine but Some people may believe the false statements given by the media and rumours spread quickly about vaccine**.** As we know twitter is one of the main social networking sites where users post their messages and interact, we aim to analyze those tweets recorded about covid virus vaccine to analyze the sentiments of people for the vaccine. By doing this sentiment analysis public will get to know about vaccine and they trust whether the vaccine is good or not.

CHAPTER 3

LITERATURE SURVEY

Mohammad Abu Kausar, Arockiasamy Soosaimanickam, Mohammad Nasar [1] did research on the sentiment analysis of covid virus. As we know that covid virus destoryed the world financially, socially so the research has a goal to observe the opinions of the people for some time duration using Twitter. They gathered the tweets by using API of twitter with the help of package called R Tweet. They also did data pre-processing to do sentiment analysis effectively. Finally, they did analysis on people's sentiments with the help of a package called syuzhet and results are displayed.

Pristiyono, Mulkan Ritonga, Muhammad Ali Al Ihsan, Agus Anjar, Fauziah Hanum Rambe[2] did research on Indonesian's opinions on covid virus vaccine in which the Indonesian government is willing to increase the production of vaccine in their country.In this research,they used algorithm called naive bayes with the help of process called crawling of data.The keywords used in this research are COVID-19 and vaccine.By crawling of the data,the analysis of sentiments performed on greater than six thousand tweets in which unfavourable tweets are more than 56%,favourable tweets are around 39% and neutral tweets are about 1%.

Jim Samuel,G. G. Md. Nawaz Ali,Md. Mokhlesur Rahman, Ek Esawi and Yana Samuel[3] have performed research on opinions of the people on covid virus.In this research,the opinions took from the people on covid virus are from United States of America.They performed analysis of sentiments of the people of USA on covid virus.They did text analysis on clean data by doing data preprocessing on raw data.The accuracy of division of tweets is 91% using naive bayes algorithm and using regression they got accuracy of 74% in this research.

Yosor Alqudeimat,Deema Alenezi,Bedour AlHajri,Heba Alfouzan,Zain Almokhaizeem,Saba Altamimi,Waleed Almansouri,Sayed Alzalzalah,Ali H. Ziyab[4] did a research on vaccine for covid virus acceptance in Kuwait.This research was performed on adults regarding whether they are interested to take vaccine for corona or not.In this research,an algorithm/approach called regression which is developed by poisson.The result is 53.1% of the adults are interested to be vaccinated in order to save their lives.

Sultan Mahmud,Md. Mohsin,Ijaz Ahmed Khan,Ashraf Uddin Mian,Miah Akib Zaman[5] researched on vaccine of covid virus acceptance and evalaute the reasons on the interest of being vaccinated in order to get rid of the destruction done by covid virus.Government of

Bangladesh like Government of India started drive on vaccination from Febrauary of this year.In the middle of Jan 30 and Feb 6 of this year,they conducted a survey on internet.They used regression to evaluate the reasons for interest on taking vaccine.The result is 61.16%(370/605) of the participants showed interest on being vaccinated.Among them 35.14% are favourable to be vaccinated as soon as possible.

Jerome Nyhalah Dinga,Leontine Kouemou Sinda and Vincent P. K. Titanji[6] from Cameroon which is a country in Africa continent conducted a research on identifying the hesitancy on vaccine of covid virus in the people of Cameroon country.First to start the research they went for ethical clearance from IRB of Brea University.After then they prepared list of questions on adults of Cameroon in the online method and also they did in person too.Under the guidance of WHO and SAGE,they worked to find the percentage of Cameroonian people on hesitancy of covid virus vaccine.The Cameroonian adults who finished list of questions are 2512.Among them 84.6% of the Cameroonian people reported as hesitant to covid virus vaccine.

Kamaran H. Manguri, Rebaz N. Ramadhan, Pshko R. Mohammed Amin[7] did research on covid 19 using sentiment analysis ,As we know the reactions of people changes by posting their feelings in the form of tweets. The data collected by them from twitter is based on 2 keywords covid 19, coronavirus which contain hashtag. The data is taken by using python programming ,library used by them is tweepy and sentiment analysis done using TextBlob library in python. The two keywords are used to determine subjectivity ,polarity and total collected tweets are 530232.Finally, by performing sentiment analysis the neutral troll regarding coronavirus, covid 19 keywords where polarity is high which is more than 50% and large records were objective of 64%

CHAPTER 4

SOFTWARE REQUIREMENTS AND SPECIFICATIONS

4.1 PURPOSE:

The purpose of this project is to get the opinions of public regarding efficiency of covid virus vaccine so that people can decide themselves whether to take vaccine or not.

4.2 SCOPE:

There is a rapid increase of covid 19 cases all over the world so there is a need of using vaccine, to use any vaccine the government needs to know whether the vaccines are useful or not to the people for that sentiment analysis needs to be done.

4.3 DEFINITIONS AND ACRONYMS:

- **PDF:** Probability Density Function.
- **CDF:** Cumulative Distribution Function
- **HTTP:** Hypertext Transfer Protocol
- **JUPYTER NOTEBOOK:** It is an web application which is helpful for data science and scientific computing for programming.

4.4 EXISTING SYSTEM:

In existing system**,** some employees from media will collect the opinions of people on vaccine.

4.5 PROPOSED SYSTEM:

In proposed system, to get the public pulse accurate, we use parts of tweets done by the people in twitter on covid 19 vaccine and we do sentiment analysis.

4.6 REQUIREMENT ANALYSIS

4.6.1 FUNCTIONAL REQUIREMENTS

1.Select the dataset which consists of parts of tweets.

2.Do data cleaning by erasing noisy data using Pandas library.

3.Start the sentiment analysis using VADER method

4.Perform analysis on data by plotting graphs for favourable, unfavourable and neutral emotions of people.

5.Perform analysis on the cut off for most favourable and unfavourable emotions.

6.Visualise most favourable and unfavourable sentiments.

4.6.2 NON-FUNCTIONAL REQUIREMENTS:

- **AVAILABILITY:** This program is easy to use and easily available.

- **PERFORMANCE:** The program is built in such a way it can work efficiently

- **SOFTWARE REQUIREMENTS:**

  PYTHON: It is a popular programming language and it is used all over the world.

  ANACONDA: It is a software which is helpful for python programmers**.**

- **HARDWARE REQUIREMENTS:**

  RAM: 4GB

  PROCESSOR: Intel core i3

  HARD DISK:500GB

  OS: Windows 10

CHAPTER 5

SYSTEM DESIGN

SYSTEM DESIGN:

System design is defined as the approach of designing system. It is useful to show different elements like modules, interfaces and data in the system.

UML DESIGN:

UML is a language which is useful to represent the functionality of the system. It's like a template in many engineering fields. It is useful to visualize the working of a system. It is useful for software developers, entrepreneurs, data analysts etc... for the design and analysis of the system. The general diagrams drawn for UML are:

1.Use Case Diagram

2.Class Diagram

3.Interaction Diagrams

i. Sequence Diagram

ii. Collaboration Diagram

4.Activity Diagram

1. USE CASE DIAGRAM:

It shows the communication between user and the system in which how system responds to the user**.**

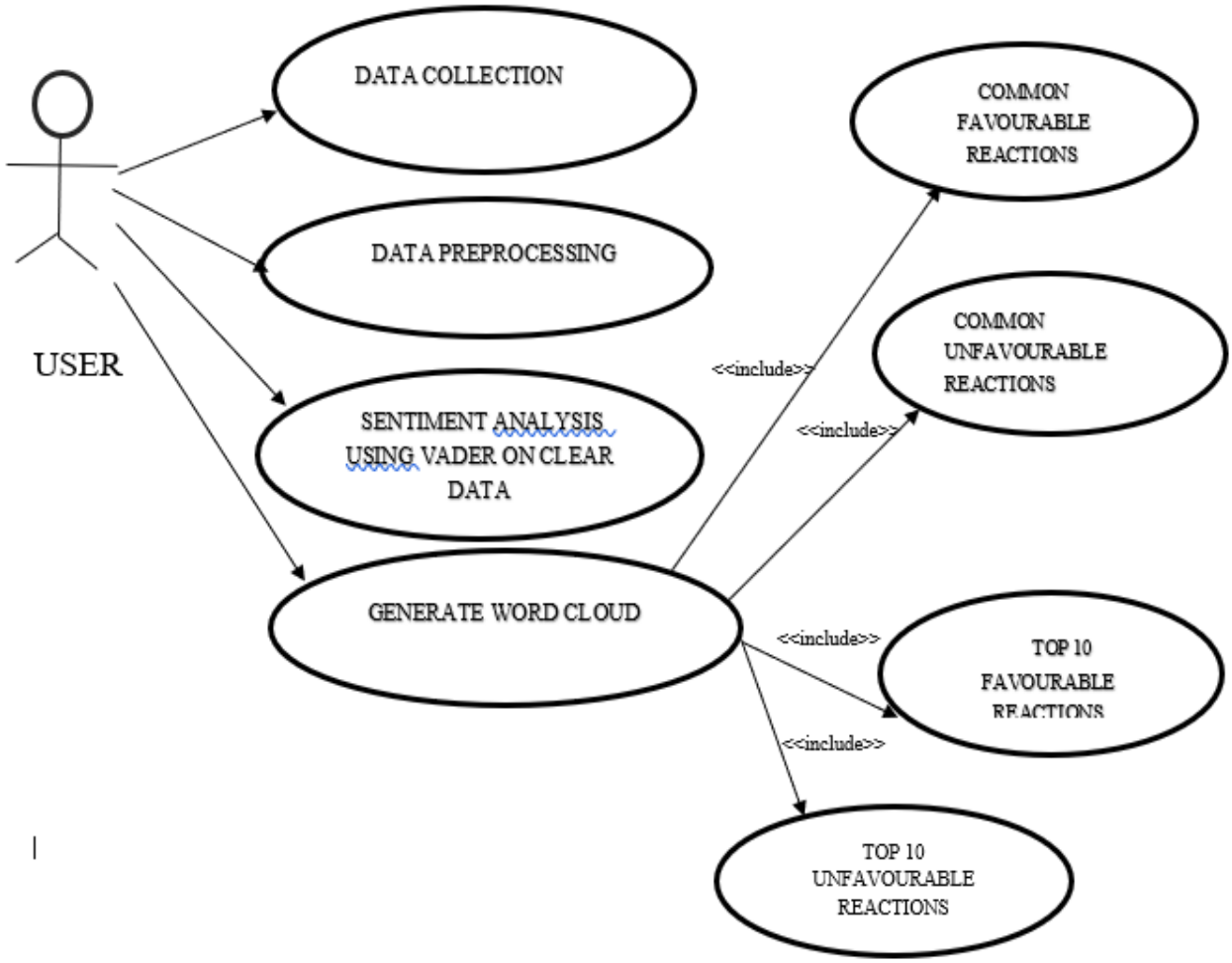


Fig: 5.1 Use Case Diagram

2.CLASS DIAGRAM:

It is the diagram to show the static form of system. It is also useful in representation of classes or objects**.**

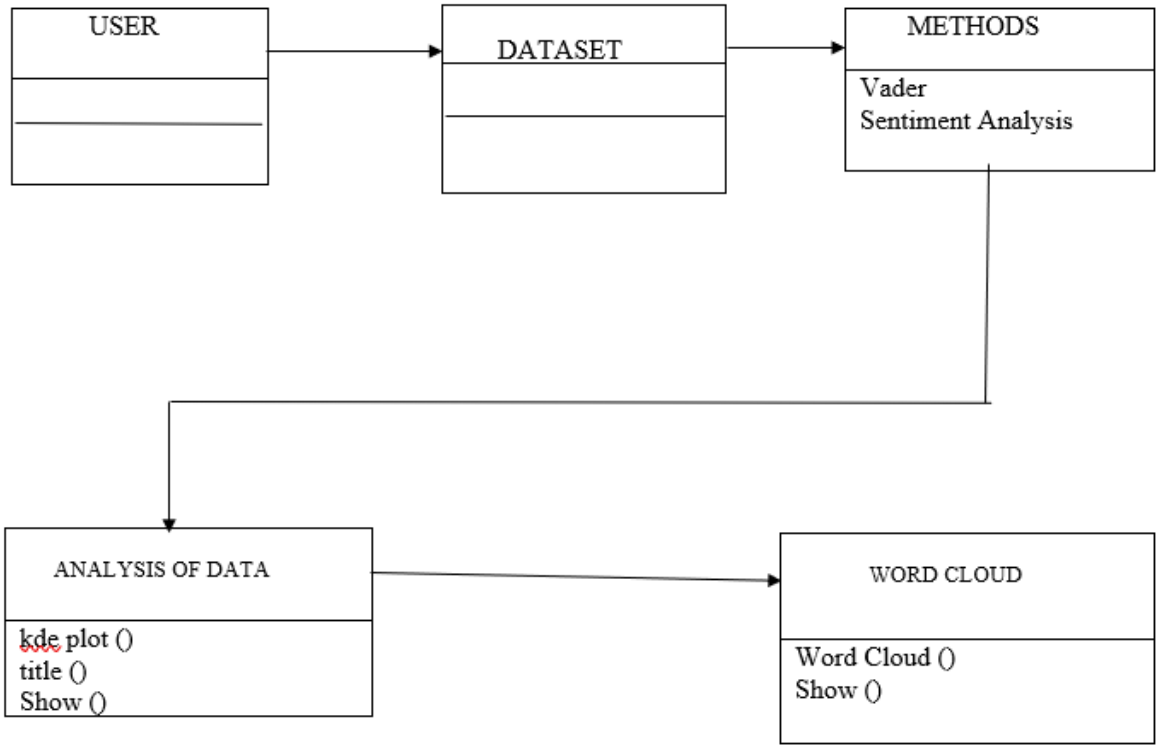


Fig: 5.2 Class Diagram

3 . INTERACTION DIAGRAM:

The purpose of interaction diagram is to represent the interactive approach of system. There are two types of interaction diagrams are sequence diagram and collaboration diagram in UML**.**

i. SEQUENCE DIAGRAM:

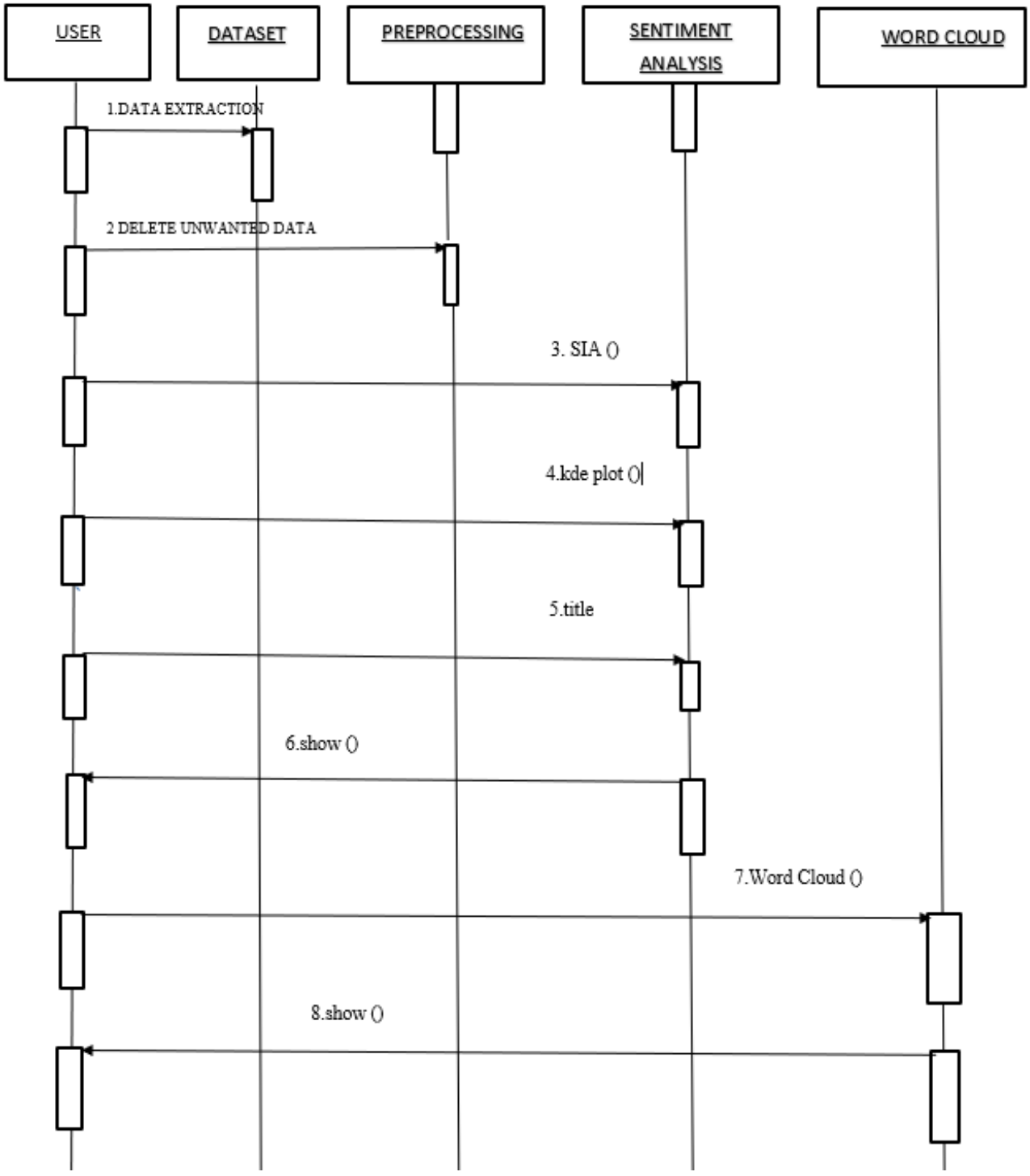


Fig: 5.3.1 Sequence Diagram

ii. COLLABRATION DIAGRAM:

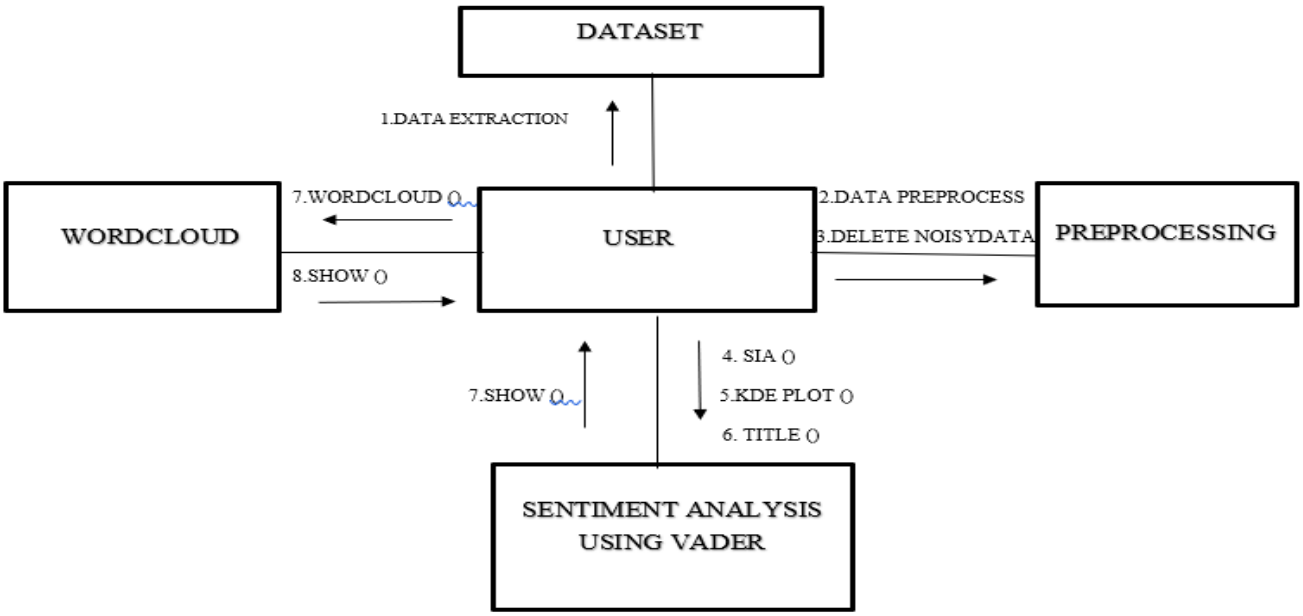


Fig: 5.3.2 Collaboration Diagram

4.ACTIVITY DIAGRAM:

It will show the work flow from starting point to ending point representing every path when activity will be executed.

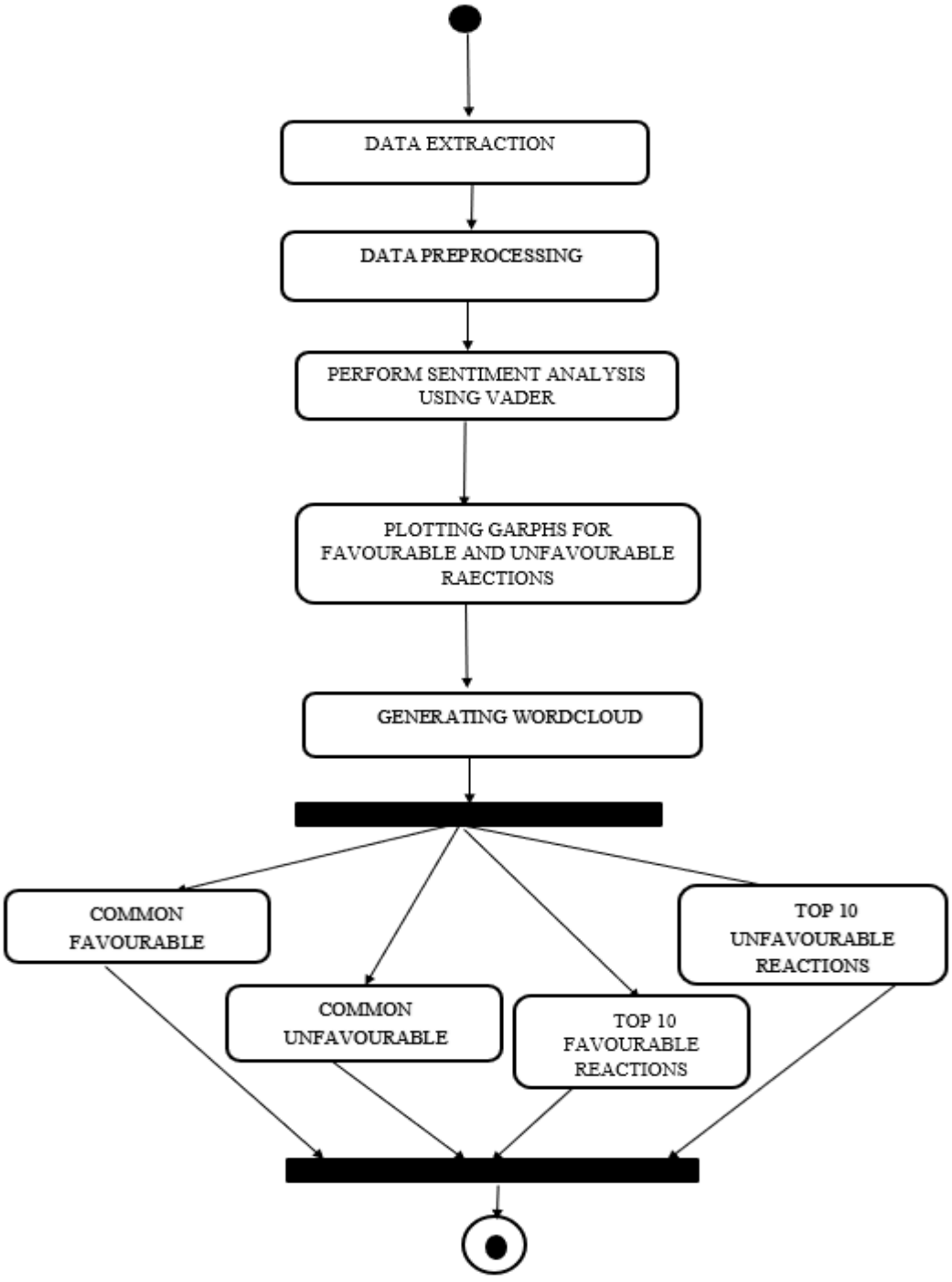


Fig: 5.4 Activity Diagram

CHAPTER 6

METHODOLOGY

The Methodology proposed is present in different phases. The phases are:

6.1 Data Extraction:

The dataset is collected using Kaggle. Dataset consists of pairs of tweets such as hashtags, source of twitter, text. The collected data is loaded into the environment.

Dataset: https://www.kaggle.com/gpreda/pfizer-vaccine-tweets/download

6.2 Data Preprocessing:

It involves transformation of raw data into clear data which is free from noisy data. To do data preprocessing, first we need to use Pandas package and import this. As our dataset is in csv format, to read this dataset there is a function in Pandas package called read_csv(). In the read_csv() function, the dataset we have in single quotes. Next, we will change the words present in the dataset into lowercase by using the function lower(). Subsequently we erase handles of twitter, hashtags, Unified Resource Locators.URL stands for Unified Resource Locator. We also erase the character which are single and we replace more spaces into space which has we can fill one character.

6.3 Sentiment analysis using Vader:

Vader is a tool which depends on dictionary that will map lexical characters to intensity of emotion. The score of the text can be get through summarization of every word present in it. Vader has the capacity/capability to comprehend the general content of the words like “dislike” as negative sentiment. If we take examples for identifying the words as positive sentiment like “Satisfactory”, “Happiness”, “Optimistic”, etc... Vader can also consider importance of uppercase letters also punctuations like “ELEVATION”. To prepare the data, first we have to use a method/function called SIA() which is helpful to analyse the intensity of the sentiments. Vader can separate/segregate the sentiments as positive, negative and neutral sentiments.

6.4 Data analysis on tweets:

In this part, we do plot on sentiments distribution will follow normal distribution. Here we can plot graphs using a method called kdeplot() which belongs to a library called seaborn. Seaborn is a built-in library of the python which is helpful in statistics. It is a visualization of data library. Then there will be cumulative distribution function applied on the sentiments of the public. To see the plot, we use show() which belongs to a library called matplotlib.

6.5 Analysis of sentiments using python:

Here, we analyse cutoff for the tweets which are favorable and the tweets which are not favorable by plotting the graphs on the screen. To analyse the cutoff, we have to sort the data by date. Then, we have to copy into another variable. For the plotting of graphs for favourable and unfavourable tweets, we use kdeplot() method/function which belongs to built-in library called seaborn.

6.6 Generate the word cloud for most positive and negative sentiments:

We generate the word cloud for more general among favorable and non-favorable tweets. To generate the word cloud, we use default class called word cloud.

6.7 Generate top 10words for positive and negative sentiments:

Here, we generate top 10 words for the favorable/positive and non-favorable/negative sentiments. To generate top 10 words in word cloud, we will sort the values of the keys in dictionary.

CHAPTER 7

SYSTEM TESTING

Testing is an approach to checking the system in which whether there are bugs present in the system or not. It's the time where we can analyze the system whether the user will be satisfied with our application or not by giving various testcases on the system.

Testcase means it is a concept in software engineering which has to follow a series of steps or actions, whether the application executed will be pass or fail in a system. By using testcase, software testers can analyze the bugs in the system such that they can take certain steps to improve the application.

Tab 7.1: Test case for setting up jupyter notebook

S No	Test case	Input values	Expected output	Obtained output	Result
1	Setting up new notebook	Create a new notebook in jupyter	Opening of new python notebook with all required specifications	Same	Success

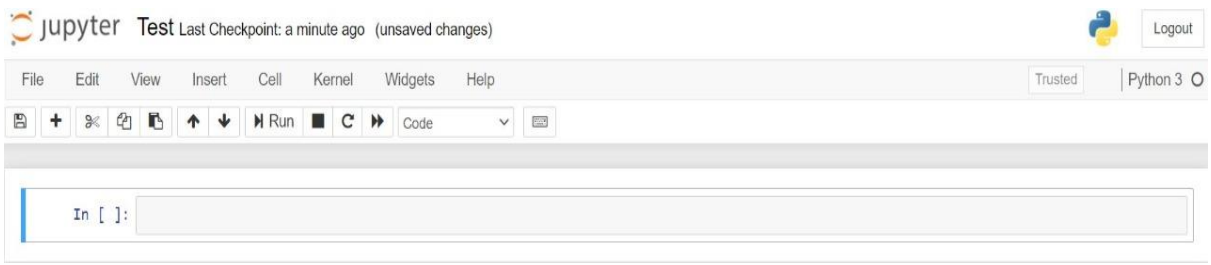


Fig 7.1 Screenshot of opening new notebook

Tab 7.2: Test case for loading dataset

S No	Test case	Input values	Expected output	Obtained output	Result
2	Data loading	File path for the dataset(vaccination_tweets.csv)	Loading the required data	Same	Success

```
i_info = pd.read_csv('vaccination_tweets.csv')
i_info.tail()
```

text	hashtags	source	retweets	favorites	is_retweet
Pfizer Inc and German partner BioNTech SE bega...	NaN	TweetDeck	0	0	False
The number of Covid-19 cases today, 26 Mar 202...	NaN	Twitter for iPhone	0	0	False
2nd jab done! ✔💉\nThank you, Qatar! QA \n#covi...	['covidvaccine', 'PfizerBioNTech']	Instagram	0	0	False
@RaHaake @Marius_Raabe "Drive-thru" 👆 \n#Pfize...	['PfizerBiontech']	Twitter Web App	0	1	False
The Western Pharma companies has repeatedly pr...	NaN	Twitter for Android	0	0	False

Fig 7.2 Screenshot for loading dataset

Tab: 7.3 Test case for clearing noisy data

S No	Test case	Input values	Expected output	Obtained output	Result
3	Data preprocessing	Find the noisy data such as twitter handlers, hashtags, URLs, special and single characters.	No noisy data present	Same	Success

```
i_info = pd.read_csv('vaccination_tweets.csv')
i_info.text =i_info.text.str.lower()
i_info.text = i_info.text.apply(lambda x:re.sub('@[^\s]+','',x))
i_info.text = i_info.text.apply(lambda x:re.sub(r'\B#\S+','',x))
i_info.text = i_info.text.apply(lambda x:re.sub(r"http\S+", "", x))
i_info.text = i_info.text.apply(lambda x:' '.join(re.findall(r'\w+', x)))
i_info.text = i_info.text.apply(lambda x:re.sub(r'\s+[a-zA-Z]\s+', '', x))
i_info.text = i_info.text.apply(lambda x:re.sub(r'\s+', ' ', x, flags=re.I))
i_info.tail()
```

	id	user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites	user_verified	date
7623	1375413472096313348	Yahoo India	World Wide Web	We endeavour to provide news that is unbiased,...	2009-08-28 04:44:40	101727	167	152	True	2021-03-26 11:45:00
7624	1375384825599684608	theSun	Petaling Jaya	Breaking news updates from the website of Mala...	2009-07-29 01:58:44	143517	461	167	False	2021-03-26 09:51:10
7625	1375378921630302209	𝓜𝓪𝓫𝓲𝓮	Doha, Qatar	Blair's Momma half pleasant half mean S...	2010-08-05 06:40:56	181	261	243	False	2021-03-26 09:27:42
7626	1375360563312726017	arniebear	NaN	... has this thing about working in a basement...	2009-07-02 00:53:06	69	88	10275	False	2021-03-26 08:14:45
7627	1375352345681260546	MAU News	Republic of Mauritius MU	MAU News is a Mauritian online media which is ...	2019-12-09 11:07:56	177	496	311	False	2021-03-26 07:42:06

text	hashtags	source	retweets	favorites	is_retweet
pfizer inc and german partner biontech se bega...	NaN	TweetDeck	0	0	False
the number of covid 19 cases today 26 mar 2021...	NaN	Twitter for iPhone	0	0	False
2nd jab done thank you qatar qatar national co...	['covidvaccine', 'PfizerBioNTech']	Instagram	0	0	False
drive thru	['PfizerBiontech']	Twitter Web App	0	1	False
the western pharma companies has repeatedly pr...	NaN	Twitter for Android	0	0	False

Fig 7.3 Screenshot showing clear noisy data

Tab: 7.4 Test case to implement VADER Sentiment Analysis

S No	Test case	Input values	Expected output	Obtained output	Result
4	Adding positive, negative, neutral sentiments	Features to be added	Added positive, negative, neutral sentiments	Same	Success

```
sid = SIA()
i_info['sentiments']           = i_info['text'].apply(lambda x: sid.polarity_scores(' '.join(re.findall(r'\w+',x.lower()))))
i_info['Positive Sentiment']   = i_info['sentiments'].apply(lambda x: x['pos']+1*(10**-6))
i_info['Neutral Sentiment']    = i_info['sentiments'].apply(lambda x: x['neu']+1*(10**-6))
i_info['Negative Sentiment']   = i_info['sentiments'].apply(lambda x: x['neg']+1*(10**-6))
i_info.drop(columns=['sentiments'],inplace=True)
i_info.tail()
```

	id	user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites	user_verified	date
7623	1375413472096313348	Yahoo India	World Wide Web	We endeavour to provide news that is unbiased,...	2009-08-28 04:44:40	101727	167	152	True	2021-03-26 11:45:00
7624	1375384825599684608	theSun	Petaling Jaya	Breaking news updates from the website of Mala...	2009-07-29 01:58:44	143517	461	167	False	2021-03-26 09:51:10
7625	1375378921630302209	𝓜𝓪𝓫𝓲𝓮	Doha, Qatar	Blair's Momma half pleasant half mean S...	2010-08-05 06:40:56	181	261	243	False	2021-03-26 09:27:42
7626	1375360563312726017	arniebear	NaN	... has this thing about working in a basement...	2009-07-02 00:53:06	69	88	10275	False	2021-03-26 08:14:45
7627	1375352345681260546	MAU News	Republic of Mauritius MU	MAU News is a Mauritian online media which is ...	2019-12-09 11:07:56	177	496	311	False	2021-03-26 07:42:06

text	hashtags	source	retweets	favorites	is_retweet	Positive Sentiment	Neutral Sentiment	Negative Sentiment
pfizer inc and german partner biontech se bega...	NaN	TweetDeck	0	0	False	0.128001	0.872001	0.000001
the number of covid 19 cases today 26 mar 2021...	NaN	Twitter for iPhone	0	0	False	0.177001	0.823001	0.000001
2nd jab done thank you qatar qatar national co...	['covidvaccine', 'PfizerBioNTech']	Instagram	0	0	False	0.217001	0.783001	0.000001
drive thru	['PfizerBiontech']	Twitter Web App	0	1	False	0.000001	1.000001	0.000001
the western pharma companies has repeatedly pr...	NaN	Twitter for Android	0	0	False	0.000001	1.000001	0.000001

Fig 7.4 Screenshot of implemented VADER Sentiment Analysis

Tab: 7.5 Test case on data analysis of sentiments of tweets

S No	Test case	Input values	Expected output	Obtained output	Result
5	Performing data analysis on sentiments	Values of Sentiments	Plotted graphs for positive, negative and neutral	Same	Success

```
plt.subplot(2,1,1)
plt.title('Distriubtion Of Sentiments Across Our Tweets',fontsize=19,fontweight='bold')
sns.kdeplot(i_info['Negative Sentiment'],bw=0.1)
sns.kdeplot(i_info['Positive Sentiment'],bw=0.1)
sns.kdeplot(i_info['Neutral Sentiment'],bw=0.1)
plt.xlabel('Sentiment Value',fontsize=19)
plt.subplot(2,1,2)
plt.title('CDF Of Sentiments Across Our Tweets',fontsize=19,fontweight='bold')
sns.kdeplot(i_info['Negative Sentiment'],bw=0.1,cumulative=True)
sns.kdeplot(i_info['Positive Sentiment'],bw=0.1,cumulative=True)
sns.kdeplot(i_info['Neutral Sentiment'],bw=0.1,cumulative=True)
plt.xlabel('Sentiment Value',fontsize=19)
plt.show()
```

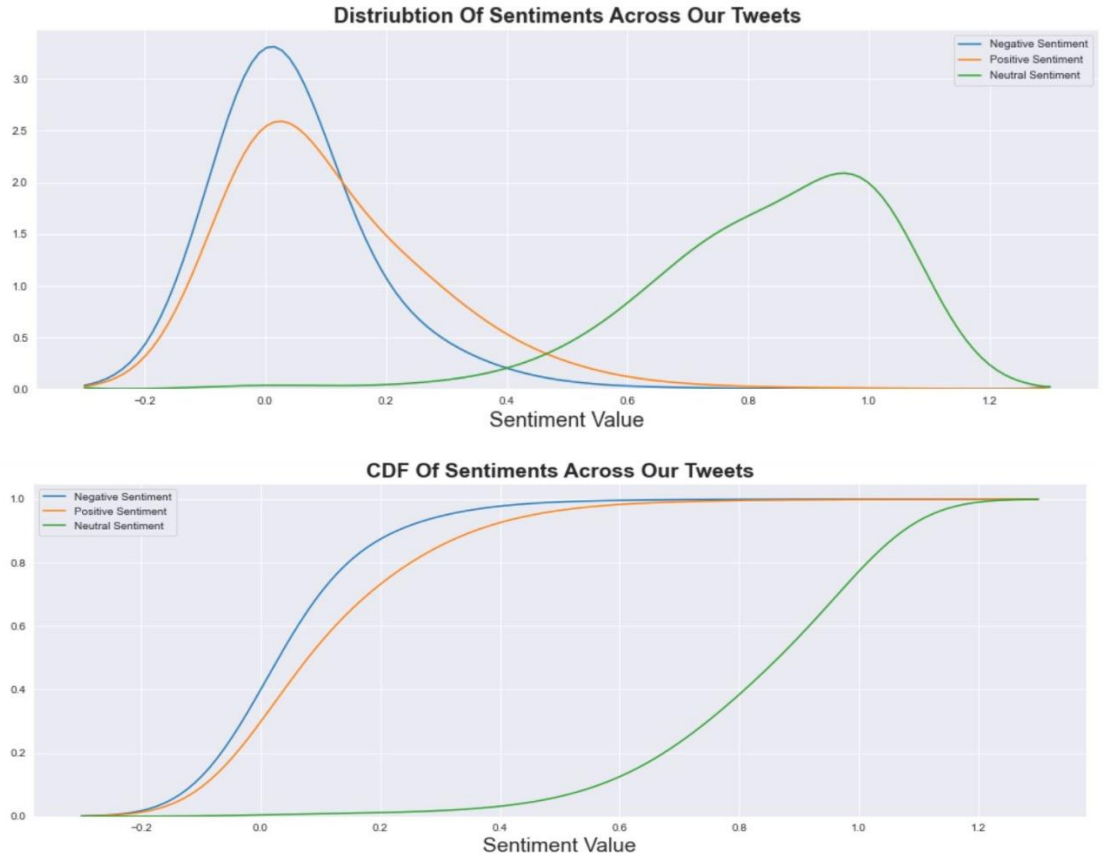


Fig 7.5 Screenshot for analyzed data on sentiment of tweets

Tab: 7.6 Testcase to analyze cut off for positive and negative emotions

S No	Test case	Input values	Expected output	Obtained output	Result
6	Analyze cutoff for most positive and negative emotions	Values of positive and negative emotions	Plotted graphs for positive and negative emotions	Same	Success

```
i_info = i_info.sort_values(by='date')
it_info=i_info.copy()
it_info['date'] = pd.to_datetime(i_info['date']).dt.date

it_info['year']         = pd.DatetimeIndex(it_info['date']).year
it_info['month']        = pd.DatetimeIndex(it_info['date']).month
it_info['day']          = pd.DatetimeIndex(it_info['date']).day
it_info['day_of_year']  = pd.DatetimeIndex(it_info['date']).dayofyear
it_info['quarter']      = pd.DatetimeIndex(it_info['date']).quarter
it_info['season']       = it_info.month%12 // 3 + 1

plt.subplot(2,1,1)
plt.title('Selecting A Cut-Off For Most Positive/Negative Tweets',fontsize=19,fontweight='bold')

aax0 = sns.kdeplot(i_info['Negative Sentiment'],bw=0.1)

kde_x, kde_y = aax0.lines[0].get_data()
aax0.fill_between(kde_x, kde_y, where=(kde_x>0.25) ,
                interpolate=True, color='b')

plt.annotate('Cut-Off For Most Negative Tweets', xy=(0.25, 0.5), xytext=(0.4, 2),
            arrowprops=dict(facecolor='red', shrink=0.05),fontsize=16,fontweight='bold')

aax0.axvline(i_info['Negative Sentiment'].mean(), color='r', linestyle='--')
aax0.axvline(i_info['Negative Sentiment'].median(), color='tab:orange', linestyle='-')
plt.legend({'PDF':i_info['Negative Sentiment'],r'Mean: {:.2f}'.format(i_info['Negative Sentiment'].mean()):i_info['Negative Senti
            r'Median: {:.2f}'.format(i_info['Negative Sentiment'].median()):i_info['Negative Sentiment'].median()})

plt.subplot(2,1,2)
```

```
aax1 = sns.kdeplot(i_info['Positive Sentiment'],bw=0.1,color='green')

plt.annotate('Cut-Off For Most Positive Tweets', xy=(0.4, 0.43), xytext=(0.4, 2),
            arrowprops=dict(facecolor='red', shrink=0.05),fontsize=16,fontweight='bold')
kde_x, kde_y = aax1.lines[0].get_data()
aax1.fill_between(kde_x, kde_y, where=(kde_x>0.4) ,
                interpolate=True, color='green')
aax1.set_xlabel('Sentiment Strength',fontsize=18)


aax1.axvline(i_info['Positive Sentiment'].mean(), color='r', linestyle='--')
aax1.axvline(i_info['Positive Sentiment'].median(), color='tab:orange', linestyle='-')
plt.legend({'PDF':i_info['Positive Sentiment'],r'Mean: {:.2f}'.format(i_info['Positive Sentiment'].mean()):i_info['Positive Senti
            r'Median: {:.2f}'.format(i_info['Positive Sentiment'].median()):i_info['Positive Sentiment'].median()})

plt.show()
```

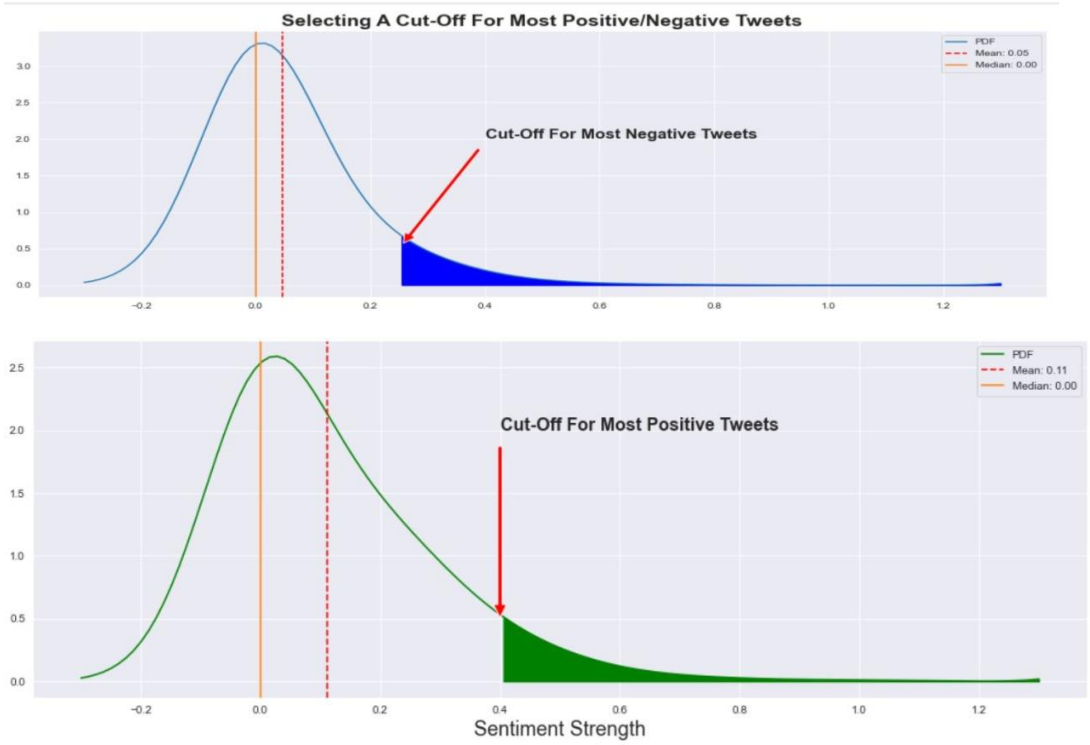


Fig 7.6 Screenshot for analyzed cutoff on most positive and negative emotions

Tab: 7.7 Test case to find most common words among most positive and negative emotions

S No	Test case	Input values	Expected output	Obtained output	Result
7	Generating word clouds for positive and negative emotions	Values present in positive and negative sentiments	Generated Word Cloud for most common positive and negative emotions	Same	Success

```
MOST_Positive = i_info[i_info['Positive Sentiment'].between(0.4,1)]
MOST_Negative = i_info[i_info['Negative Sentiment'].between(0.25,1)]
MOST_Positive_text = ' '.join(MOST_Positive.text)
MOST_Negative_text = ' '.join(MOST_Negative.text)
pwc = WordCloud(width=600,height=400,collocations = False).generate(MOST_Positive_text)
nwc = WordCloud(width=600,height=400,collocations = False).generate(MOST_Negative_text)
plt.subplot(1,2,1)
plt.title('Common Words Among Most Positive Tweets',fontsize=16,fontweight='bold')
plt.imshow(pwc)
plt.axis('off')
plt.subplot(1,2,2)
plt.title('Common Words Among Most Negative Tweets',fontsize=16,fontweight='bold')
plt.imshow(nwc)
plt.axis('off')
plt.show()
```

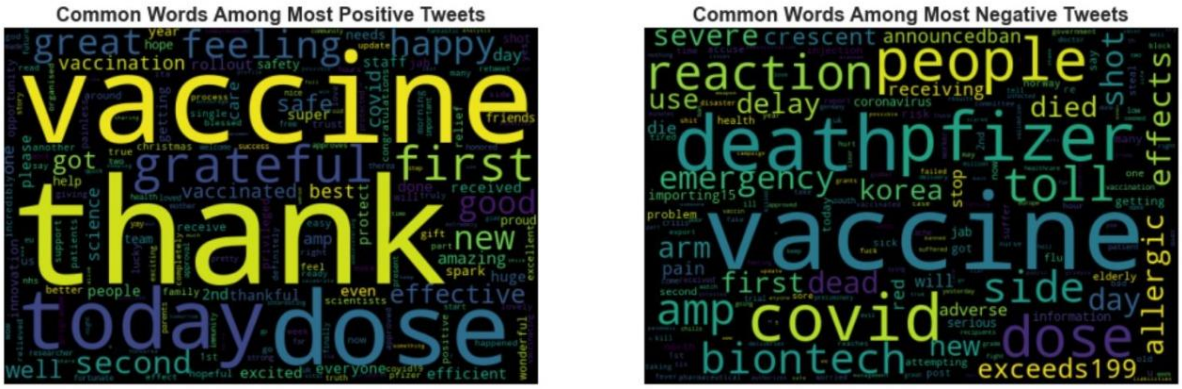


Fig 7.7 Screenshot of most common words among most positive and negative emotions

Tab: 7.8 Test case to implement word cloud for top 10 most positive and negative emotions

S No	Test case	Input values	Expected output	Obtained output	Result
8	Generating word cloud for top 10 most positive and negative emotions	Text present in the dataset	Generated Word Cloud for top10 most positive and negative emotions	Same	Success

```
t_l = MOST_Positive_text
w11_dict = dict()
for word in t_l.split():
    w= word.strip()
    if w in STOPWORDS:
        continue
    else:
        w11_dict[w] = w11_dict.get(w,0)+1
w11_dict = {k: v for k, v in sorted(w11_dict.items(), key=lambda item: item[1],reverse=True)}
t_l = MOST_Negative_text
w22_dict = dict()
for word in t_l.split():
    w= word.strip()
    if w in STOPWORDS:
        continue
    else:
        w22_dict[w] = w22_dict.get(w,0)+1
w22_dict = {k: v for k, v in sorted(w22_dict.items(), key=lambda item: item[1],reverse=True)}
TOP_10_pos = list(w11_dict.keys())[:10]
TOP_10_neg = list(w22_dict.keys())[:10]
plt.subplot(1,2,1)
w_c = WordCloud(width=600,height=400,collocations = False,colormap='nipy_spectral').generate(' '.join(TOP_10_pos))
plt.title('Top 10 Words In Most Positive Tweets',fontsize=19,fontweight='bold')
plt.imshow(w_c)
plt.axis('off')
plt.subplot(1,2,2)
w_c = WordCloud(width=600,height=400,collocations = False,colormap='nipy_spectral').generate(' '.join(TOP_10_neg))
plt.title('Top 10 Words In Most Negative Tweets',fontsize=19,fontweight='bold')
```

```
plt.imshow(w_c)
plt.axis('off')
plt.subplot(1,2,2)
w_c = WordCloud(width=600,height=400,collocations = False,colormap='nipy_spectral').generate(' '.join(TOP_10_neg))
plt.title('Top 10 Words In Most Negative Tweets',fontsize=19,fontweight='bold')
plt.imshow(w_c)
plt.axis('off')
plt.show()
```

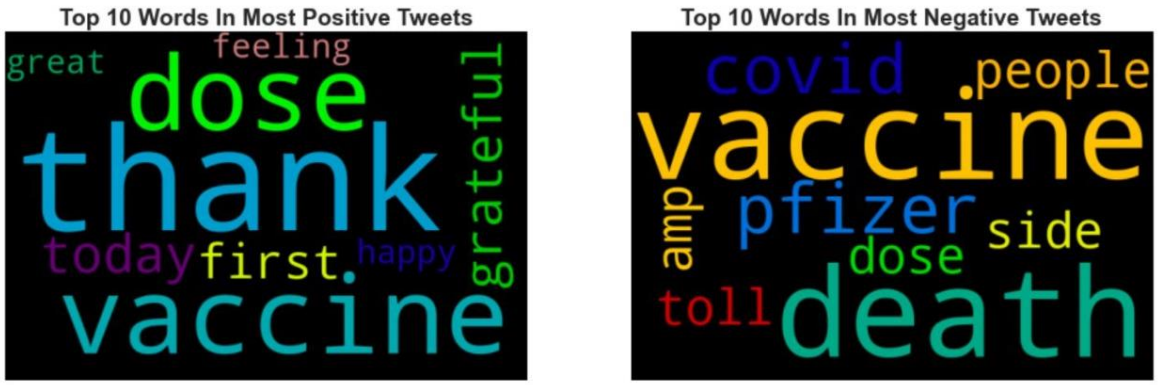
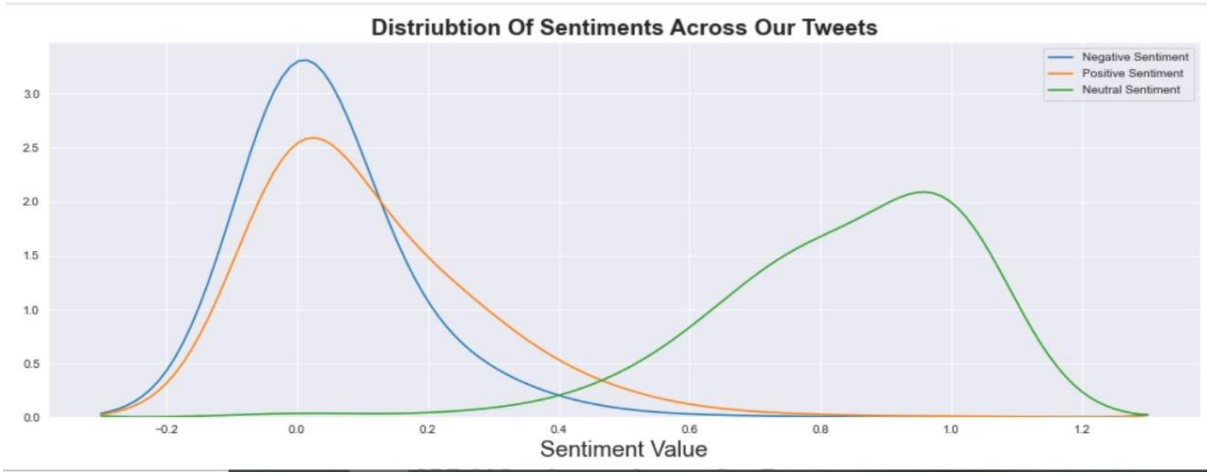
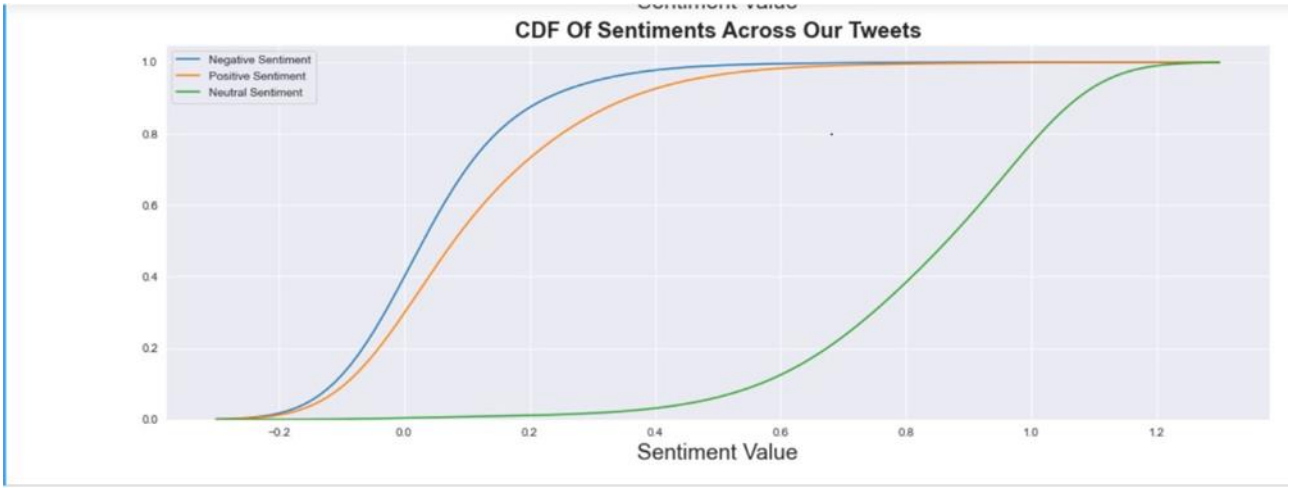
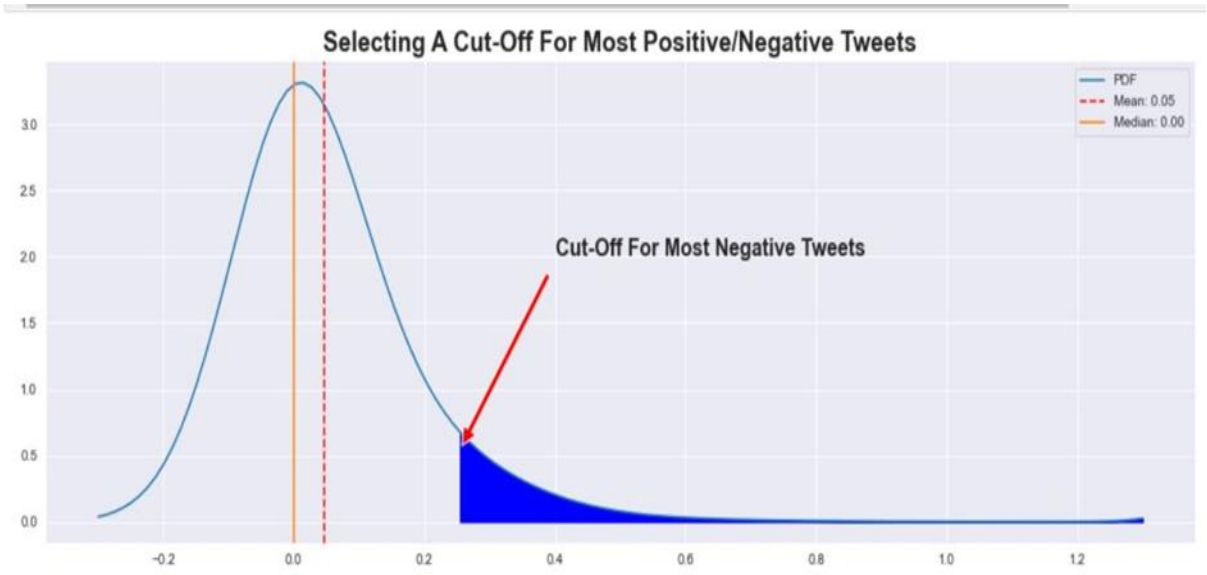


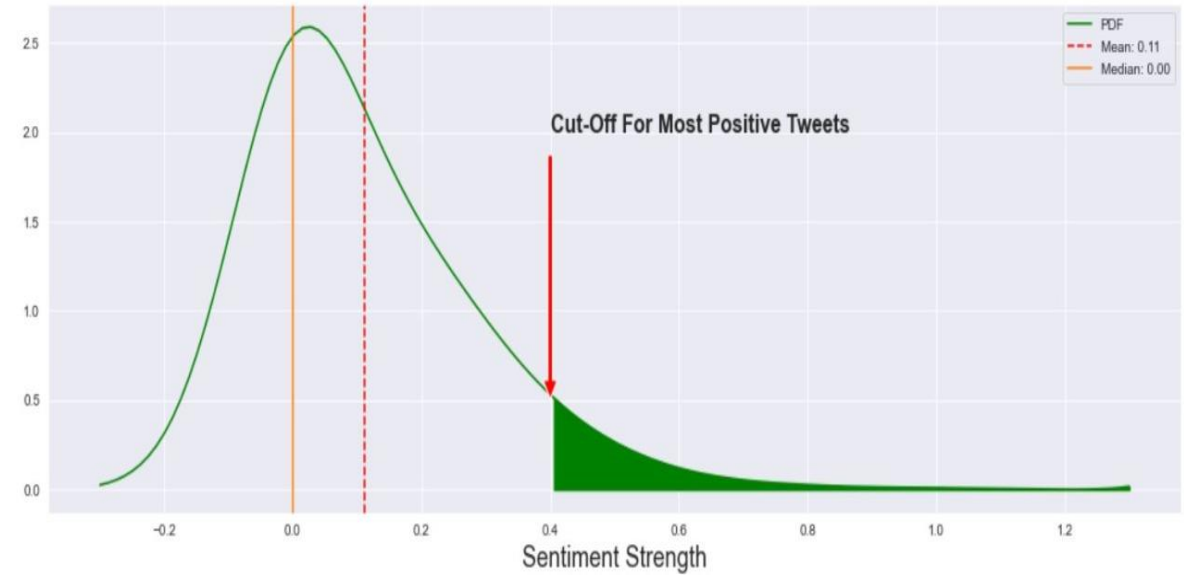
Fig 7.8 Screenshot of top 10 most positive and negative emotions

RESULT ANALYSIS:

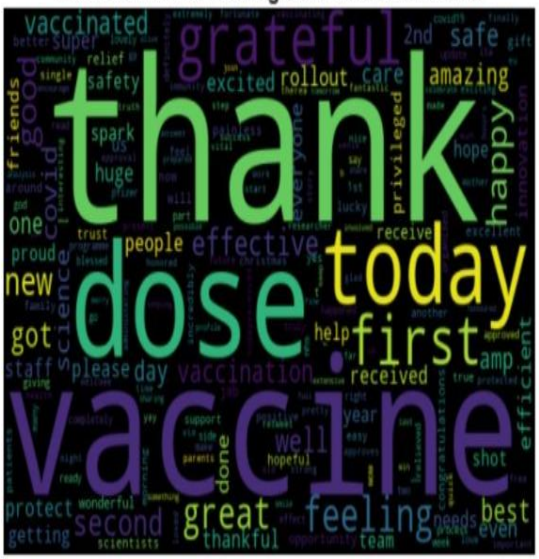





Selecting A Cut-Off For Most Positive/Negative Tweets
PDF
Mean: 0.05
Median: 0.00
Cut-Off For Most Negative Tweets
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2

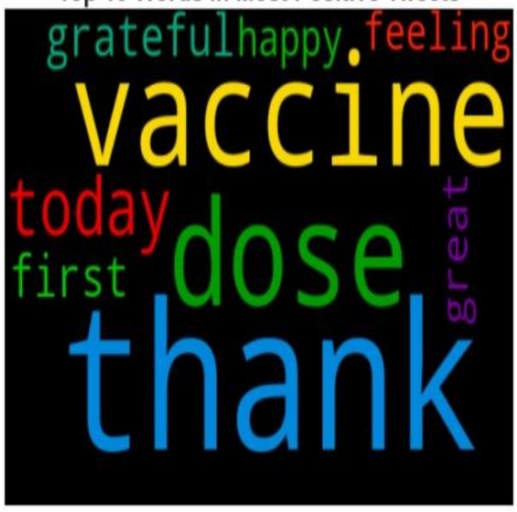

PDF
Mean: 0.11
Median: 0.00
Cut-Off For Most Positive Tweets
2.5
2.0
1.5
1.0
0.5
0.0
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Sentiment Strength

Common Words Among Most Positive Tweets

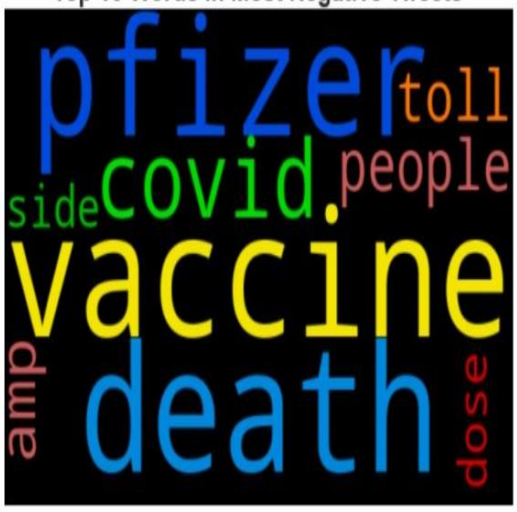


Common Words Among Most Negative Tweets

Top 10 Words In Most Positive Tweets



Top 10 Words In Most Negative Tweets



CHAPTER 8

CONCLUSION AND FUTURE WORK

Covid-19 vaccine sentiment analysis is done using tweets done by the people. Finally, we generated word clouds for more general favourable and unfavourable words and also top 10 favourable and unfavourable words using a library called Word Cloud. We also plotted graphs for positive, negative and neutral emotions. As a future work, we can tie up with governments to know the opinions of the people on the vaccine.

REFERENCES

[1] Mohammad Abu Kausar, Arockiasamy Soosaimanickam, Mohammad Nasar “Public Sentiment Analysis on Twitter Data during covid-19 Outbreak", International Journal of Advance Computer Science and Applications, (Vol.12, No. 2),2021.

[2] Pristiyono, Mulkan Ritonga, Muhammad Ali Al Ishan, Agus Anjar, Fauziah Hanum Rambe “Sentiment Analysis of covid-19 vaccine in Indonesia using Naive Bayes Algorithm ",Annual Conference on Computer Science and Engineering Technology (Iop Conf.Series: Materials Science and Engineering 1088(2021)012045,doi:10.1088/1757-899X/1088/1/012045),2021.

[3] Jim Samuel ,G. G. Md. Nawaz Ali, Md. Mokhlesur Rahman, Ek Esawi and Yana Samuel "covid 19 Public Sentiment Insights and Machine Learning for Tweets Classification "(Information 2020,11,314;doi:10.3390/info11060314), June 11,2020

[4] Yosor Alqudeimat, Deema Alenezi, Bedour AlHajri, Heba Alfouzan, Zain Almokhaizeem, saba Altamimi, Waleed Alaman souri,Sayed Alzalzalah, Ali H. Ziyab "Acceptance of a covid-19 Vaccine and its related Determinants among the General Adult Population in Kuwait"(med Princ Pract DOI:10.1159/000514636), January 22,2021.

[5] Sultan Mahmud,Md. Mohsin,Ijaz Ahmed Khan,Ashraf Uddin Mian,Miah Akib Zaman "Acceptance of COVID-19 Vaccine and Its Determinants in Bangladesh",2021.

[6] Jerome Nyhalah Dinga, Leontine Kouemou Sinda and Vincent P.K. Titanji,"Assessment of Vaccine Hesitancy to a covid-19 Vaccine in Cameroonian Adults and Its Global Implication ",(Vaccines 2021,9,175. http://doi.org/10.3390/vaccines 9020175),February 19,2021.

[7] Kamaran H.manguri,Rebaz N.Ramadhan,Pshko R.Mohammad Amin "Twitter sentiment analysis on worldwide covid 19 outbreaks ",Kurdistan journal of Applied research (Doi 10.24017/covid.8),2020.

Code:

```
import numpy as np

import pandas as pd

import re

import string

import nltk

import matplotlib pyplot as plt

import seaborn as sns

sns.set_style('darkgrid')

nltk.download('vader_lexicon')

from nltk.sentiment.vader import SentimentIntensityAnalyzer as SIA

from wordcloud import WordCloud,STOPWORDS

from pandas.plotting import autocorrelation_plot

from statsmodels.graphics.tsaplots import plot_acf

from statsmodels.graphics.tsaplots import plot_pacf

from statsmodels.tsa.seasonal import seasonal_decompose

from nltk.util import ngrams

from nltk import word_tokenize

from nltk.stem import PorterStemmer

from nltk.stem import WordNetLemmatizer

import random

plt.rc('figure',figsize=(17,13))

i_info = pd.read_csv('vaccination_tweets.csv')

i_info.text =i_info.text.str.lower()
```

```
i_info.text = i_info.text.apply(lambda x:re.sub('@[^\s]+','',x))

i_info.text = i_info.text.apply(lambda x:re.sub(r'\B#\S+','',x))

i_info.text = i_info.text.apply(lambda x:re.sub(r"http\S+", "", x))

i_info.text = i_info.text.apply(lambda x:' '.join(re.findall(r'\w+', x)))

i_info.text = i_info.text.apply(lambda x:re.sub(r'\s+[a-zA-Z]\s+', '', x))

i_info.text = i_info.text.apply(lambda x:re.sub(r'\s+', ' ', x, flags=re.I))

sid = SIA ()

i_info['sentiments'] = i_info['text'].apply(lambda x: sid.polarity_scores('
'.join(re.findall(r'\w+',x.lower()))))

i_info['Positive Sentiment']   = i_info['sentiments'].apply(lambda x: x['pos']+1*(10**-6))

i_info['Neutral Sentiment']    = i_info['sentiments'].apply(lambda x: x['neu']+1*(10**-6))

i_info['Negative Sentiment']   = i_info['sentiments'].apply(lambda x: x['neg']+1*(10**-6))

i_info.drop(columns=['sentiments'],inplace=True)

plt.subplot(2,1,1)

plt.title('Distriubtion Of Sentiments Across Our Tweets',fontsize=19,fontweight='bold')

sns.kdeplot(i_info['Negative Sentiment'],bw=0.1)

sns.kdeplot(i_info['Positive Sentiment'],bw=0.1)

sns.kdeplot(i_info['Neutral Sentiment'],bw=0.1)

plt.xlabel('Sentiment Value',fontsize=19)

plt.subplot(2,1,2)

plt.title('CDF Of Sentiments Across Our Tweets',fontsize=19,fontweight='bold')

sns.kdeplot(i_info['Negative Sentiment'],bw=0.1,cumulative=True)

sns.kdeplot(i_info['Positive Sentiment'],bw=0.1,cumulative=True)

sns.kdeplot(i_info['Neutral Sentiment'],bw=0.1,cumulative=True)

plt.xlabel('Sentiment Value',fontsize=19)
```

```
plt.show()


i_info =i_info.sort_values(by='date')

it_info=i_info.copy()

it_info['date'] = pd.to_datetime(i_info['date']).dt.date

it_info['year']  = pd.DatetimeIndex(it_info['date']).year

it_info['month'] = pd.DatetimeIndex(it_info['date']).month

it_info['day']  = pd.DatetimeIndex(it_info['date']).day

it_info['day_of_year'] = pd.DatetimeIndex(it_info['date']).dayofyear

it_info['quarter'] = pd.DatetimeIndex(it_info['date']).quarter

it_info['season'] = it_info.month%12 // 3 + 1

plt.subplot(2,1,1)

plt.title('Selecting A Cut-Off For MOST Positive/Negative
Tweets',fontsize=19,fontweight='bold')

aax0 = sns.kdeplot(i_info['Negative Sentiment'],bw=0.1)

kde_x, kde_y = aax0.lines[0].get_info()

aax0.fill_between(kde_x, kde_y, where=(kde_x>0.25) , interpolate=True, color='b')

plt.annotate('Cut-Off For Most Negative Tweets', xy=(0.25, 0.5), xytext=(0.4, 2),
arrowprops=dict(facecolor='red', shrink=0.05),fontsize=16,fontweight='bold')

aax0.axvline(i_info['Negative Sentiment'].mean(), color='r', linestyle='--')

aax0.axvline(i_info['Negative Sentiment'].median(), color='tab:orange', linestyle='-')

plt.legend({'PDF':i_info['Negative Sentiment'],r'Mean: {:.2f}'.format(i_info['Negative
Sentiment'].mean()):i_info['Negative Sentiment'].mean(), r'Median:
{:.2f}'.format(i_info['Negative Sentiment'].median()):i_info['Negative Sentiment'].median()})

plt.subplot(2,1,2)
```

```
aax1 = sns.kdeplot(i_info['Positive Sentiment'],bw=0.1,color='green')
plt.annotate('Cut-Off For Most Positive Tweets', xy=(0.4, 0.43), xytext=(0.4, 2),
        arrowprops=dict(facecolor='red', shrink=0.05),fontsize=16,fontweight='bold')
kde_x, kde_y = aax1.lines[0].get_info()
aax1.fill_between(kde_x, kde_y, where=(kde_x>0.4) ,interpolate=True, color='green')
aax1.set_xlabel('Sentiment Strength',fontsize=18)
aax1.axvline(i_info['Positive Sentiment'].mean(), color='r', linestyle='--')
aax1.axvline(i_info['Positive Sentiment'].median(), color='tab:orange', linestyle='-')
plt.legend({'PDF':i_info['Positive Sentiment'],r'Mean: {:.2f}'.format(i_info['Positive
Sentiment'].mean()):i_info['Positive Sentiment'].mean(), r'Median:
{:.2f}'.format(i_info['Positive Sentiment'].median()):i_info['Positive Sentiment'].median()})
plt.show()


MOST_Positive = i_info[i_info['Positive Sentiment'].between(0.4,1)]
MOST_Negative = i_info[i_info['Negative Sentiment'].between(0.25,1)]
MOST_Positive_text = ' '.join(MOST_Positive.text)
MOST_Negative_text = ' '.join(MOST_Negative.text)
pwc = WordCloud(width=600,height=400,collocations =
False).generate(MOST_Positive_text)
nwc = WordCloud(width=600,height=400,collocations =
False).generate(MOST_Negative_text)
plt.subplot(1,2,1)
plt.title('Common Words Among MOST Positive Tweets',fontsize=16,fontweight='bold')
plt.imshow(pwc)
```

```
plt.axis('off')
plt.subplot(1,2,2)
plt.title('Common Words Among MOST Negative Tweets',fontsize=16,fontweight='bold')
plt.imshow(nwc)
plt.axis('off')
plt.show()


t_l = MOST_Positive_text
w11_dict = dict()
for word in t_l.split():
    w= word.strip()
    if w in STOPWORDS:
        continue
    else:
        w11_dict[w] = w11_dict.get(w,0)+1
w11_dict = {k: v for k, v in sorted(w11_dict.items(), key=lambda item:
item[1],reverse=True)}
l_t = MOST_Negative_text
w22_dict = dict()
for word in t_l.split():
    w= word.strip()
    if w in STOPWORDS:
        continue
    else:
        w22_dict[w] = w22_dict.get(w,0)+1
```

```
w22_dict = {k: v for k, v in sorted(w22_dict.items(), key=lambda item:
item[1],reverse=True)}

TOP_10_pos = list(w11_dict.keys())[:10]

TOP_10_neg = list(w22_dict.keys())[:10]

plt.subplot(1,2,1)

w_c = WordCloud(width=600,height=400,collocations =
False,colormap='nipy_spectral').generate(' '.join(TOP_10_pos))

plt.title('TOP 10 Words In MOST Positive Tweets',fontsize=19,fontweight='bold')

plt.imshow(w_c)

plt.axis('off')

plt.subplot(1,2,2)

w_c = WordCloud(width=600,height=400,collocations =
False,colormap='nipy_spectral').generate(' '.join(TOP_10_neg))

plt.title('Top 10 Words In MOST Negative Tweets',fontsize=19,fontweight='bold')

plt.imshow(w_c)

plt.axis('off')

plt.show()
```