Social Media Sentiment Study on COVID-19 Outbreak

by

Anik Karmaker
17101294
Nayeem Mohammad Hafiz
17101244
Anir Chanda
17301121

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
June 2021

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Nayeem

Nayeem Mohammad Hafiz
17101244

Anik

Anik Karmaker
17101294

Anir

Anir Chanda
17301121

Approval

The thesis/project titled "Social Media Sentiment Study on COVID-19 Outbreak" submitted by

1. Anik Karmaker (17101294)

2. Nayeem Mohammad Hafiz (17101244)

3. Anir Chanda (17301121)

Of Spring, 2021 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on June 08, 2021.

Examining Committee:

Supervisor:
(Member)

Mostafijur Rahman Akhond
Lecturer
Department of Computer Science and Engineering
Jashore University of Science and Technology

Program Coordinator:
(Member)

Md. Golam Rabiul Alam,PhD
Associate Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi
Associate Professor and Chairperson
Department of Computer Science and Engineering
Brac University

Abstract

On 11th March 2020, the World Health Organization announced the COVID19 outbreak as a pandemic. Starting from China, this virus has infected and killed thousands of people from Italy, Spain, the USA, Iran, and other European countries as well. While this pandemic has continued to affect the lives of millions, several countries have resorted to complete lockdown. During this lockdown, people have taken social networks to express their feelings and find a way to calm themselves down. For accurate measurement of awareness of the people, it is necessary to successfully categorize the dataset of social media. The objective of this study is to use different data from twitter and filter these data for awareness measurement and to develop a model for evaluating the awareness among the people of all around the world by analyzing the collected social media opinions. In this research work, we have collected the data in a JSON file format and extracted the data into various criteria. We will parse the JSON file format data to the CSV file format and clean the data to use in our model. We will take English language-based data only. Then we will use the algorithm TextBlob to analyze the social media sentiment. We will finally apply that method to determine the awareness among the people and analyze that they are taking this pandemic very seriously or not. We will also analyze the graphical representation of our special keyword dataset. The result indicates that the methodology can be used to determine people's awareness and give us an idea about the sentimental issues of people. It will also help the government to take necessary steps for example psychological campaign, counseling program to make the people stronger to handle any pandemic.

Keywords: COVID19, corona virus, Twitter tweets, Job, Education, Medical, TextBlob, Matplotlib, Tweepy, Natural Language Processing, Sentiment analysis, Text Polarity

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis have been completed without any major interruption.
Secondly, to our supervisor Mostafijur Rahman Akhond sir and our co-supervisor Dr. Mohammad Zavid Parvez sir for their kind support and advice in our work. They helped us whenever we needed help.
And finally to our parents without their throughout support it may not be possible. With their kind support and prayer we are now on the verge of our graduation.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

API Application Programming Interface

$BERT$ Bidirectional Encoder Representations from Transformers

$COVID-19$ Corona Virus Disease of 2019

$HTML$ Hypertext Markup Language

ICU Intensive Care Unit

$IEDCR$ Institute of Epidemiology,Disease Control and Research

NLP Natural Language Processing

URL Uniform Resource Locator

USA United States of America

WHO World Health Organization

Chapter 1

Introduction

1.1 Introduction

COVID-19, a new corona virus strain, was discovered in Wuhan, China in December 2019. The World Health Organization declared the COVID-19 outbreak a pandemic on March 11, 2020. COVID-19, a highly infectious virus, has crossed borders at a pace that no one could have predicted, turning the world upside down. COVID19 has already claimed the lives of over 0.3 million people, with approximately 6 million people infected in over 200 countries or territories by May 31, 2020. It has wreaked such havoc on the global economic and social structures that it is difficult to imagine life returning to normal. Several countries have resorted to full lockdown as a means of controlling the situation. When the entire planet is infected with COVID-19, the lockout has become the only way to maintain power (Ghosal et al., 2020)[1].

Since then, there has been a lot of discussion about how the countries can ease the lockdown policy. Since lockout isn't the best option, relaxing the controls is the only way to keep the economy running and social functioning. Reopening government decisions is a big step to take because the planet is still not free from the deadly clutches of the corona, which continues to claim human lives. Via classified information, politicians and the general public can gain a clearer understanding of the situation by polling public opinion on the reopening decision. This COVID-19 epidemic has not only wreaked havoc on the global economy (Fernandes, 2020)[2],, but it has also had a significant psychological effect. Social networking networks are one of the best sources for capturing human emotions (De Choudhury and Counts, 2012)[3]. It is the quickest way for people to express themselves, and as a result, the news stream is saturated with data showing the current state of people's minds. People use these media to disseminate knowledge as well as share their opinions. During the lockdown, people have turned to social media to share their emotions and thereby find a way to distress. This outbreak will last 3-4 months.

Corona has caused global consternation Many pandemics have occurred in history, but this one is causing serious economic problems on a national and micro level. Individuals can develop psychotic symptoms as a result of the pandemic, and nations may undergo economic stagnation as a result of people being restricted from traveling, as all economic activities have been shut down, as well as social distancing. By mid-August 2020, nearly 773,072 people worldwide had tested positive for COVID-

19. COVID-19 is now spreading rapidly, especially in countries such as the United States and India. COVID-19 has impacted over 215 countries and will continue to do so until August 18th, 2020. As of August 17th, 2020, the top ten countries seriously impacted by COVID-19 are the United States (5,566,632 patients), Brazil (3,340,197 patients), India (2,647,663 patients), Russia (922,853 patients), South Africa (587,345 patients), Peru (535,946 patients), Mexico (522,162 patients), Colombia (468,332 patients), Chile (385,946 patients), and Spain (358,843 patients) [4].

The widespread use of social media platforms such as Twitter has sped up the process of exchanging information and expressing opinions about community events and health crises [5]–[8] . COVID-19 was a top trending topic on Twitter in January 2020, and it's still being discussed. People have become more reliant on different social media platforms to obtain news and share their opinions as more countries have implemented quarantine steps. Twitter data is useful for highlighting public conversations and emotions about hot topics, as well as real-time awareness of pandemics in the making. Several government entities around the world are using Twitter to communicate with the general public about the current COVID-19 pandemic [9]. Since the COVID-19 outbreak, a growing number of studies have used Twitter data to better understand public responses and conversations about COVID19 [10]–[15]. Abd-Alrazaq and colleagues, for example, use Tweets collected between February 2nd and March 15th, 2020, to pursue subject modeling and sentiment analysis to understand key issues and feelings surrounding COVID-19 [10]. Doctors and people who have been affected by epidemics are more likely to talk about them on social media sites like Twitter, which have become increasingly important in our daily lives.

Researchers from all over the world are sifting through Twitter data to see how people are reacting to corona virus-related issues like lockdown and protection, and safety measure. (Barkur et al., 2020) analyzed Twitter data and discovered that the citizens of India were supportive of their government's decision to flatten the curve by imposing a lockdown[16]. (Dubey, 2020) conducted an opinion survey on tweets about COVID19 from March 11 to March 31 in twelve separate countries: the United States, Italy, Spain, Germany, China, France, the United Kingdom, Switzerland, Belgium, the Netherlands, and Australia [17]. They discovered that, while most countries are adopting an optimistic outlook toward the situation, people are also experiencing negative emotions such as anxiety and sadness. In comparison to other nations, France, Switzerland, the Netherlands, and the United States of America have shown more mistrust and resentment. (Rajput and colleagues, 2020) They conducted a comparative study of Twitter data from the Coronavirus outbreak [18]. During February and March, the number of tweet ids tweeting about coronavirus increased exponentially, reaching several peaks. Coronavirus, Covid19, and Wuhan were the most frequently used terms in the texts, according to an empirical study. The massive amount of tweet messages sent over a span of 2-3 months, as well as the frequency of these phrases, clearly demonstrate the dangers to which the global population is subjected.

(Abd-Alrazaq and colleagues, 2020) The aim was to find the most popular topics on Twitter about the COVID-19 pandemic between February 2 and March 15,

2020. Users on Twitter discussed 12 main topics related to COVID-19, which were divided into four main themes: 1) the virus's origins; 2) its sources; 3) its impact on people, countries, and the economy, which was represented by six topics: deaths, fear, and stress, travel bans and warnings, economic losses, panic, increased racism, and 4) ways of reducing the risk of infection [19]. Cinelli et al. (2020) conducted a comparative study of knowledge diffusion about COVID19 on five different social media platforms: Twitter, Instagram, YouTube, Reddit, and Gab [20]. They decided to look at how knowledge is disseminated during a crisis. Their findings indicate that the spread of information is more dependent on the engagement patterns of users who are interested in the subject than on the information's reliability.

In [21], they used coronavirus-specific tweets to identify public sentiment and provide insights into how fear sentiment developed over time, especially when COVID-19 reached peak levels in the United States. As they classified tweets using sentiment 2 analysis, they discovered that fear was the most prevalent emotion across the board, and that its seriousness became apparent by the end of March, as the fear curve showed a steep increase.However,they did not focus much on people sentiment regarding Covid's effect to their life based on certain factors like-job,health,economy,medical appliances and others.So,there are further scopes to study in these areas.

Since there are further scopes to research on the social media sentiment analysis on the Coronavirus issue we thought of doing the followings: 1.We collect data from social medias. 2.We use the latest technologies of NLP to utilize the data. 3.We analyze the data to find out what has been discussed over the social media over the sub-continent specially on social media. 4.We figure out some certain factors and find out what people have been discussing about those in the social media. 5.Then we figure out the rate of positively and negativity of those discussion among the people. 6.Finally,we will make a more detailed analysis from here.
The structure of this paper is organized as follows. Chapter 2 presents literature review. chapter 3 contains the system model. Chapter 4 describes the data collection. Chapter 5 describes result analysis. Finally, Chapter 6 concludes the paper with future works.

Chapter 2

Literature Review

Machine learning is the analysis of various types of algorithms and statistical data that a computer system employs to perform various tasks without the use of specific instructions and instead relying on patterns. To make decisions for a specific task, machine learning algorithms build a mathematical data model using different sample data, referred to as "Training Data." Artificial Intelligence is also regarded as a subset of it. To perform various tasks, different types of machine learning methods are used. Algorithms for machine learning are used in a wide range of applications. Machine learning algorithms are linked to statistical statistics, so it focuses on making various forms of predictions with computers. Machine learning algorithms have a wide range of uses in today's scientific world. Various forms of robotic jobs, automated personal assistants, video monitoring, early forecasts, and social media platforms are examples of machine learning applications. A machine learning algorithm can be applied in a variety of ways. A nave Bayes classifier is a "probabilistic classifier" that uses the Bayes theorem to classify features with clear independence assumptions. Nave Bayes and Kernel density estimation can be combined to improve accuracy. In a supervised learning process, Nave Bayes classifiers can be trained to perfection. The maximum likelihood method is used as the parameter assumption for Nave Bayes, which means that anyone can work with this model without acknowledging Bayesian probability or using any Bayesian method. K-means clustering is a signal processing method of vector quantization that divides n observations into k clusters, each of which corresponds to the cluster's closest mean and serves as a model. The data is partitioned into Voronoi cells as a function of the result. In data mining, it is a common method for cluster analysis. NLP (Natural Language Processing) is a branch of computer science concerned with the interaction of a computer and human language in order to process vast amounts of data. It facilitates the extraction of data and keywords.

COVID19 is causing a lot of headaches these days. COVID19 is an infectious disease that was first recorded in Wuhan, Hubei, China in December 2019 and is now a global pandemic, according to Wikipedia. This disease was only found in China at the time. However, it is gradually being adopted by the majority of the world's countries [22].

More than 188 countries and territories were affected in October 2020, with 33.3 million confirmed cases. More than 1.01 million people died as a result of this coronavirus, but 23.6 million people were recovered. 1st This pandemic spread quickly across Europe. The reported cases and death cases, according to the WHO (World Health Organization), are listed below [23]:

Regions	Confirmed Cases	Death Cases
Americas	16,624,745	555,956
South-East Asia	7,071,811	115,583
Europe	5,822,105	236,512
Eastern Mediterranean	2,405,123	61,910
Africa	1,182,927	25,881
Western Pacific	614,623	13,415

Table 2.1: Confirmed cases and death cases

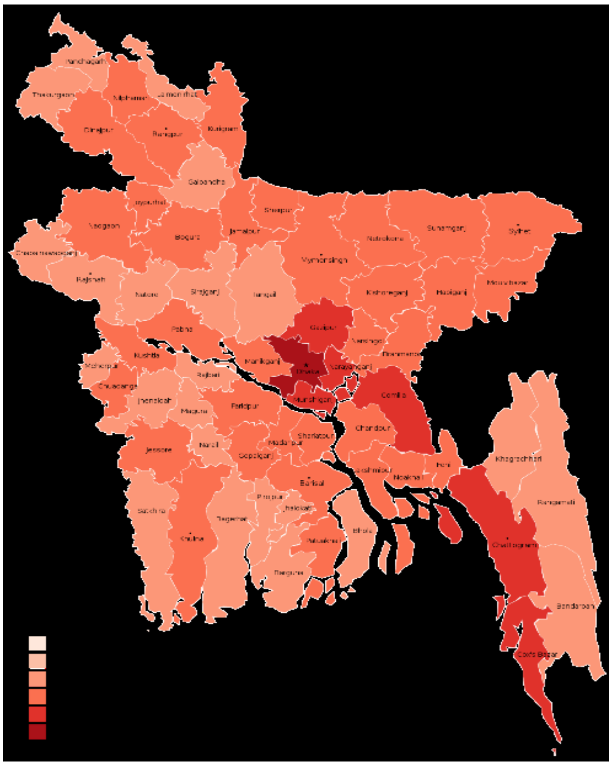

Figure 2.1: Affected area in Bangladesh

Cases that have been confirmed and cases that have resulted in death are shown in Table 2.1 , Figure 2.1 World Health Organization (WHO) (1 October 2020) Google is the source of this information.
This pandemic has also struck Bangladesh. There is a piece of information on Wikipedia that states when coronavirus was confirmed in Bangladesh. IEDCR (Institute of Epidemiology, Disease Control and Research) announced the first three confirmed cases on March 8, 2020 [24].

Since they were unaware of the disease at the time and interacted with a large number of people, the pandemic spread throughout the world. As a result, the virus spread to everyone. Bangladesh has conducted a total of 19,59,075 studies, with 3,64,987 confirmed cases as of October 1, 2020, according to Worldometers. There were a total of 5,272 deaths and 2,77,078 bodies recovered [25].

Bangladesh is improving in terms of technological support, and the government is digitizing all information. We use social media sites such as Facebook, Twitter, and Instagram, as well as a variety of blogging platforms. People share their perspectives on the current pandemic on social media, making it a good source of knowledge. As a result, we've decided to use social media to gather data and conduct research on the pandemic.

Many researchers have completed studies on this pandemic using social media evidence. Lopez, et al. claimed in their paper that they gathered data in a variety of languages, such as twitter data from Twitter, and compared the differences in comments over time. They also investigated how facts and disinformation was disseminated through Twitter. To compare the variations, they used Natural Language Processing, Text Mining, and Network Analysis [26].

According to Singh and colleagues, there was a lot of discussion on social media. As a result, they concentrated on very specific aspects that aided their study. They looked at the total number of social media conversations, as well as the topics discussed, the central location of these conversations, and the links between these conversations and other high or low quality information on the internet through shared URLs. They discovered a Spatio-temporal relationship between the long-running conversation and the high- or low-quality details, as well as a Spatio-temporal relationship between the conversations and the new cases.

They gathered data and knowledge from a number of outlets, including blogs, journals, social media, and medical organizations. They used two sets of data from Twitter. The first collection was created using Twitter data and grouped by country. For instance, if both the USA and covid19 were listed in the data, the tweet was categorized as belonging to the USA. This collection of tweets was used as a conversation location post. Their second collection of data consisted of geotagged tweets. Users who used the longitude and latitude position as a tag were able to access these tweets using the Twitter API. They have used the WHO's final data package (World Health Organization). They gathered common domains to evaluate mutual connections, and there are over 60,000 domains that people share in their tweets. They also gathered data of varying quality from shared URLs. They compared the data to arrive at their final conclusion. Their job is broken down into many parts [27].

The COVID19 pandemic has elicited both positive and negative reactions on social media. Karisani et al. conducted their research using data from the Twitter API. They used user postings on social media, with the intention of using machine learning and linguistic resources to better understand the effects of the pandemic in China. They used a cutting-edge machine learning model to assess the positive COVID19 reports. They gathered the data using the Twitter search API, with the keyword ”keyword” as the initial search term. At first, they used the words ”coronavirus” and ”corona virus.” However, they searched their data using the terms ”COVID-19” and ”COVID 19 keyword” in mid-March.

On Twitter, there is a wealth of information in a variety of languages. As a result, they only received English-language tweets, except retweets and responses. In their experiment, they used seven different machine learning algorithms. Four models based on the state-of-the-art model Bidirectional Encoder Representation from Transformers were used: the Naive Bayes model, Logistic Regression, commonly used Neural Network model, and four models based on the state-of-the-art model Bidirectional Encoder Representation from Transformers. BERT-Corona has the F1 meaning, according to their findings. As a result, they discovered that model initialization through pretraining is more effective than increasing model complexity [28].

People’s mental health is not healthy due to the coronavirus pandemic. As a result, several tweets and sentiments have emerged during the early stages of every pandemic. Medford et al. used Twitter behavior, content, and sentiment to calculate and identify early changes in the COVID19 pandemic. Their research was an observational study, and they collected data using the Twitter platform. They highlighted the hashtags and calculated the frequency of keywords associated with illness prevention, vaccination, and racial discrimination. After gathering the information, they converted it to plain text and used sentiment analysis to determine the emotional boundary (positive, neutral, or negative) and prevailing emotion for each tweet (anger, fear, joy, etc). They used a Latent Dirichlet Allocation model, which creates contexts automatically based on the distribution of terms. Fear was the most commonly used word in nearly 49.5 percent of all tweets, according to their findings. “Surprise” was the second most popular term, appearing in 29.3 percent of all tweets. As a result, they discovered that tweets with a negative emotion increased at a faster pace day by day than neutral and optimistic tweets [29].

Human psychology and behavior modification as a result of the COVID19 pandemic, according to Arpaci et al. They looked at 43 million tweets from Twitter to demonstrate the public’s interest in COVID19. The most frequently used terms were ”death,” ”test,” ”spread,” and ”lockdown.” This condition reflects people’s internal fear of being infected and contracting the infection. As a result, they hypothesized that social media messages may have the greatest impact on human psychology and actions. To find the most frequently used words on Twitter, they used MATLAB and clustering algorithms. They concluded that social media posts and comments can have a significant impact on human emotions using this tool, and that this finding can aid in the development of a more successful communication strategy in the event of a pandemic [30].

To predict the COVID19 outbreak, Jahanbin and Rahmanian conducted research using twitter and web news mining. They clarified that Twitter's unstructured data were segregated and then subjected to text washing, filtering scanning, and classification operations. To get their final result, they used Eclass1-MIMO, a fuzzy rule-based evolutionary algorithm. They obtained the user's exact demographical and geographical position, which aids in predicting anguish rates in each location and drawing policymakers' attention to the need to improve the health-care system and minimize community deaths [31].

Samuel et al. reported in their paper that they used the core twitter posts and R statistical tools, as well as its sentiment analysis packages, to identify public sentiment relevant to the pandemic. As the spread of coronavirus infection increased, they noticed an increase in the rate of fear. To classify coronavirus tweets, they used two machine learning algorithms. For short tweets, they achieved a classification accuracy of 91 percent using the Naive Bayes system, and a classification accuracy of 74 percent using the logistic regression model. They had a terrible time with long tweets. Their study concluded that more Coronavirus data is needed to construct a reliable and empirical model. They also mentioned that their model and results could be applied to a possible local and global pandemic [21].

Shammi et al. conducted their study from the perspective of Bangladesh. They looked at how people reacted to the long lockout scenario and how it affected COVID19's long-term growth priorities and strategic management regime. They used circumstances that were based on scenarios. As a result, they conducted a survey in order to create a dataset based on some questionnaires. They used statistical techniques such as classical test theory, concept component analysis, hierarchical cluster analysis, Pearson's correlation matrix, and linear regression analysis to analyze the data [32].

As we previously stated, social media is the most effective way to gather information about people's opinions and expressions on any subject. As a result, Kaur et al. looked at social media data to see how people felt about COVID19. They used the Twitter API to collect tweets about COVID19 and analyzed them using a machine learning algorithm. They also use the tools of positive, neutral, and negative. They used Python programming on the Twitter API and NLTK (Natural Language Toolkit) for tweet preprocessing, evaluating the dataset with TextBlob, and displaying the results as positive, neutral, and negative parameters. Their neutral sentiment percentage was approximately 43.9 percent, negative sentiment was 32.1 percent, and positive sentiment was approximately 24 percent. As a result, their final result was a neutral sentiment [33], which was significantly higher than the other percentages.

The value of Twitter in social media was claimed by Sarlan, et al. They conducted sentiment analysis research. Twitter is a quick and effective way to gather data. It provides a method for analyzing people's attitudes toward the impossible to achieve. As a result, they extracted a large number of tweets from Twitter and used their prototyping for this project. They divided their findings into positive and negative categories, which they displayed in a pie chart and HTML tab. Their software also intended to create a web application, but due to Django's limitations, they decided to postpone this for now [34].

Chapter 3

Proposed System

Social media is a platform that can help us to visualize people's opinion and sentiment based on the current situation. As we have decided to do a research on twitter sentiment analysis so we selected twitter for collecting our data. There are a huge number of users on twitter around the world. So, we can get a vast amount of data for doing our research. People share their opinions and views on a daily basis on twitter about COVID19. We collected data from twitter and implemented those data to extract positive, negative and neutral sentiments among people over time.

3.1 Platform Design:

After doing problem identification and background study we found some way to do the progress of our study

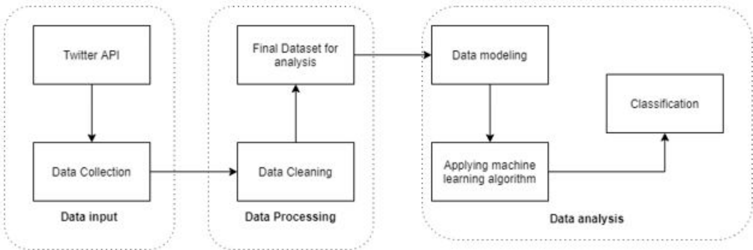

Figure 3.1: Proposed Model

We gathered the approaches together and designed our platform. We collected twitter API from twitter developer portal from our account. For accessing the data, we need consumer key, consumer secret key, access token and access token secret from twitter API. Twitter API is used to programmatic collect and analyze the data. Twitter API gives the access to collect different resources from twitter.

- Tweets
- Users
- Direct Messages
- Lists

- Trends
- Media
- Places

We requested the twitter developer to give access to those keys for gathering the data. After four days of request, we got the approval from the twitter developer and got the access keys. Then we collected the tweets from twitter and the tweets of data were from December 2019 to February 2021. We got huge chunks of data and stored the data into our data set. We cleaned the data for our future implementation. After cleaning all the inappropriate words and junk data we used the data is usable for analyzing the result. We implemented the processed data into the TextBlob for getting polarity of positive, negative and neutral. TextBlob is a Python library for processing textual data. It provides a simple API for diving into common natural language processing such as parts-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation and many more. Sentiment analysis gives polarity after analyzing the data and the polarity score is a range within (-1.0 to 1.0). Then we got positive, negative and neutral polarities from that analysis. Then we have counted the positive, negative and neutral texts. After counting the numbers of final data, we plotted the tweets over a time interval and ended up in a result.

Chapter 4

Data Preparation

4.1 Data Collection:

Coronavirus infection is increasing day by day. We can assume how much affected people are in the whole community. So, people have their own perceptions and opinions regarding this issue. We have decided to take these opinions as our dataset to analyze the people's awareness and the most frequent topics they have discussed. Twitter is the famous social media site where people from all around the world use this social media and share their views and opinion. These views and opinions are known as tweets. Twitter helps us to access its API through python Library called Tweepy. Tweepy helps us to extract data from twitter of any user. We collected the data from twitter using twitter API. We requested the twitter developer portal for consumer key, consumer secret key, access token and access token secret key. After getting the approval twitter developer portal generated those keys for us. To generate these keys, we had to follow some procedure:

1. Log in to Twitter developer site.
2. Create an App
3. Fill the details in the form
4. Create own Twitter Application
5. Details of the new twitter application will be shown with the consumer key and the consumer secrete key.

Tweepy: Tweepy is the open source which enables python to communication with the twitter and helps us to use its API. We imported a tweepy library of python and implemented the library in our code.

Python Programming Language: Python is a programming language that is popular and it is being used for text mining and data analysis. It is object oriented and high-level programming language.

We used these keys to collect our own data from twitter. We collected the total data from December 2019 to February 2021 time interval. We used many hashtags for collecting the data. For example: "covid19", "covid", "coronavirus", "stayhome", "lockdown", "pandemic", "covid19job", "social distance", "covid19medical" etc. After tracking all the hashtags, we got the data into a JSON format. Then we converted the JSON file format data into CSV file format.

```python
import json
import sqlite3
import tweepy
import time
from tweepy.streaming import StreamListener
from tweepy import Stream
from html.parser import HTMLParser
import json
# My Key
consumer_key = 'Enter Consumer key';
consumer_secret = 'Enter Consumer secret key';
access_token = 'Enter Access token key';
access_token_secret = 'Enter Access token Secret key';
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth)
save_file = open('Enter any name.json', 'a')
class MyListener(StreamListener):
    def __init__(self, api=None):
        super(MyListener, self).__init__()
        self.num_tweets = 0
    def on_data(self, data):
        try:
            save_file.write(data)
            save_file.write('\n')
            self.num_tweets += 1
            print(self.num_tweets)
        except BaseException as e:
            print("Error on_data: %s" % str(e))
        except AttributeError:
            print("tribute error nontype")
        return True

    def on_error(self, status):
        print(status)
        return True
twitter_stream = Stream(auth, MyListener())
twitter_stream.filter(track=['Enter name to keep track of data'])
```

Figure 4.1: Code snippet for twitter data collection

4.2 Data Cleaning

After converting the JSON file into CSV file format there are multiple error data, incomplete data, language problems, junk and inappropriate data. As this was a huge chunk of data, there were many different languages in the data and we considered English language so it is important to remove other data except English language. We filtered the data based on English language. The data may not have complete data for our research so we need to filter that data for example there is a userID, user name of the user but there is no valid information or captions that can help us to do our research. So, we excluded that data from our csv file. We also used regular expressions for cleaning our dataset. As there were many hyperlinks, special characters so we used simple regular expression statement to clean the dataset.

```
def clean_tweet(self, tweet):
    return ' '.join(re.sub("(@[A-Za-z0-9]+)|([^0-9A-Za-z \t])|(\w+:\/\/\S+)", " ", tweet).split())
```

Figure 4.2: Regular expression for data cleaning

4.3 Data Processing

We extracted tweets from out dataset. Then we used our tweets from the dataset to do sentiment analysis for getting positive, negative and neutral polarity using TextBlob.

4.4 Sentiment Analysis

Sentiment analysis helps us to consider the mood and emotions of the people and gather insight information based on the situation. We can get an assumption of the sentiment of people based on social media context. TextBlob library is an approach to analyze the sentiment of people. TextBlob is a python library for natural language processing. Natural Language Processing gives an easy way to get a log of lexical resources and allows users to work with the categorization, classification and other tasks. TextBlob is a simple library that does complex analysis and operations on textual data. For lexicon-based approaches, a sentiment is defined by its semantic orientation and the strength of each word in the sentence. The text message is represented by some bag of words to calculate the sentiment. This method needs a pre-defined dictionary which classifies negative and positive words. After giving separated scores to all the word clouds, final sentiment is calculated by some pooling methodology like calculating the average of all the sentiments. TextBlob returns polarity and subjectivity of a sentence. The polarity varies between (-1,1). So, it means that the negative sentiment will be represented as -1 and the positive sentiment will be represented as 1. If the polarity is 0 then it means the sentiment is neutral. Subjectivity calculates the amount of personal perspective in the text of a person. If the subjectivity gets higher value, then the text contains personal opinion rather than factual information. So, we implemented the TextBlob python library to calculate the positive, negative and neutral polarity of our total dataset. We imported the TextBlob python library and implemented a method that calculates the polarity of our dataset.

```
from textblob import TextBlob
for m in tweets:
    tem = 3
    analysis = TextBlob(m)
    if analysis.sentiment.polarity > 0:
        tem = 2
    text_pos = text_pos + 1
    elif analysis.sentiment.polarity == 0: \
        tem = 0
        text_confused = text_confused + 1
elif analysis.sentiment.polarity < 0:
        tem = 1
        text_neg = text_neg + 1
```

Figure 4.3: Code snippet of using TextBlob

Then we searched for the special keywords from the tweets and it is important for us to get our result. After searching special keywords for example: "Job", "Medical", "House rent", "Death tolls", "Education", "Poverty" etc. we calculated the positive, negative and neutral tweets from our total tweets and extract those calculation to plot in python matplotlib library. After plotting the tweets against different time intervals, we got different types of graphs.

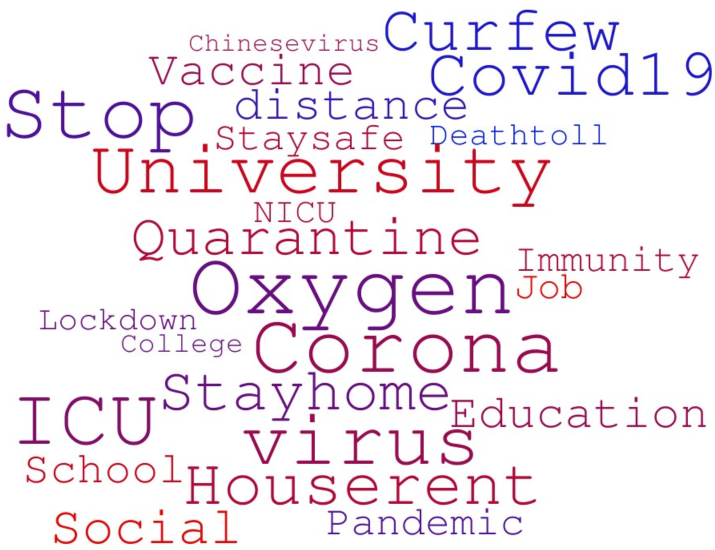

Figure 4.4: Wordcloud of special keywords

Chapter 5

Result Analysis

We implemented the TextBlob library on our dataset of tweets and got the polarity of positive, negative and neutral. We also calculated the positive, negative and neutral tweets to plot the graph and determine the variation of sentiment of the people in the social media. Data presented in a visual format makes greater sense than data presented in a textual format. Data visualization in the form of graphs is more presentable and aids in the comprehension of facts and figures better. Matplotlib is a python library for charting(2D). It generates quality of figures in a variety of formats. It also aids in the creation of various graphs for the viewing of data.

```
import matplotlib.pyplot as plt
import pandas as pd
df = pd.read_csv('enter csv file')
plt.plot(df.date, df.total_tweets, label = "total tweets")
plt.plot(df.date, df.positive_tweets, label = "postive tweets")
plt.plot(df.date, df.negative_tweets, label = "negative tweets")
plt.plot(df.date, df.neutral_tweets, label = "neutral tweets")
plt.title("title of the graph")
plt.xlabel('date')
plt.ylabel('total tweets')
plt.legend()
plt.show()
```

Figure 5.1: Code snippet of line using Matplotlib

Our specific keywords help us to get the different types of graphs. The specific keywords are:

• **Jobs:** Jobs are affected because of coronavirus. At the beginning of the pandemic no one thought that jobs would be affected by coronavirus. Many people across the world lost their jobs. Businesses did not run accordingly because of the lockdown throughout the world. Many countries faced very bad situation due to the increase of covid19.

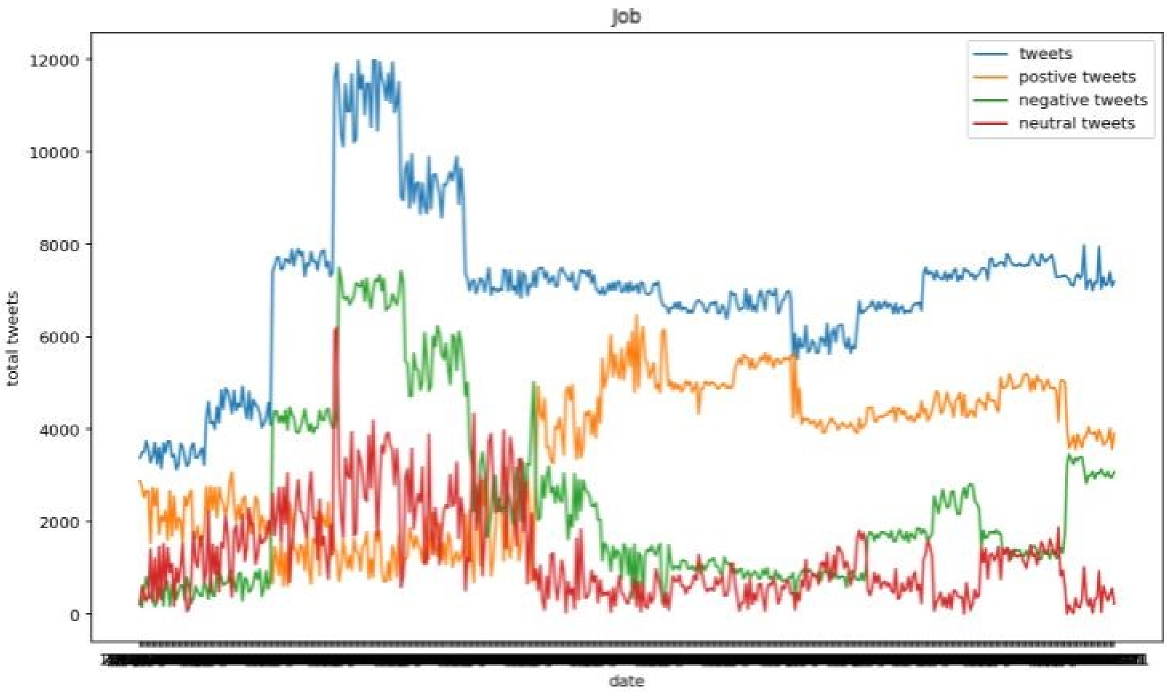

Figure 5.2: Graph of tweets over time due to have an impact on job

Figure 5.2 shows the differences of the positive, negative and neutral sentiment throughout the 1 year time interval. The time duration of the collected tweet is from December 2019 to February 2021. The X-axis of the graph represents the time interval and y axis represents the total tweets. The blue color line represents the tweets, the orange color line represents the positive tweets of the total tweets, the green color line represents the negative tweets and the red color line represents the neutral tweets.

So, we can see from the graph that at the beginning of the pandemic the number of tweets were very low. When the time increases the number of tweets increases. For example: In December 2019 the positive tweets were more than the negative tweets. So, the positivity among the people was higher than the negative sentiment and the orange line is higher than the all-other colored lines.

In January 2020 also the positivity was higher than negative number of tweets. Mostly people were either positive or neutral about this issue in this month.

In February 2020, people's negativity grew more and more. So, in this time period people were mostly negative or neutral regarding this issue. When the situation was getting too bad in March-April 2020 and people were losing their job then the negative sentiment increased. We can see from the graph that after some time interval the negativity increased which can be represented by the green line from the graph. This means that the negative sentiment was increased in that time among the people because of losing their job due to Covid19.

Again, in the middle of the year people were mostly happy about the job issue as things were coming back to normal slowly and job sector was also coming back to its old form. So, people were turning back to be happy about the job issue and they were gaining back hope.

In the December 2020 to January 2021 there is a increment of the positive line. So, we can say that positive sentiment and tweets were raised in that time among the people. The neutral sentiment means the polarity of the tweets of people was 0. Some people took that situation as normal.

• **Education:** Education sector was also affected by the coronavirus. In most of the country's schools, colleges and universities are still closed because of the coronavirus. For example: we can see that Bangladesh has also stopped the education sector for 1 year. The government officials still closed the school, colleges and universities in the concern of the students. The other countries also stopped the education sector for increasing Covid19.

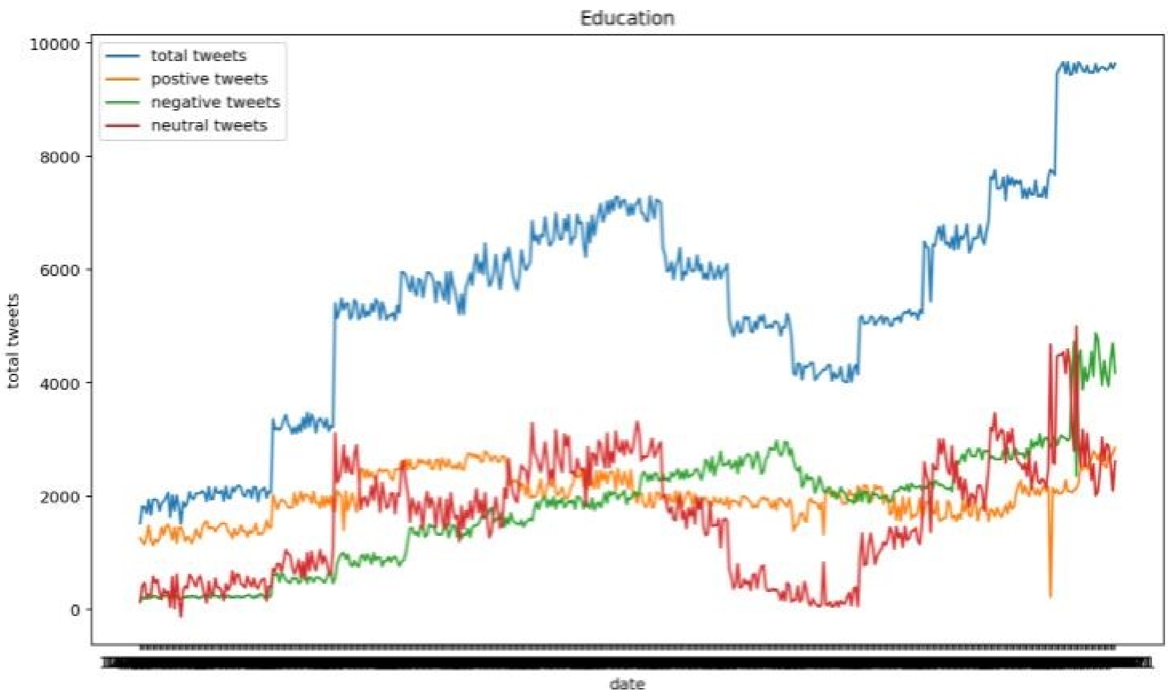

Figure 5.3: Graph of tweets over time due to have impact on education

Figure 5.3 shows the differences of the positive, negative and neutral sentiment throughout the 1 year time interval from December 2019 to February 2021. The X-axis of the graph represents the time interval and y axis represents the total tweets. The blue color line represents the tweets, the orange color line represents the positive tweets of the total tweets, the green color line represents the negative tweets and the red color line represents the neutral tweets.

So, we can see from the graph that at the beginning of the pandemic the number of tweets was very low. When the time increases the number of tweets increases. At the beginning of the graph there was a high rate of positive and neutral sentiment polarity among the people.

When the lockdown was announced we can see in the middle of the graph (February 2020 to April 2020) the tweets increased more than before and positivity also increased. People thought that this lockdown was a right decision and they took that decision as a positive decision all over the world.

But in the middle of the year 2020, people was turning to have negativity among them as they were totally unaware about what was going to happen. As, being an under-developed country, it was unable to conduct online teaching at a full pace so most of the students were lagging behind which made people worry that much. Moreover, it was far easy for city people to get access to fast and secured internet but it was very tough for the students living in rural area to get the same benefits as city people. So people were thinking that the education system was having a huge drawback and that made them worry that education may rise up as a huge trouble in upcoming future. So, that created an effect on their sentiment for which we are being able to notice a rise on negative polarity at the middle of the year.

But in the last of the graph in September 2020 to October 2020 the negativity and neutral opinion was increased in that time intervals. The green line and the red line raised high in that time interval and it represents that the negativity and neutrality were increased in that time among the people and they posted negative and neutral tweets both in the context of lockdown and covid19.

• **Medical:** Medical sector has a huge impact for Covid19. It is one of the main sectors that can help to save the affected people and decrease the death of the covid patients. Medical sectors are mainly called hospitals, doctors who are known as the front liners, oxygen availability, ICU's of the hospital, medical facilities. page

Figure 5.4 is the people's sentiment on the medical sector due to the Covid19 situation against the time interval. The X-axis represents the date and the Y-axis represents the total number of tweets. The blue color line represents the tweets, the orange color line represents the positive tweets of the total tweets, the green color line represents the negative tweets and the red color line represents the neutral tweets.

At the beginning of the graph, we can see that the opinions were positive, nega-

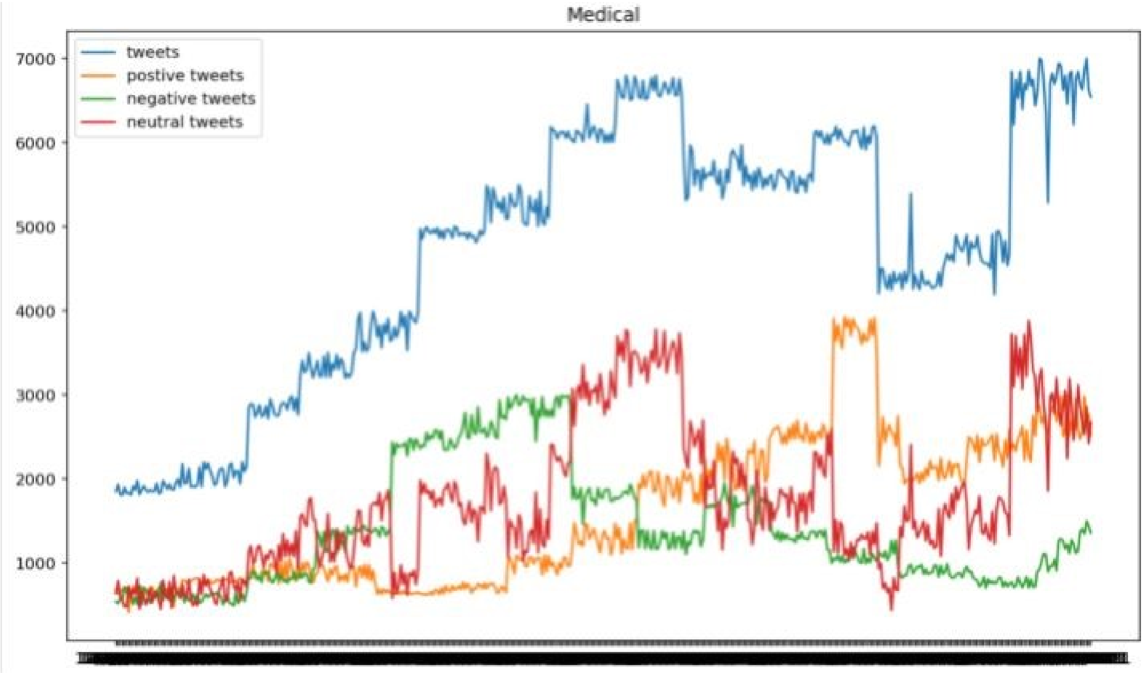

Figure 5.4: Graph of tweets over time of medical sector

tive and neutral. So, all the three colored lines are closed among each other. In the January-February month things were also quite the same as the scenario did not change drastically. Still then, the mass people did not have the slightest idea about the things that were waiting for them. They were thinking that this disease may go away very soon but their thoughts were not on the track.

Covid was waiting for them with more surprises. Later on, Covid helped us to realize how terrific we were on our medical sector. The medical sector of South-Asia region was not even slightly prepared for Covid. We never thought what will happen to our health sector if such pandemic arises. So we did not put any emphasis on it. As a result, when Covid stroke hard on us the whole infrastructure fall down and as a result we lost a significant amount of lives which could have been save to some extent.

There were certain issue in the medical sector which was needed to be worked upon but in January-February things were not such drastic like later on so people were less bothered about it as they still did not have to face the consequences. But when there was a rush of death in March 2020 to July 2020 in all over the word and in south Asian region then the negativity increased for the medical sector among the people on twitter social media. They thought that the medical sector was not capable of handling this situation around the world. Especially if we see the condition of South Asia then we can understand the bad condition of the hospitals.

When the capability of the hospitals and medical sectors increased in November 2020 to December 2020 then the opinion was positive and the positive curve was rising high. For example: If the hospital has the good facilities of oxygen and ICU's or there was less rate of death of covid patients then we can consider the tweet as positive. If the situation stands against the condition, then the sentiment will be negative and the curve represents the negative sentiments of the people and if the negative tweets increase then the curve will also rise according to the tweets.

• **Death toll:** Death toll is known as the number of deaths from a particular cause. If the death toll increases then the people will be more panicked and it will give an effect on the social media posts. There will be more opinions regarding this topic. It is also dependent on the medical sector.

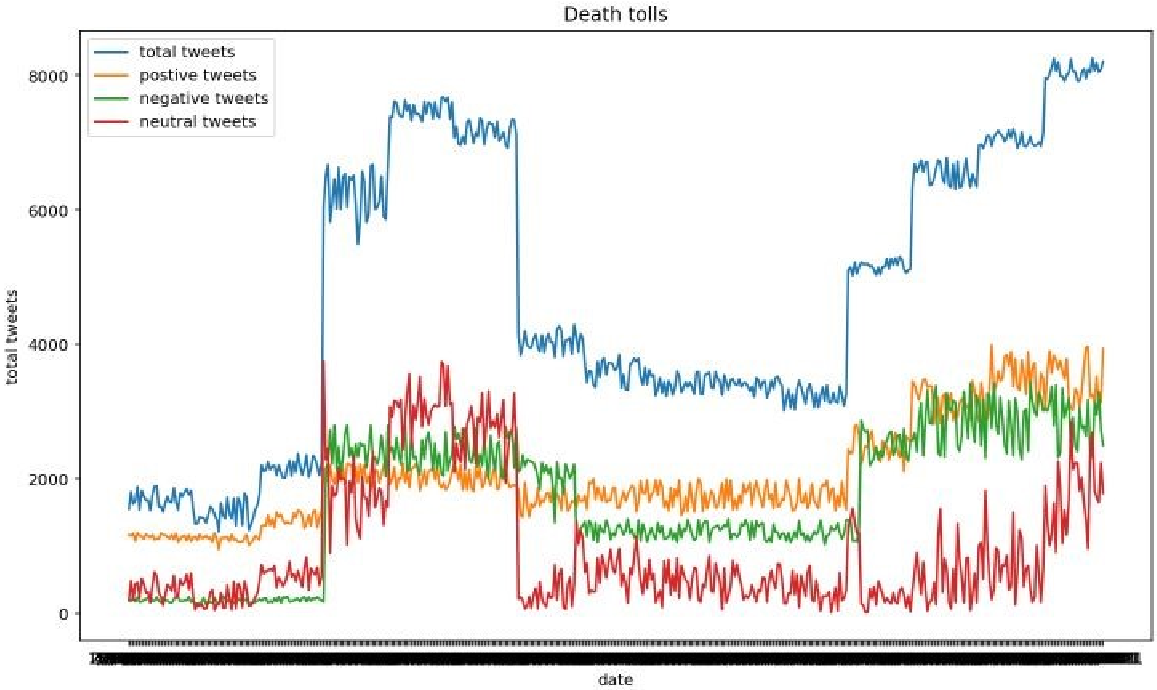

Figure 5.5: Graph of tweets over time due to Death tolls

Figure 5.5 is the people's sentiment on the death tolls because of the Covid19 situation against the time interval. The X-axis represents the date and the Y-axis represents the total number of tweets. The blue color line represents the tweets, the orange color line represents the positive tweets of the total tweets, the green color line represents the negative tweets and the red color line represents the neutral tweets. When the pandemic started in December then the people's perspective was very positive and people thought that the disease was nothing but a normal disease. So it will not have any impact on people's health. So, the sentiment of the people was very positive and neutral.

In Jaunary the number of tweets was same as December. The perspective of people all around the world did not change. Actually, the disease was new to everyone. No one knows about it. So, they did not understand what will be the next situation. They shared their positive perspectives in the social medias.

But after increasing the death of affected patients in the mid of the year (March 2020 to May 2020) people's opinions and views were changed in the social media.

The negative line was raised higher than the positive curve.

In the month of June, the rush of the patients death was not improved and people were seeing massive death all around the world. So, the green color line which represents the negative sentiment was raised higher than the other lines. People were sharing their negative views about Covid19 in the social media.

If we see the graph then we can visualize that the positive and negative lines are close enough at some point in November 2020 to January 2021 time intervals.

• **House rent:** House rent has also affected by the coronavirus situation. In the South Asia, many people lost their house. As there was lockdown in many countries and people did not able to continue their business and they had no income. So, they could not pay their rent and because of that situation they lost their house. Covid19 has affected many parts of the general life.

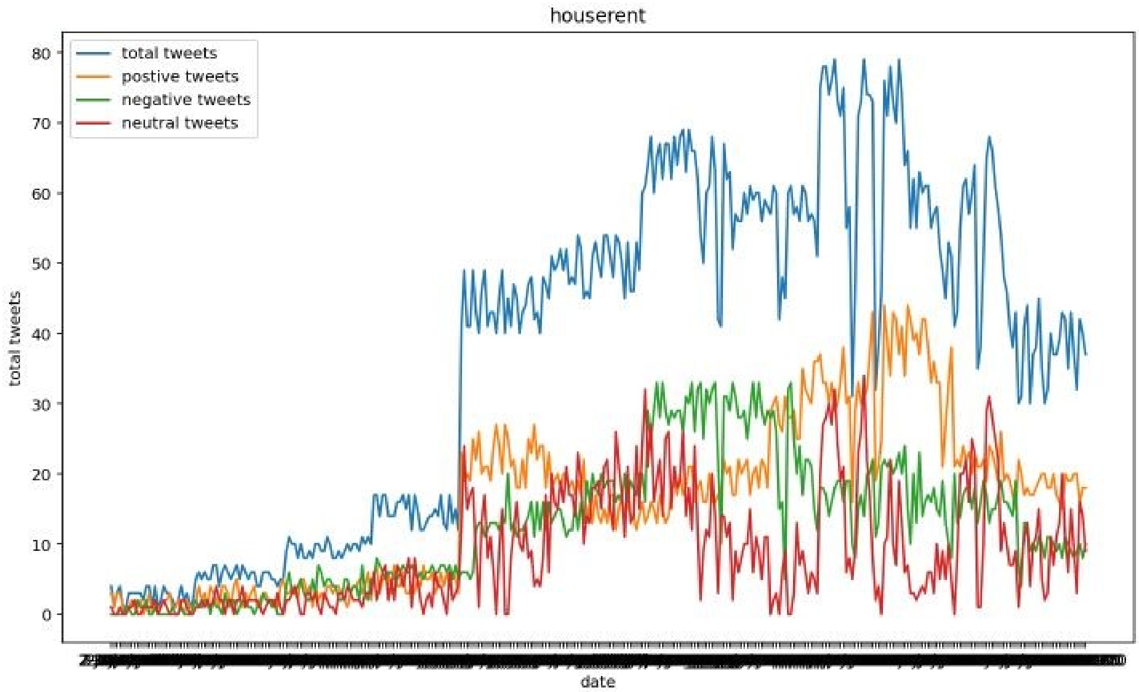

Figure 5.6: Graph of tweets over time due to house rent

Figure 5.6 represents the house rent over time intervals. The X-axis represents the date and the Y-axis represents the total number of tweets. The blue color line represents the tweets, the orange color line represents the positive tweets of the total tweets, the green color line represents the negative tweets and the red color line represents the neutral tweets. As the total tweets are not more enough in twitter but we saw many posts and opinions in other social media and blogs. So, we considered this parameter as a matter of concern and we took data from the twitter to measure the sentiment of the people around the world.

when the pandemic started in December the disease was only in the china region and it did not spread in other countries. So, people did not concern about it and they shared positive and neutral views and opinions in the social media. The orange line and the red line were close to each other in that time.

When the pandemic was increasing day by day then the daily livelihood situation of the people was getting worse in the March 2020 to June 2020 time interval. They had no income and it is the reason of getting an impact on house rent. House owners were not satisfied and they started to tell the renters to leave the home. We saw many opinions like this in the social media in that pandemic time. So people took it as negative sentiment. In that time there were many negative opinions in the social media than the positive and the neutral. So, the negative line increased in the middle of the time interval of the graph.

In the time of July and August, the days were passing and the situation of the pandemic was not developing. The views and opinions were also increased in that time in the social media. But the perspectives of people were not changed as the situation was not improved. So they posted the negative opinions and the negative line was higher in that time.

In September and October, The situation was getting under control. Lock downs was stopped in some regions because of the decreasing rate of affected patients. People went into their regular life and earn their livelihood. So, The positive opinions were increased in that time and people's sentiments were positive. We can see in the graph that the orange line which represents the positive tweets was raised higher than the other two lines.

- **Poverty:** Covid19 situation has a big impact on poverty. Many of the new poor will live in nations where poverty levels are already high. Significant numbers of people will fall below the extreme poverty line in a number of middle-income countries. Extreme poverty, which has traditionally afflicted people in rural regions, is likely to affect an increasing proportion of city dwellers. It also has an impact social media and there are many opinions of the people.

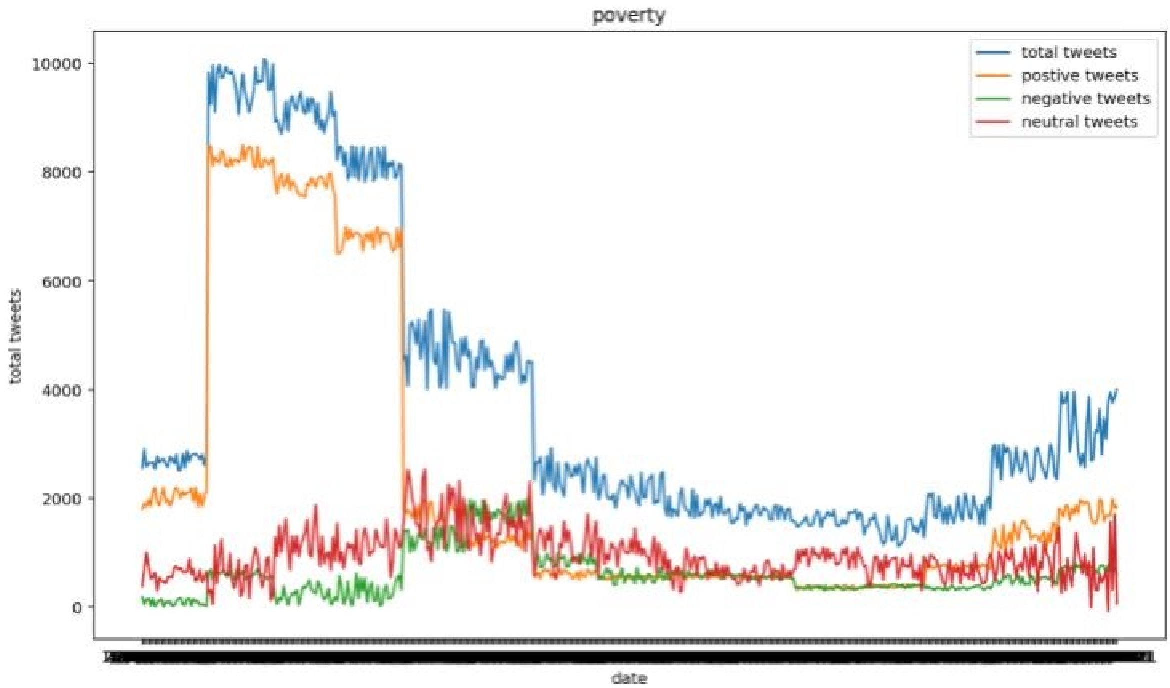

Figure 5.7: Graph of tweets over time due to poverty

Figure 5.7 represents the house rent over time intervals. The X-axis represents the date and the Y-axis represents the total number of tweets. The blue color line represents the tweets, the orange color line represents the positive tweets of the total tweets, the green color line represents the negative tweets and the red color line represents the neutral tweets. The lines of the positive, negative and neutral tweets have been changed over time.

In the beginning of December 2019 to February 2020 there is highly positive line but after some time intervals the line changed in April 2020 to May 2020. The negative line of the total tweets has been raised. So, we can assume that the sentiment or the opinion of the people has been changed due to the condition of poverty and Covid19 has affected many people daily life and income.

In the month of June 2020, the negative number tweets and the neutral number of tweets were increased. So, the green and the red line raised higher than the orange line which is positive.

In July 2020, the number of positive and negative tweets were close to each other. These two numbers varied in a very little gap. So, the orange and red lines are close to each other in the graph but the red line which represents the neutral tweets is higher than the other two lines.

In august, September and October 2020, The positive and negative lines are also close to each other. So, we can say that people were posting their positive and negative opinions in the social media.

In January and February 2021, total tweets were increased than the previous months.

But the orange line which represents the positive tweets has raised higher than the other two lines and we can say that the sentiment of the people was positive in that time than the negative sentiment.

But it a matter of concern that this situation is getting worse now a days. If the pandemic situation remains same then the poverty around world specially in lower income countries and middle countries will face difficulties in the future time.

5.1 Discussion:

We have analyzed the twitter sentiment on the basis of positive, negative and neutral tweets from the social media. We have collected data from twitter for our sentiment analysis for 1 year time interval. As the covid19 situation is not getting better, it will have a future impact on the sentiment of the people. There will be more different kinds of data on social media. We have analyzed our collected data from twitter social and got a result. We had many barriers to do our research because of the lack of premium accounts. We collected data from general twitter account so we had to collect the data on a daily basis. On an another note, the actual number of tweets at the considered date can be larger than the shown amount as there are scopes to miss some hashtags and variation of the intended keywords. We tried to retrieve data from twitter as much as we can but we might be unable to retrieve some more important data that can help us to get better results. We are not claiming that our dataset is big enough to do the study but we are sure that we got a better result from our collected data from twitter. We have also collected data from other social media and we are in the process of data preparation. It is not ready to be shared but it is our future work to do.

Chapter 6

Conclusion

In conclusion, we tried to find out the sentiment of general people on coronavirus which is possibly originated from Wuhan, China with the support of certain methods. This deadly virus has already claimed the lives of many people and affected over 200 countries both economically and socially. Lockdown is imposed to counter this deadly virus but it has stained on lives of people both physically, mentally, and financially. People also urged to relax lockdown to some extend for society to function properly in social media. That's why Social media like Facebook and Twitter are the best platform to evaluate psychological effects as people tend to share their pieces of knowledge and views regarding this virus.

This Covid19 has created a global crisis not only physically but also mentally as people cannot express their freedom properly and move with restriction. Then, their financial crisis has also led to mental breakdowns. People have depended on social media and obtain news regarding the coronavirus of other countries and implemented quarantine steps. Moreover, public awareness has been at the peak of social media. Although people are experiencing negative emotions however they support the lockdown to curve down the rising graph of covid positive patients. Platforms like Twitter divided the covid related topic into four themes like the virus's origin and its source. Then, it impacts people of countries and countries economy and then ways of reducing the risk of infection. Twitter also found out that keyword like Wuhan has been repeatedly used in covid related tweets in the western nation which show indication of risk increase in racism violent and fear among people with other antisocial activities.

To research on social media sentiment analysis firstly, we collected data from social media like Facebook and Twitter. Then, we use the latest technologies to utilize data. Then, we analyzed data to find out the most discussed coved social media with keyword like Covid. Finally, we rate out the positivity and negativity in those discussed topics. As the occurrence linked to COVID-19 will continue to spread like wildfire more severe possibly more strict lockdown will be imposed and more people will be mentally strained. Our goal is to recognize their physical and mental meltdown so with proper steps can be taken by the medical and social workers to ensured aid with help of analyzed data to the people of Bangladesh.

Infection with the Coronavirus is on the rise. We can estimate how many people in the community are affected. As a result, everyone has their own viewpoints and thoughts on the subject. We've decided to use these comments as our dataset for analyzing people's awareness and the most often talked issues. Twitter is a well-known social media platform where individuals from all over the world exchange their thoughts and opinions. Tweets are a kind of communication that expresses one's thoughts and opinions. Tweepy is a Python library that allows us to access Twitter's API. Tweepy allows us to gather data from any user's Twitter account. Using the Twitter API, we gathered data from Twitter. We requested consumer key, consumer secret key, access token, and access token secret key from the Twitter developer portal. The keys were generated for us from the Twitter developer portal after we received clearance.

We have also thought of using other social media platform for future research purpose as we were unable to utilize it in this research since this method can be applied in other social media platform as well. Moreover, recent a new variant of Covid is on the rise and also a new type of disease named "Black Fungus" has also been introduced. So we are also thinking to use this system model for identifying the sentiment of people for these diseases also. So we are assuming that further research can be done on these topics to get more details idea regarding it.

Bibliography

[1] S. Ghosal, R. Bhattacharyya, and M. Majumder, "Impact of complete lockdown on total infection and death rates: A hierarchical cluster analysis," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 14, no. 4, pp. 707–711, 2020.

[2] N. Fernandes, "Economic effects of coronavirus outbreak (covid-19) on the world economy," *Available at SSRN 3557504*, 2020.

[3] M. De Choudhury and S. Counts, "The nature of emotional expression in social media: Measurement, inference and utility," *Human Computer Interaction Consortium (HCIC)*, 2012.

[4] *COVID Live Update: 171,724,414 Cases and 3,570,765 Deaths from the Coronavirus - Worldometer*, en. [Online]. Available: https://www.worldometers.info/coronavirus/ (visited on 06/01/2021).

[5] H. Rosenberg, S. Syed, and S. Rezaie, "The twitter pandemic: The critical role of twitter in the dissemination of medical information and misinformation during the covid-19 pandemic," *Canadian journal of emergency medicine*, vol. 22, no. 4, pp. 418–421, 2020.

[6] D. Scanfeld, V. Scanfeld, and E. L. Larson, "Dissemination of health information through social networks: Twitter and antibiotics," *American journal of infection control*, vol. 38, no. 3, pp. 182–188, 2010.

[7] M. Guy, P. Earle, C. Ostrum, K. Gruchalla, and S. Horvath, "Integration and dissemination of citizen reported and seismically derived earthquake information via social network technologies," in *International symposium on intelligent data analysis*, Springer, 2010, pp. 42–53.

[8] M. Cheong and V. C. Lee, "A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via twitter," *Information Systems Frontiers*, vol. 13, no. 1, pp. 45–59, 2011.

[9] S. R. Rufai and C. Bunce, "World leaders' usage of twitter in response to the covid-19 pandemic: A content analysis," *Journal of Public Health*, vol. 42, no. 3, pp. 510–516, 2020.

[10] A. Abd-Alrazaq, D. Alhuwail, M. Househ, M. Hamdi, and Z. Shah, "Top concerns of tweeters during the covid-19 pandemic: Infoveillance study," *Journal of medical Internet research*, vol. 22, no. 4, e19016, 2020.

[11] W. Ahmed, J. Vidal-Alaball, J. Downing, and F. L. Seguı, "Covid-19 and the 5g conspiracy theory: Social network analysis of twitter data," *Journal of medical internet research*, vol. 22, no. 5, e19458, 2020.

[12] A. Alvarez-Risco, C. R. Mejia, J. Delgado-Zegarra, S. Del-Aguila-Arcentales, A. A. Arce-Esquivel, M. J. Valladares-Garrido, M. R. Del Portal, L. F. Villegas, W. H. Curioso, M. C. Sekar, *et al.*, "The peru approach against the covid-19 infodemic: Insights and strategies," *The American journal of tropical medicine and hygiene*, vol. 103, no. 2, pp. 583–586, 2020.

[13] H. Budhwani and R. Sun, "Creating covid-19 stigma by referencing the novel coronavirus as the "chinese virus" on twitter: Quantitative analysis of social media data," *Journal of Medical Internet Research*, vol. 22, no. 5, e19301, 2020.

[14] E. Chen, K. Lerman, and E. Ferrara, "Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set," *JMIR Public Health and Surveillance*, vol. 6, no. 2, e19273, 2020.

[15] T. Mackey, V. Purushothaman, J. Li, N. Shah, M. Nali, C. Bardier, B. Liang, M. Cai, and R. Cuomo, "Machine learning to detect self-reporting of symptoms, testing access, and recovery associated with covid-19 on twitter: Retrospective big data infoveillance study," *JMIR public health and surveillance*, vol. 6, no. 2, e19509, 2020.

[16] G. Barkur and G. B. K. Vibha, "Sentiment analysis of nationwide lockdown due to covid 19 outbreak: Evidence from india," *Asian journal of psychiatry*, vol. 51, p. 102 089, 2020.

[17] A. D. Dubey, "Twitter sentiment analysis during covid19 outbreak," *Available at SSRN 3572023*, 2020.

[18] N. K. Rajput, B. A. Grover, and V. K. Rathi, "Word frequency and sentiment analysis of twitter messages during coronavirus pandemic," *arXiv preprint arXiv:2004.03925*, 2020.

[19] A. Abd-Alrazaq, D. Alhuwail, M. Househ, M. Hamdi, and Z. Shah, "Top concerns of tweeters during the covid-19 pandemic: Infoveillance study," *Journal of medical Internet research*, vol. 22, no. 4, e19016, 2020.

[20] M. Cinelli, W. Quattrociocchi, A. Galeazzi, C. M. Valensise, E. Brugnoli, A. L. Schmidt, P. Zola, F. Zollo, and A. Scala, "The covid-19 social media infodemic," *Scientific Reports*, vol. 10, no. 1, pp. 1–10, 2020.

[21] J. Samuel, G. Ali, M. Rahman, E. Esawi, Y. Samuel, *et al.*, "Covid-19 public sentiment insights and machine learning for tweets classification," *Information*, vol. 11, no. 6, p. 314, 2020.

[22] *COVID-19 pandemic*, en, Page Version ID: 1026362330, Jun. 2021. [Online]. Available: https://en.wikipedia.org/w/index.php?title=COVID-19_pandemic&oldid=1026362330 (visited on 06/01/2021).

[23] *WHO Coronavirus (COVID-19) Dashboard — WHO Coronavirus (COVID-19) Dashboard With Vaccination Data*. [Online]. Available: https://covid19.who.int/ (visited on 06/01/2021).

[24] *COVID-19 pandemic in Bangladesh*, en, Page Version ID: 1026229460, Jun. 2021. [Online]. Available: https://en.wikipedia.org/w/index.php?title=COVID-19_pandemic_in_Bangladesh&oldid=1026229460 (visited on 06/01/2021).

[25] *COVID Live Update: 171,827,043 Cases and 3,573,701 Deaths from the Coronavirus - Worldometer*, en. [Online]. Available: https://www.worldometers.info/coronavirus/ (visited on 06/01/2021).

[26] C. E. Lopez, M. Vasu, and C. Gallemore, "Understanding the perception of covid-19 policies by mining a multilanguage twitter dataset," *arXiv preprint arXiv:2003.10359*, 2020.

[27] L. Singh, S. Bansal, L. Bode, C. Budak, G. Chi, K. Kawintiranon, C. Padden, R. Vanarsdall, E. Vraga, and Y. Wang, "A first look at covid-19 information and misinformation sharing on twitter," *arXiv preprint arXiv:2003.13907*, 2020.

[28] N. Karisani and P. Karisani, "Mining coronavirus (covid-19) posts in social media," *arXiv preprint arXiv:2004.06778*, 2020.

[29] R. J. Medford, S. N. Saleh, A. Sumarsono, T. M. Perl, and C. U. Lehmann, "An "infodemic": Leveraging high-volume twitter data to understand early public sentiment for the coronavirus disease 2019 outbreak," in *Open forum infectious diseases*, Oxford University Press US, vol. 7, 2020, ofaa258.

[30] I. Arpaci, S. Alshehabi, M. Al-Emran, M. Khasawneh, I. Mahariq, T. Abdeljawad, and A. E. Hassanien, "Analysis of twitter data using evolutionary clustering during the covid-19 pandemic," *Computers, Materials & Continua*, vol. 65, no. 1, pp. 193–204, 2020.

[31] K. Jahanbin, V. Rahmanian, *et al.*, "Using twitter and web news mining to predict covid-19 outbreak," *Asian Pacific Journal of Tropical Medicine*, vol. 13, no. 8, p. 378, 2020.

[32] M. Shammi, M. Bodrud-Doza, A. R. M. T. Islam, and M. M. Rahman, "Strategic assessment of covid-19 pandemic in bangladesh: Comparative lockdown scenario analysis, public perception, and management for sustainability," *Environment, Development and Sustainability*, vol. 23, no. 4, pp. 6148–6191, 2021.

[33] C. Kaur and A. Sharma, "Twitter sentiment analysis on coronavirus using textblob," EasyChair, Tech. Rep., 2020.

[34] A. Sarlan, C. Nadam, and S. Basri, "Twitter sentiment analysis," in *Proceedings of the 6th International conference on Information Technology and Multimedia*, IEEE, 2014, pp. 212–216.