www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.03
TPI 2019; 8(3): 590-595

www.thepharmajournal.com
Received: 20-01-2019
Accepted: 24-02-2019

Dr. Shivani Dubey
Assistant Professor, Computer Science & Engineering, Lingaya's Vidyapeeth, Faridabad, Haryana, India

Correspondence Author:
Dr. Shivani Dubey
Assistant Professor, Computer Science & Engineering, Lingaya's Vidyapeeth, Faridabad, Haryana, India

Systematic study of twitter sentiment analysis on COVID-19 datasets using various techniques

Dr. Shivani Dubey

DOI: https://doi.org/10.22271/tpi.2019.v8.i3j.25398

Abstract
In this research we have done COVID-19 sentiment analysis using twitter datasets. This research paper gives an empirical analysis of public tweets and speeches related to the COVID-19 pandemic. Using natural language processing techniques, we analysed a number of tweets, news articles, and other online content to understand attitudes and sentiments related to the pandemic. Our research shows that there are many different opinions about COVID-19, including fear, uncertainty, anger and hope. We also identified key themes driving this sentiment, including government policies, vaccine development and social distancing measures. Findings from this study can inform healthcare communication strategies and help policymakers better understand public sentiment in times of crisis.

Keywords: COVID-19, twitter, sentiment analysis, positive, negative and neutral sentiments, subjectivity and polarity of data

Introduction
Since the COVID-19 pandemic first appeared in December 2019, it has had a huge influence on the world. When the outbreak started, it was in Wuhan, China, and it swiftly spread to neighboring nations. Due of its severity and global distribution by March 11, 2020, the World Health Organization classified it as a pandemic. In numerous nations, including the United States, France, Germany, Russia, Brazil, India, and others, the virus has killed thousands and infected millions of people. As a result, many nations have turned to partial or total lockdowns to stop the virus's spread. People used social media platforms at this time to express their feelings and ideas and look for calming activities [1].

This study analyses the sentiment of tweets made by citizens of afflicted nations. Over 20,000 tweets with the hashtags #Lockdown, #COVID19, #CORONAVIRUS, #CORONA, #COVID-19, #StayHome, #COVID_19, #COVIDPandemic, #COVID19, #CoronaVirus, #quarantine, #Coronavirusoutbreak, #StayHomeStaySafe, #COVID, and others were used in the analysis. The tweets were gathered, pre-processed, and then submitted to sentiment analysis and several algorithms, which produced intriguing findings [2].

The purpose of the study was to comprehend and examine the various coping mechanisms and emotional expressions used by individuals from various afflicted nations. Every part of the world was impacted by the epidemic, which led to state- and microeconomic issues. Psychotic symptoms, economic collapse, travel limitations, and closed economies with social isolation may affect both individuals and the population [3]. As of mid-August 2020, over 21 million people globally had tested positive for COVID-19, with an estimated 773,072 deaths.

The top ten countries affected by COVID-19 until August 17, 2020, were the USA, France, Russia, Brazil, India, South Africa, Mexico, Chile, Peru, Colombia, Chile, and Spain. Social networking sites, such as Twitter, have been used to increase information sharing and communication during social events and health crises. Twitter has been one of the top trends since January 2020, and with more countries adopting quarantine measures and policy, people have become more reliant on social media to receive or share messages related to COVID-19 [4].

As the main channel of communication for sharing COVID-19-related measures, policy changes, and general information, many governments all around the world use Twitter. The number of tests has increased, and analysis of Twitter data reveals popular sentiment and conversations on COVID-19. For instance, to analyse and comprehend the main topics and attitudes regarding COVID-19, some researchers gathered tweets from February 2 to March

15, 2020, and then used sentiment analysis [6].

Twitter messages are called tweets and are publicly available. Therefore, they should be considered as raw material to generate ideas, analyze customer performance, and conduct various evaluations of policy schemes and the results of studies on emotions. Today, people's thoughts and opinions about various products are also based on their internet buying experiences. As a result, marketing and purchasing departments at businesses ought to devote more effort to assessing the client experience [5].

The sentiment analysis research was conducted on tweets from infected countries, provided insights into the feelings of people affected by COVID-19. The study aimed to determine how people in different countries felt and shared their emotions during the pandemic. The results of the study highlighted the different emotions expressed by people in different countries, such as fear, anxiety, sadness, hope, and optimism [6]. The objective of this research is to examine people's emotions reflected in tweets posted on Twitter within a specific timeframe. The study aims to address the following things:

1. How can tweets be collected? Which hashtags will be used to collect the tweets?
2. What data cleaning processes will be used to preprocess the tweets, including removal of white spaces, links, punctuations, stop words, tokenization, and retweets?
3. How can sentiment analysis be performed and what are the key findings?

This study will focus on tweets written in English to explore the sentiment of people from different countries affected by the COVID-19 pandemic. The collected tweets will be subjected to various text mining algorithms to extract meaningful insights that can inform our understanding of how people are coping with this pandemic [7].

Related Works

Numerous academics have studied and worked on sentiment on various social media platforms, such as Instagram, Facebook, and Twitter, in particular, as the epidemic spreads over the world. These efforts have helped to uncover user mood, thinking, and attitude. This section concludes some significant papers that served as references.

Researchers have also examined Twitter data before to detect different epidemics. They also analyzed Twitter activity related to the COVID-19 pandemic to understand people's emotions and opinions. The researchers gathered pertinent tweets using the Twitter API and then used machine learning to categorize them as having positive, negative, or neutral emotions [8]. They also used the NLTK library to preprocess the tweets and the Text-Blob dataset for evaluation. Results were presented through visualizations. One study focused on using machine learning algorithms to analyze global sentiment about COVID-19, with the Naive Bayes approach being deemed particularly effective. In a different study, researchers created a list of hashtags associated with COVID-19 in order to gather certain tweets across a two-week period in early 2020 [8]. Using an API, the researchers collected tweets and stored them as plain text. They then identified relevant keywords and assessed their significance, including topics such as infection control measures, vaccination, and discrimination based on race. These keywords were analyzed to better understand their impact and relevance to the overall study.

The researchers used sentiment analysis to determine the emotional tone of tweets related to COVID-19, and analyzed how anxiety levels changed over time as the pandemic reached its peak in the United States. They used appropriate text visualization to describe their findings. The study also compared two important classification methods for sentiment analysis in terms of their effectiveness for classifying different lengths of coronavirus-related tweets, with Naïve Bayes process achieving 91% classification accuracy for short tweets [9].

In another study, the authors proposed an efficient platform called MISNIS (Intelligent Mining of Social Networks' Impact on Companies) for collecting, storing, managing, and analyzing Twitter data. This platform helped non-technical users to mine data quickly and achieved high success in collecting Portuguese tweets. The researchers emphasized the importance of understanding the emotions associated with text analysis and highlighted the rich and significant knowledge that Twitter provides. The report also mentioned that sentiment analysis is a useful method for analyzing huge volumes of Twitter data and that Twitter covers a wide range of issues, including politics and education. A natural language processing tool called sentiment analysis divides text into various groups according to its emotional tone.

Research Methods

As a relatively new area of study, sentiment analysis in microblogging has been the subject of extensive prior research on user reviews, web posts, documents, and articles as well as general sentiment analysis at the phrase level. These departures from Twitter are primarily caused by the 280-character limit per tweet, which forces users to convey their opinions succinctly and concisely. Launched in March 2006, Twitter is a platform for social networking and microblogging. Twitter is the most with 330 million monthly active user’s reliable and popular social networking platform. The researchers are encouraged by Twitter to determine feelings on almost everything, including public health sentiments information, products, equality, natural calamities, digital technology, movies, politics etc. [10].

We gathered datasets from Kaggle, stored them in CSV format, then performed sentiment analysis on tweet text to identify the overwhelming emotions (happy, anger, contempt, fear, sorrow, etc.) and emotional valence (positive, neutral, or negative) of each tweet. Then, in order to categorize and assess pertinent themes over time inside a tweet, we performed topic modelling using an unsupervised machine learning approach.

Data Collection

Kaggle datasets were applied to a batch of tweets. The hashtag used to collect the tweet for example was #COVID-19, #COVID19, #CORONAVIRUS, #CORONA, #StayHomeStaySafe, #StayHome, #StayHomeSaveLives, #COVID_19, #COVIDPandemic, #COVID19, #CoronaVirus, #Lockdown, #Quarantine, #quarantine, #CoronavirusOurel and #COVID; and the collected tweets were stored in a CSV file. To prevent duplicate tweets, the messages, retweets, and responses were removed during the gathering of tweets. The data cleaning procedure was carried out, and stop words, spaces, and punctuation were eliminated, much like a full database [11].

B. Sentiment Analysis

It entails calculating and quantifying people's emotions, or their ideas regarding a certain situation, such as product reviews, etc. A procedure called sentiment analysis uses letters that are either favorable, negative, or neutral. Natural language processing (NLP) and machine learning (ML) techniques are used in a sentiment analysis system for text analysis to assign computed sentiment scores inside a sentence or phrase to themes, entities, topics, and many categories. It is thought that in order to fully understand human emotions, we must also use well calibrated algorithms in an autonomous sentiment analysis [12]. The polarity scores of the tweets were divided into three groups (Positive, Neutral, and Negative) at this step, which was an extension of the previous stage. Positive emotions have a polarity range of (0.1), negative emotions have a polarity range of (-1.0), and neutral emotions have a polarity range of 0.0. By simply exploring the dataset and using filters, this phase was completed. The sentiment dataset made use of each of these three classifications separately [13].

```
removing url

remove_url= lambda x: re.sub(r"https\S+", "", str(x))
text_en_lr = df['text'].apply(remove_url)
text_en_lr.head()

0    If I smelled the scent of hand sanitizers toda...
1    Hey @Yankees @YankeesPR and @MLB - wouldn't it...
2    @diane3443 @wdunlap @realDonaldTrump Trump nev...
3    @brookbanktv The one gift #COVID19 has give me...
4    25 July : Media Bulletin on Novel #CoronaVirus...
Name: text, dtype: object
```

Fig 1: Removing URL

```
needed_columns = ['user_name', 'date', 'text']
df =df[needed_columns]
df.head()
```

	user_name	date	text
0	ᏉᎥ☻լꂅᏆ	2020-07-25 12:27:21	If I smelled the scent of hand sanitizers toda...
1	Tom Basile us	2020-07-25 12:27:17	Hey @Yankees @YankeesPR and @MLB - wouldn't it...
2	Time4fisticuffs	2020-07-25 12:27:14	@diane3443 @wdunlap @realDonaldTrump Trump nev...
3	ethel mertz	2020-07-25 12:27:10	@brookbanktv The one gift #COVID19 has give me...
4	DIPR-J&K	2020-07-25 12:27:08	25 July : Media Bulletin on Novel #CoronaVirus...

Fig 2: Selecting columns

Punctuation and Lower Case: Any punctuation that is not a letter, such as "#," "123," "$," etc., is deleted from the text and replaced with a space. The headlines are then changed to lowercase to avoid the inconsistency produced by Python's case sensitivity, which interprets uppercase and lowercase terms differently.

```
Convert tweets to lowercase

text_en_lr_lc = text_en_lr.apply(lambda x: x.lower())
text_en_lr_lc.head()

0    if i smelled the scent of hand sanitizers toda...
1    hey @yankees @yankeespr and @mlb - wouldn't it...
2    @diane3443 @wdunlap @realdonaldtrump trump nev...
3    @brookbanktv the one gift #covid19 has give me...
4    25 july : media bulletin on novel #coronavirus...
Name: text, dtype: object
```

Fig 3: Converting tweets to lowercase

```
import string

text_en_lr_lc_pr = text_en_lr_lc.apply(lambda x: x.translate(str.maketrans('', '', string.punctuation)))
text_en_lr_lc_pr.head()

0    if i smelled the scent of hand sanitizers toda...
1    hey yankees yankeespr and mlb  wouldnt it have...
2    diane3443 wdunlap realdonaldtrump trump never ...
3    brookbanktv the one gift covid19 has give me i...
4    25 july  media bulletin on novel coronavirusup...
Name: text, dtype: object
```

Fig 4: String Punctuation

Eliminating stop-words: Words that don't contribute to the sentiment of the text and are rarely used by sentiment analysis programs. As a result, they are merely unnecessary textual clutter.

```
# removing stopwords

word_list= [word for line in text_en_lr_lc_pr for word in line.split()]
word_list[:5]

['if', 'i', 'smelled', 'the', 'scent']
```

Fig 5: Removing stop-words

Text Blob: TextBlob is a Python package used for natural language processing (NLP). TextBlob actively employed the Natural Language ToolKit (NLTK) to carry out its functions. The NLTK library gives users easy access to a vast array of lexical resources while enabling them to do classification, categorization, and a range of other tasks. A nice library that offers complex processing and textual analysis is TextBlob.The polarity and subjectivity of a statement are returned by TextBlob. The polarity scale ranges from [-1,1], with [-1] denoting a negative feeling, [0] denoting neutral, and [1] denoting a positive emotion. In TextBlob, semantic markers make analysis easier [14].

```
from textblob import TextBlob

# Create a function to get to get the subjectivity and Polarity
def getSentiment(text):
    blob = TextBlob(text)
    sentiment_polarity = blob.sentiment.polarity
    sentiment_subjectivity = blob.sentiment.subjectivity

    if sentiment_polarity > 0:
        sentiment_label = 'Positive'
    elif sentiment_polarity < 0:
        sentiment_label = 'Negative'
    else:
        sentiment_label = 'Neutral'

    result = {'polarity' : sentiment_polarity,
              'subjectivity' : sentiment_subjectivity,
              'sentiment': sentiment_label}
    return result

getSentiment(df['text'].iloc[0])

{'polarity': -0.25, 'subjectivity': 0.25, 'sentiment': 'Negative'}
```

Fig 6: Subjectivity and Polarity

Polarity finding: One of the most crucial pre-processing procedures in the entire process was this one. We calculated and determined the polarity score of the tweets in the dataset using Python's Text-Blob package. Using Text-Blob, the cleaned tweets from the previous step were submitted to several ranking models, and for the provided tweets, a generalized polarity score was discovered. This score was closely tied to the collection of words in the text, such as trigrams, bigrams, and unigrams. Text-Blob produces a value between [1, +1], where + 1 denotes high positive polarity, 1 denotes high negative polarity, and 0 denotes neutral. The dataset used then included this score as a distinct new attribute [15].

```
getSentiment(df['text'].iloc[0])

{'polarity': -0.25, 'subjectivity': 0.25, 'sentiment': 'Negative'}

df['sentiment_results'] = df['text'].apply(getSentiment)
df['sentiment_results'].head()

0    {'polarity': -0.25, 'subjectivity': 0.25, 'sen...
1    {'polarity': 0.5, 'subjectivity': 0.5, 'sentim...
2    {'polarity': 0.0, 'subjectivity': 0.0, 'sentim...
3    {'polarity': 0.0, 'subjectivity': 0.3571428571...
4    {'polarity': 0.0, 'subjectivity': 0.0, 'sentim...
Name: sentiment_results, dtype: object
```

Fig 7: Sentiment Results

Results and Observation

Twitter sentiment analysis assists in detecting patterns or trends in the sentiment expressed and can offer insightful information about how people feel about a certain subject or event. The sentiment analysis can be divided into following parts:

Positive emotion: Tweets with positive sentiment typically utilize exclamation points, more positive language, and terms to show joy or satisfaction [16].

Negative emotion: Tweets with a bad attitude typically use more negative language, show grief or irritation, and include terms like "hate" or "dislike."

Neutral Emotion: Tweets with neutral attitude often utilize more neutral vocabulary, provide factual information, and steer clear of emotive language.

Here are some figures and charts representing the results obtained after working on the COVID-19 datasets [17].

Word Cloud: A word cloud is a visual representation of text data where each word's size corresponds to how frequently it appears in the text. A word cloud may be used to visualize the most frequently used terms in a collection of tweets connected to a certain topic or event in the context of Twitter sentiment analysis. Refer Fig.8, Fig.9, Fig.10 given below.

Here as we see in the word cloud COVID-19 is shown big as it was used the most during COVID-19 pandemic. Other words are comparatively smaller in size than the word COVID-19.

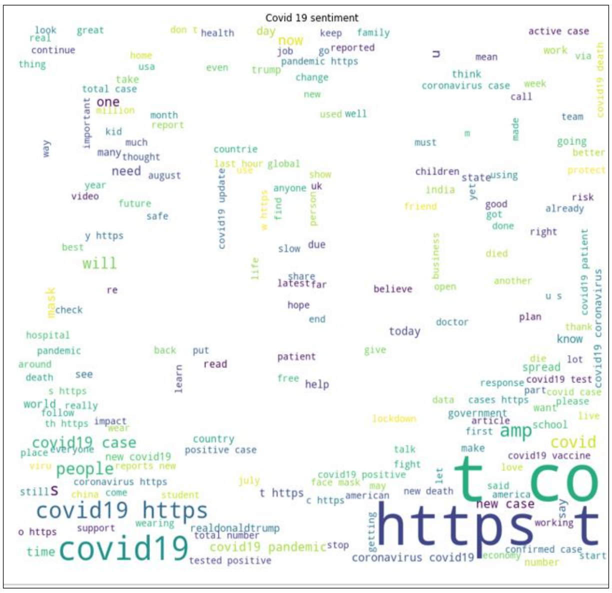

Fig 8: Word cloud - I for COVID 19 Sentiment analysis dataset

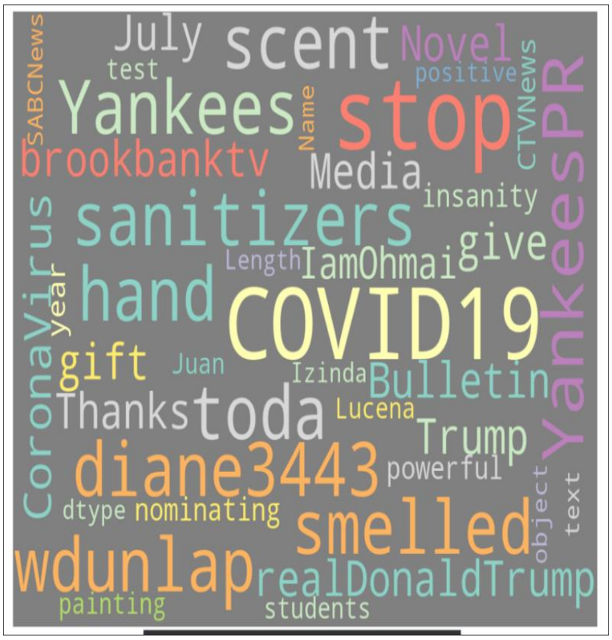

Fig 9: Word cloud -II for COVID 19 Sentiment analysis dataset

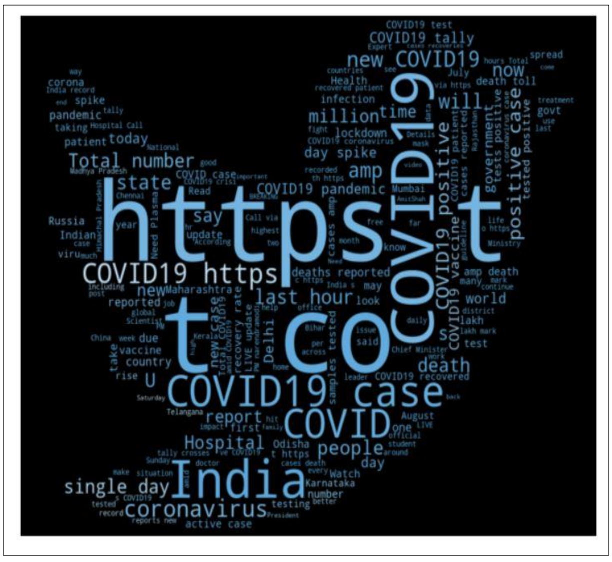

Fig 10: Word cloud – III for COVID 19 Sentiment analysis

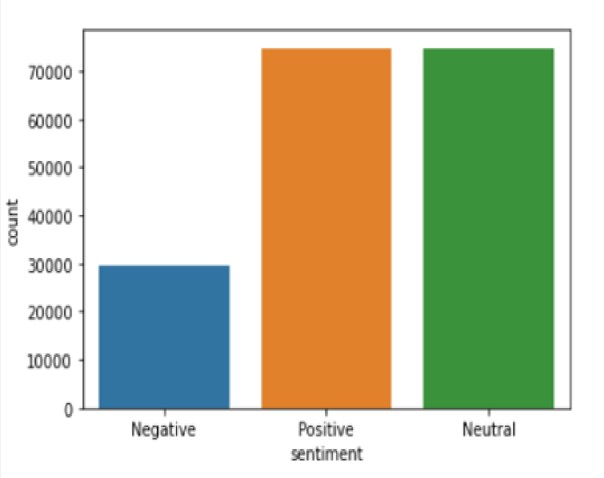

Fig 11: Public Sentiments during COVID-19 pandemic

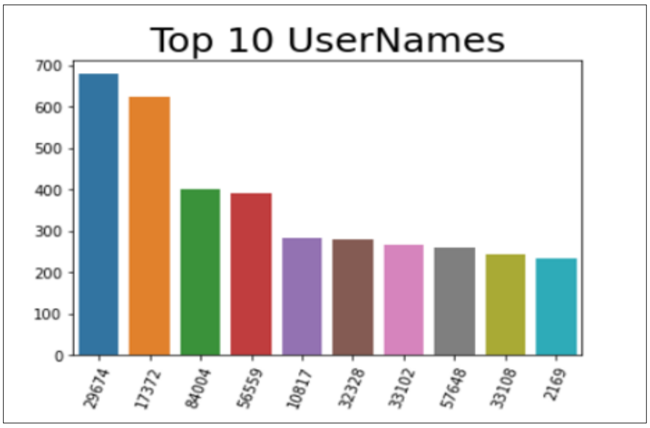

Fig 12: Top 10 Usernames

These are some of the most used words during COVID-19 pandemic among which the word "COVID-19" is used the most. It also shows the most used positive, negative and neutral words used during the pandemic [18]. Which helps us understand the sentiments, emotions and feelings of people during the pandemic time. Refer Fig.13, Fig.14, Fig.15 given below.

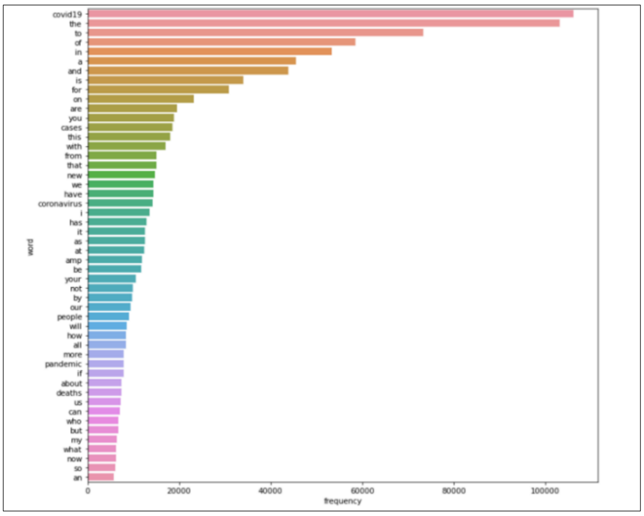

Fig 13: Most common used words

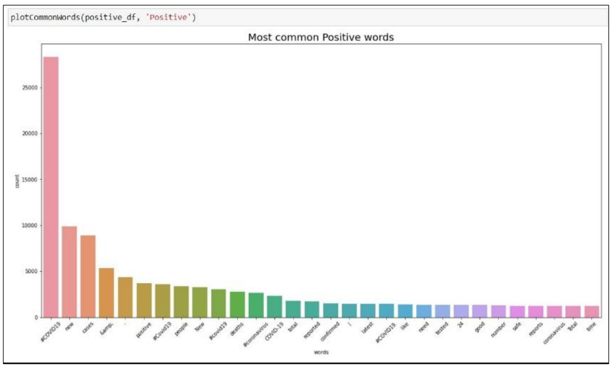

Fig 14: Most common Positive words

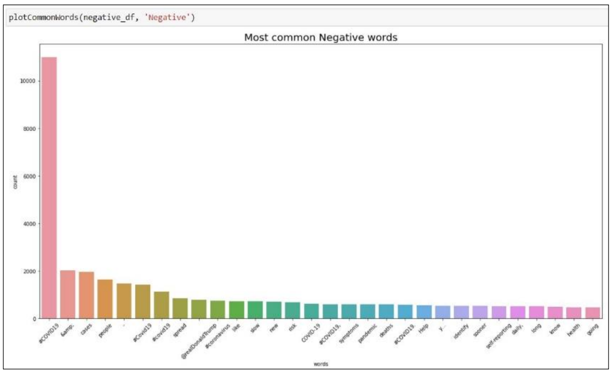

Fig 15: Most common Negative words

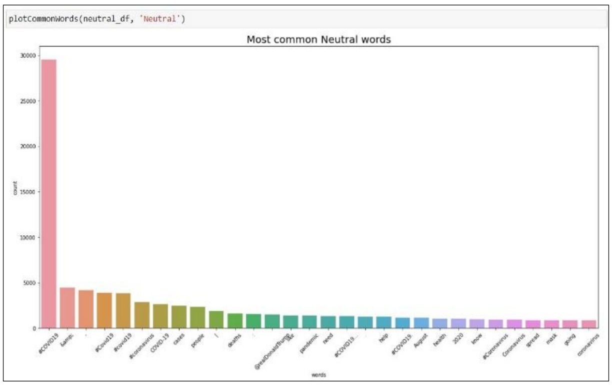

Fig 16: Most common neutral words

The goal of sentiment analysis is to determine whether the overall sentiment expressed in the tweets is positive, negative, or neutral which is represented by these word graphs.

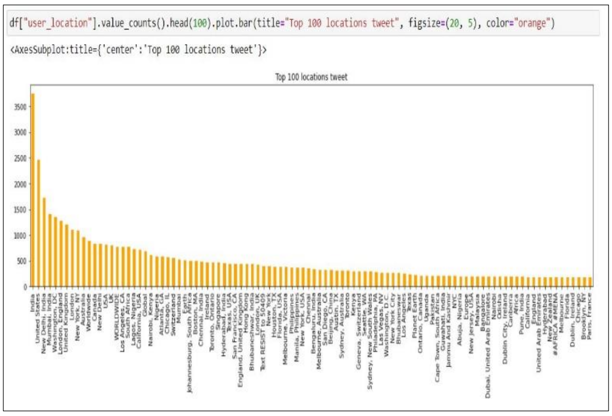

Fig 17: Top 100 locations trend

Countries affected the most during COVID-19 can be seen by the following graph. The countries with most tweets and related tweets to COVID-19 represents location of top affected locations or countries by COVID-19 [18].

Conclusion& Future Scopes

The research paper on the sentiment analysis of COVID-19 aims to analyze the opinions, emotions and attitudes expressed by people about COVID-19. The conclusion of such a study would depend on the methodology, data sources and analytical techniques used in the study. The methodology used in the study may involve collecting data from various sources such as social media platforms, online reviews, news articles or surveys. The collected data can then be analyzed using natural language processing (NLP) techniques to identify the sentiment (positive, negative or neutral) expressed by people regarding COVID-19.

A study conclusion can summarize key findings and highlight any notable trends or patterns observed in sentiment and topics discussed. For example, a study may find that people express more negative sentiment toward COVID-19 than positive sentiment. It can also identify the most common topics discussed, such as government response, health measures, vaccine effectiveness, etc. The study would also help vaccination process by looking at the trends and records. Implications of the findings for public health officials, policy makers, and other stakeholders in addressing the COVID-19 pandemic may also be discussed in the conclusion. For example, if a study finds that there is a high rate of vaccine hesitancy among certain groups, the conclusion may recommend strategies to increase vaccine uptake among those groups.

This research paper was focused on sentiment analysis and emotions people had during the COVID-19 pandemic it was successfully executed. To maintain credibility of the data as well as the ease of taking or extracting user tweets, Twitter platform was then selected for the study.

Because everyone was familiar with COVID-19 and recovery rates were increasing at the time, it was discovered during the study that nearly all nations tweeted about COVID-19 with favorable feedback. Similarly, it was discovered by analyzing word clouds from various nations that individuals tweet words like virus, COVID, health, pandemic, fight, remain safe, help, emergency, masks, hospital, and death with a range of emotions. We can now comprehend and accept that practically everyone in the globe thinks at roughly the same level thanks to the excellent sentiment analysis and people's thoughts on COVID-19 supplied by this study. In summary, the conclusion of a research paper on sentiment analysis of the pandemic (COVID-19) would depend on research question, data sources and analysis techniques used in the study. It would aim to summarize key findings, identify any notable trends or patterns, and discuss the indications and inferences of the findings for addressing the pandemic.

References

1. Kaushik P, Yadav R. Reliability design protocol and block chain locating technique for mobile agent Journal of Advances in Science and Technology (JAST). 2017;14(1):136-141. https://doi.org/10.29070/JAST
2. Kaushik P, Yadav R. Deployment of Location Management Protocol and Fault Tolerant Technique for Mobile Agents. Journal of Advances and Scholarly Researches in Allied Education [JASRAE]. 2018;15(6):590-595. https://doi.org/10.29070/JASRAE
3. Kaushik, P, Yadav R. Mobile Image Vision and Image Processing Reliability Design for Fault-Free Tolerance in Traffic Jam. Journal of Advances and Scholarly Researches in Allied Education (JASRAE). 2018;15(6):606-611. https://doi.org/10.29070/JASRAE
4. Kaushik P, Yadav R. Traffic Congestion Articulation Control Using Mobile Cloud Computing Journal of Advances and Scholarly Researches in Allied Education (JASRAE). 2018;15(1):1439-1442. https://doi.org/10.29070/JASRAE
5. Kaushik P, Yadav R. Reliability Design Protocol and Blockchain Locating Technique for Mobile Agents Journal of Advances and Scholarly Researches in Allied Education [JASRAE]. 2018;15(6):590-595. https://doi.org/10.29070/JASRAE