 doi:10.20944/preprints202205.0248.v1

Article

COVID-19 Real Time Live Tweet Sentimental Analysis Using Deep Learning Methods, Its Effect on Phases of COVID-19 Waves

Shruti Thakur [1] and Ubio obu [1*]

[1] Computer Science and Engineering, G H Raisoni College of Engineering, Nagpur, India
*Corresponding author: obu_ubio.cs@ghrce.raisoni.net

Abstract: The Covid-19 also known as the Coronavirus is a viral disease from the SARS-CoV-2 family of virus, as at December 2019 the first case of this virus infection was identified at Wuhan, China, this seemingly isolated case soon became a global pandemic, whose effect was felt globally which also had colossal effects on both health, economic and politics . As at the time of this research about 4.5 million people have died of the Coronavirus and over 215 million people already infected by it. This pandemic stood out not just for its scale but for how social media was a major contribution to its spread as well as to curbing it. The power of social media was used to spread misinformation as well as to spread awareness on the subject, with both having massive impact on the people. In this paper we will be running a sentimental analysis on twitter under the keyword “Covid-19 and Coronavirus”, twitter is a powerful social media tool that is known for its ability to keep trends in the form of tweets, we will be drawing correlations between the peaks of tweet with the peak of infection. We will also be analyzing to know what the impact of these tweets are having on the rate of the infection and vice versa. We will also be analyzing what people are tweeting most about, what are the talking points, comparing both real time and past tweets with real time infection and death rates using deep different learning methods to access what information can be derived from it.

Keywords: Covid-19; sentimental analysis; deep learning

1. Introduction

Covid-19 will forever be a defining period of time, in the epoch of human history. In referring to the technological, historical and even cultural events of the twenty first century, the pre covid-19 era, the covid-19 pandemic era and the post covid-19 era will be suitable adjectives to refer to these period, as each of these eras have had defining moments that will forever shape the effects of this periods. Man has had to imbibe new cultures, technologies and way of life as a result of the Covid-19 pandemic, many of which may become norms in the future.

As at the time of this research, it is unfortunate that we do not know where we stand in the events of history, can we sufficiently say we are at the edge of the end of the pandemic? Hopefully so, but data has shown that the embers of Covid-19 all over the world weans slower than expected, and may take even up to a year for it to burn out, but one thing is sure, we are not yet in the Post Covid-19 era, neither are we at the beginning of the pandemic where the pandemic was strange, and man was oblivious of how to live with this Virus. This research is coming at a point when man has gotten some knowledge about the virus, has developed some vaccines like the Modern vaccine, and Pfizer-BioNtech vaccine[1].

This means that a lot of information has been gotten about the virus so far, unfortunately with this mass of information is a mass of misinformation about the virus, vaccine, prevention etc. as well. The most potent power of Coid-19 has been not just its ability to mutate and build stronger variance like the Delta Variant, but in its ability to mutate through social media to spread misinformation and half-truths. The Covid-19 pandemic has been one pandemic which has shown globally that viruses cannot just be fought with vaccines and drugs only, but there is a need for the social media to be incorporated while thinking of vaccines and prevention mechanisms to be employed against these Viruses[2].

Social media is of course broad, capturing the total sentiments on all social media may be a bit hard, however we used twitter, a micro blogging social media, for our sample data. Twitter is associated with trends, and have a way of depicting the average sentiments that is observed on other social media platforms. This makes it a good fit for doing a covid-19 sentimental analysis, because the result obtained here is a suitable representation of the result elsewhere.

In this paper we engaged into a lot of questions, these questions defined the different sentimental analysis we did, it also informed the different methodologies that we employed in our research paper. These questions are pivotal to this research, and are the questions that form the heart of this topic, viz;

Q1: (i) Which Keywords were most common as related to Covid-19?
(ii) Which new keywords are making the trends with respect to Covid-19

Q2: (i) What were the effects of tweets on Covid-19 cases?
(ii) Is there any observable correlation between tweets and number of cases, vaccination or number of death cases?
(iii) What does the real time tweets tell us of the real time cases?

These are some basic questions that this paper seeks to answer. For this research we needed two basic data, first we needed a repository of previous tweets collected over the period of 23 march 2020 to July 15 2020 and then compared this with a live tweet repository we built for collection and display of tweets on an Heroku app and using Text Blob for sentimental analysis of the related Covid-19 tweets collected live within seven days (September 5 through September 12).

2. Literature Review

1. Undoubtedly the increase in Covid-19 cases, the constant deep and peaks that defiled scientific predictions, the cyclic waves and recurrent variations has of course led many scientist and researchers to seek possible explanations for these events, many opinions have been hypothesized by many researchers to explain this phenomena, but among them are some who took the discussion beyond the micro level but into the macro interaction of how the social media played a vital part in the waves of cases with special reference to twitter[3].
2. Chakraborty K et al 2020 discuses on the power of twitter as a tool, he asserts that twitter has gone beyond just being a social media but a social tool controlling the opinion of many, everyday according to him more than 500 million tweets are posted on twitter and about 200 billion each year, the gross implication of this is that societal and global opinions and conversations are ongoing in this platform and over the years have become the hub for both information and misinformation. Chakraborty did a sentiment analysis of most of Covid-19 tweets on twitter and concluded that most of the tweets were in positive sentiments, but unfortunately the positive words had little clues to offer[4].
3. Shadi S. et al were a little bit more brute in presenting their findings and assertion of social media influence and role in the pandemic. They directly linked how the spread of conspiracy theories, fake news and rumors directly contributes to the spread of the pandemic. In their method they developed an automatic machine learning model that automatically detects framework of tweets containing fake news and misinformation. They also explored how politic and political inclinations and ideologies is affecting public health[5].
4. Sural et al discussed on the direct correlation between Trait Emotional Intelligence (TEI) and Problematic Social Media Usage (PSMU) and how this indirectly influences people into taking sides with popular opinion. They carried out a sample study on

 doi:10.20944/preprints202205.0248.v1

about 444 people who had active social media platforms as at the time of the study, one of the key research outcomes of their study showed that TEI was more positive for younger social media users than with the older ones, insinuating that older generations are much more prone to misinformation and fake news from online platforms than the younger ones[6].

5. Talwar et al in their study focused on the dark effects of social media with a direct link between social media and fake news. They went ahead to propose a model which can help scientist and researchers in identifying fake news, ten such hypotheses were drawn from rational psychological theories. This study becomes very useful as it will help researchers with a scientific indication to identify fake news and social media misinformation[7].

6. Nalini Chintalapudi et al, concentrated on using BERT a deep learning analytical tool to further the ongoing discussion on the general sentiments of Covid-19 tweets, they discovered a more positive sentiment with respect to the tweets, and implied a resonating encouragement towards the work of health workers. They went on to validate the works of prior authors on same domain but with a more precision, using the BERT deep learning model[8].

3. Methodology

In the course of this study, our first objective was to take live tweets from twitter API on the subject of Covid-19, we took a random sample of tweets consistently for 7 days, starting from September 5 through September 11. These data will later on give us insight into the different issues and will be relevant in our analysis. We also got cleaned tweets of Covid-19 from the period of intense lockdown last year, when the cases were on a spike, to compare with our live tweets, for comparison and data analytical purpose. Below is an image of the daily Covid-19 cases, deaths etc. Collected in the duration of this study.

Table 1. Daily report of observed covid-19 live tweets from September 5th through 11th 2021.

DAY	POSITIVE TWEETS	NEGATIVE TWEETS	NEUTRAL
September 5	69	27	105
September 6	49	38	114
September 7	47	41	113
September 8	88	30	83
September 9	71	34	96
September 10	102	25	74
September 11	105	21	75

Table 2. Daily report of covid-19 cases from September 5th through 12th 2021.

Date	Total Deaths	New Cases	Change in Total
September 5	4,591,358	538,820	7,608
September 6	4,598,874	504,475	7,516
September 7	4,607,974	540,496	9,072
September 8	4,618,196	643,015	10,250
September 9	4,628,414	626,387	10,218
September 10	4,646,941	609,773	9,602
September 11	4,654,073	550,176	8,952
September 12	4,654,073	486,906	7,132

First it was essential that we collected live tweets from twitters API. To take the live tweets from the twitter API we had to first apply for a twitter developer account. The twitter developer account gives users access to some of twitters functionality which isn't accessible to regular users. One of such is the ability to read data from twitters API.

At the developer dashboard we are given unique consumer_key, consumer_secret, access_token and access_token_secret, these codes we saved and used in connecting our build to the twitter API so as to be able to query it. The queried data was stored in a DataFrame (df) to help with easy retrieval of data. Of course data collected directly from twitter API into df can never be clean, as it contains symbols and tags that are irrelevant to our data. Remove such irrelevant tags it was essential that we preprocess and cleaned the data, to clean the extracted data from tags, links and symbols like (@,#,) etc.

3.1. Live Deployment on StreamIt

After the tweets had been cleaned we went ahead to use TextBlob for our sentimental analysis. TextBlob was suitable for this project due to its huge libraries, its classification of sentences and simple polarity. The polarity of TextBlob ranges from -1 to 1, where:

-1= Negative sentiments

0= Neutral sentiments

1= positive sentiments

Immediately the sentiments have been extracted from the tweet we were able to analyze it and visualize it as well, for us we paid attention to analyzing the tweet sentiments, the word cloud as well as the count plot of the different sentiments. For convenience and ease, since we needed to collect the data for 7 days we created an API using StreamIt and Flask and deployed the project on Heroku, so that we could simply use the key word "Covid-19" to query the API and get the Relevant analysis.

The images of the results gotten from the live data are seen as follows.

Live Visualization of COVID-19 Cases

Our objective as stated earlier in our objective is to place side by side, results and analysis of Live Covid-19 tweets together with live Covid-19 cases in the same period, so as to analyze and observe the necessary correlations, we confined our data to take an average of about 200 different tweets from different locations so as to sample the average sentiments of the overall topic, considering how large a data set on covid-19 can be and how lengthy it will be to analyze it, we reduced and set our parameters to an average of 200 tweets . It was therefore important for us that in this period we also were collection live covid-19 cases globally. Our approach was simple we simply collected the live updates of this cases from September 5th through September 11th, as reported by the World health organization. We recorded the data on a simple CSV file, the live data of the daily infection cases, the death cases and total cases for each day was recorded on each day and the corresponding results plotted into simple visualization. The report and visualizations can be seen in fig 2. In our result and more of it in our discussion section of this report.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 May 2022 doi:10.20944/preprints202205.0248.v1

Below is the Extracted Data :

	Tweet	Likes	RT	User_location	clean_tweet	Sentiment
0	to …	0	0	Pakistan	you have the power to …	Positive
1	:ina…	0	0	India	india s august vaccina…	Positive
2	Adri…	0	0	West of you	1000 horgan dix and he…	Positive
3	_rad…	0	0	United Kingdom	radio the consequences…	Positive
4	.IES…	0	0		lies lies lies and mor…	Positive
5	and…	0	0	Taguig City, Philippin…	get vaccinated now and…	Positive
6	wil…	0	0	Sønderborg Denmark	from 9 10 covid 19 wil…	Neutral
7	give…	0	0		kelly khumalo has give…	Positive
8	IMM…	0	0	Taguig City, Philippin…	taguige os read how ac…	Neutral
9	a MP…	0	0	Accra, Ghana	dafiama bussie issa mp…	Positive
10	ippy…	0	0	India	the hungry happy hippy…	Positive

Figure 1. Sample of covid-19 tweets taken from different accounts and different countries.

Table 3. Covid-19 tweets taken from different accounts and different countries.

	Date	User	Tweet	Likes	Rt	User Location	Clean Tweet	Sentiment
1	2021-09-05 23:22:58	Hoff	Delta variety never stoo	0	1	USA	delta variety never stoo	Neutral
2	2021-09-05 23:22:57	Grace Asaf	Presenting, reporting,	0	0	India	presenting, reporting,	Neutral
3	2021-09-05 23:22:56	Hammad Coughlin	"I'm SHOCKED, SHOCKED th	0	0	Pakistan	"I'm shocked, shoched th	Negative
4	2021-09-05 23:22:56	Julian Martalog	$cemi new Covid-19 vari	1	0	Miami, FL	cemi new covid 19 varia	Positive
5	2021-09-05 23:22:56	rodriQuez	Dying COVID-19 Patient R	0	0	San Antonio, TX	dying covid 19 patient	Neutral
6	2021-09-05 23:22:56	Gary Duffy	Covid-19: 172 new cases	0	0	New Zealand	covid 19 172 new cases	Positive
7	2021-09-05 23:22:56	J.M.M.	Justin Trudeau - Virus V	0	0	Instagram: pastexpirycom	justin trudeau virus var	Neutral
8	2021-09-05 23:22:56	Rick Pittman Rick Pittman Rick Pittman Rick Pittman	UK secures 114 million	0	1	UK	uk secures 114 million	Postive
9	2021-09-05 23:22:56	Galvo	Please be careful out t	0	0	London	please be careful out t	Negative

3.2. Past Sentiment of COVID-19 Cases

To compare the performance of our live twitter sentimental analysis, it was important that we compare this with the performance of past covid-19 tweets and compare and contrast the general sentiments between the live tweets and the past tweets. There are several methods that can be used in scraping past covid-19 tweets, we can import an API GetOldTweet3, then proceed with cleaning the data, then analysis. Or simply use preprocessed cleaned data from github.com, while we go ahead with just analyzing the data. Either method is fine, but we chose to use the later method for this work because it is more time efficient.

We decided to concentrate our analysis to just India (As a model for basic generalization), and for the duration of March 2020 to June 2020, which were critical periods in the phase of Covid-19 worldwide. These period is when the world had the first wave of the covid-19 pandemic, as such the analysis will be very reflective of the sentiments of that period.

A github.com repository was found (https://github.com/gabrielpreda/CoViD-19-tweets) which contains cleaned data of tweets from the period of March 2020 to June 2020, it contained about 3090 tweets on subjects related to Covid-19 like “Coronavirus”, “Covid-19”, “lock down”, “pandemic”, etc. Model On

3.3. Analysis on Past Covid-19 Tweets

The collected data as already highlighted above contained cleaned data and was very easy to manipulate as well as analyze it to our preference. We

Simply hard coded the data using simple manual methods, then categorized the data from 0 to 3, according to the major themes observed in the data.

Where 0 represented words mapped in the category of fear, 1 represented words mapped in the category of sadness, 2 represented words mapped in the category of anger and 3 represented words mapped in the category of Joy. Fig 2 below shows a visualization of the data.

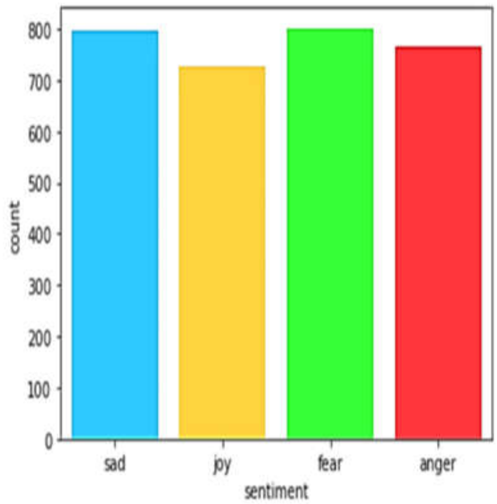

Figure 2. A plot of sentiment against count.

4. Results and Discussion

The following were the daily visualizations and count plot of sentimental analysis between 5th to 11th of September 2021.

 doi:10.20944/preprints202205.0248.v1

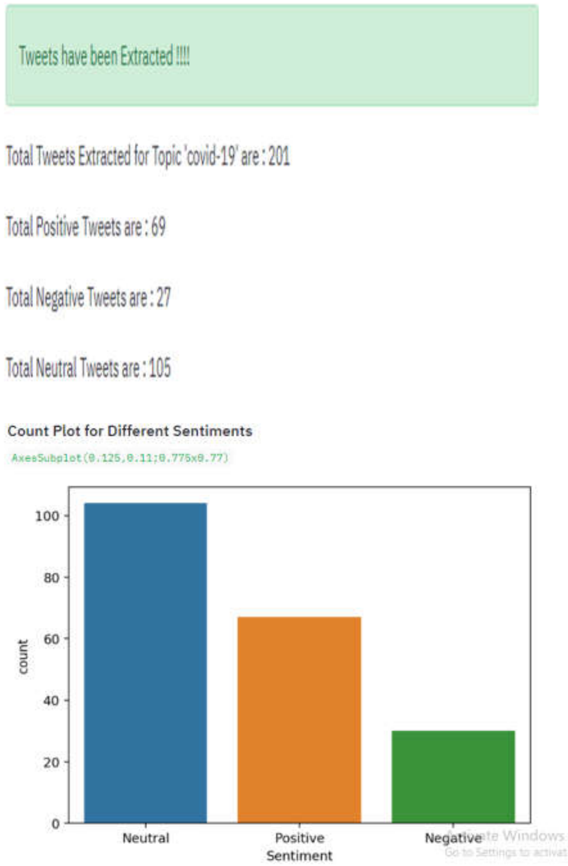

Figure 3. September 5.

 doi:10.20944/preprints202205.0248.v1

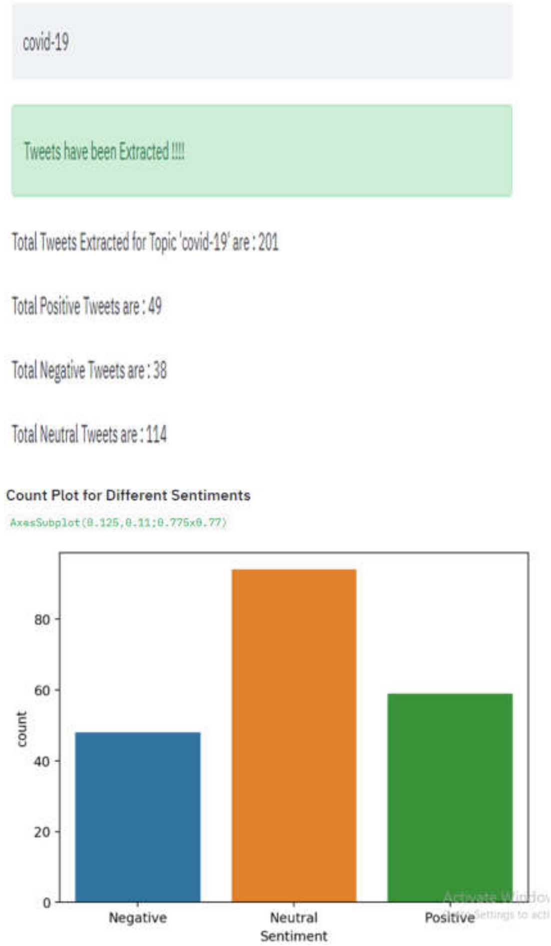

Figure 4. September 6.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 May 2022 doi:10.20944/preprints202205.0248.v1

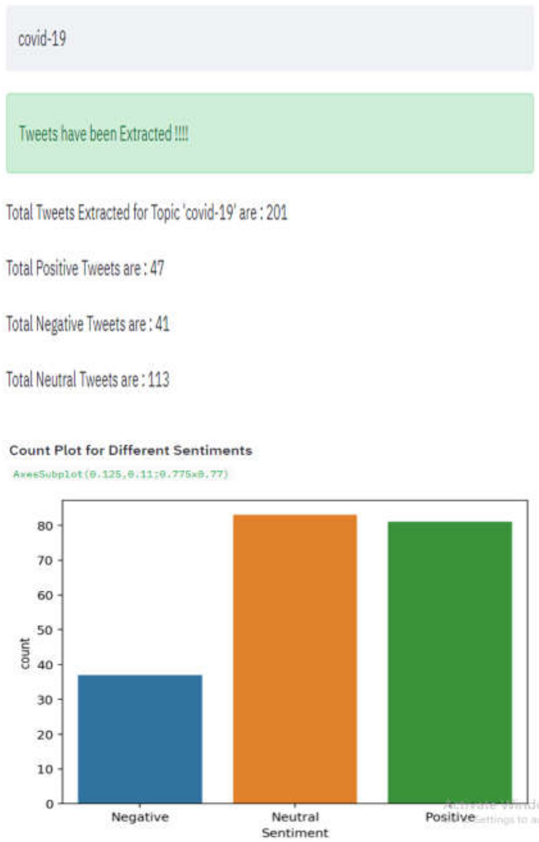

Figure 5. September 7.

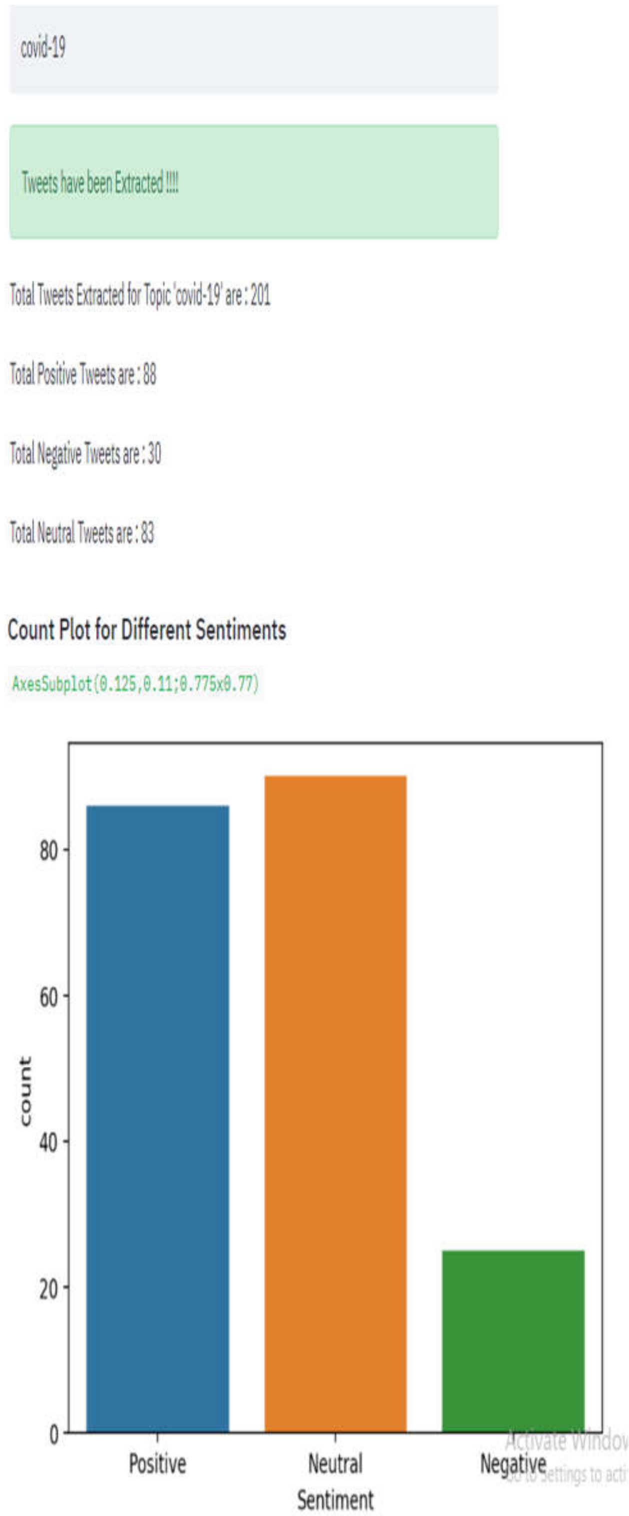

Figure 6. September 8.

 doi:10.20944/preprints202205.0248.v1

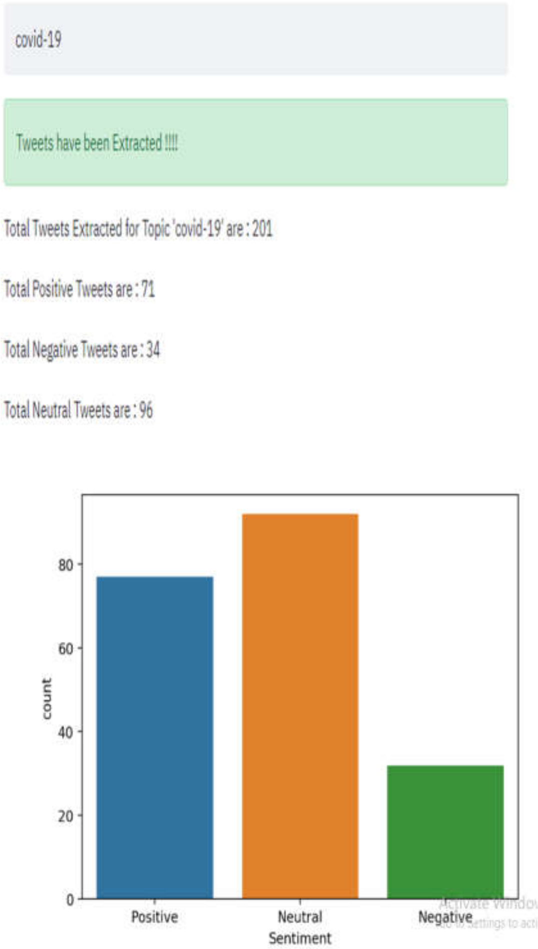

Figure 7. September 9th.

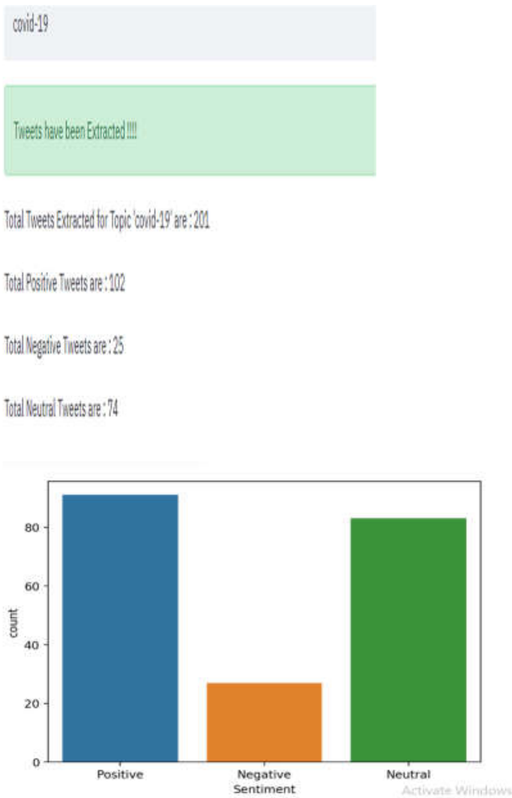

Figure 8. September 10.

 doi:10.20944/preprints202205.0248.v1

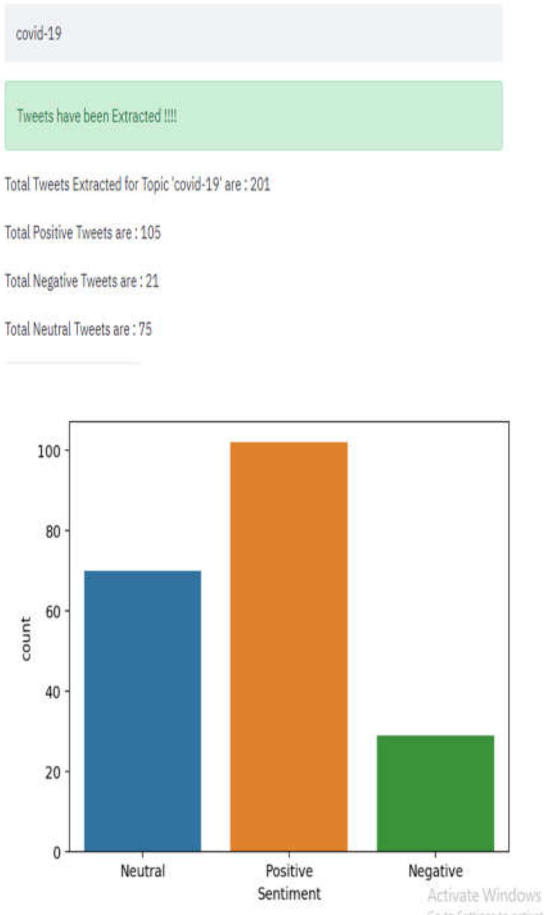

Figure 9. September 11th.

The daily records of Covid-19 cases, new cases and death for 5th to 11th of September 2021, has been plotted and the result can be found below.

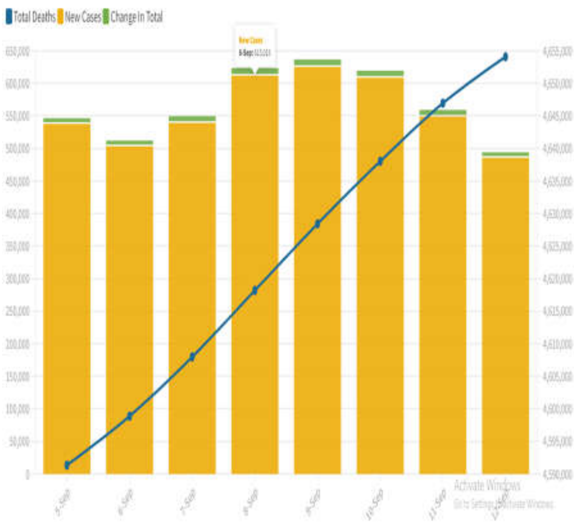

Figure 10. Daily-19 cases from 5th to 11th september 2021.

For the period under consideration the word cloud giving a general sentiment of the most common topics and words endemic among twitter users can be found below.

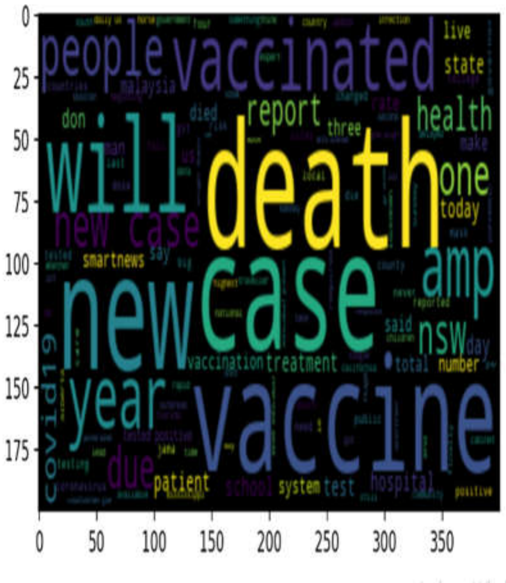

Figure 11. word cloud.

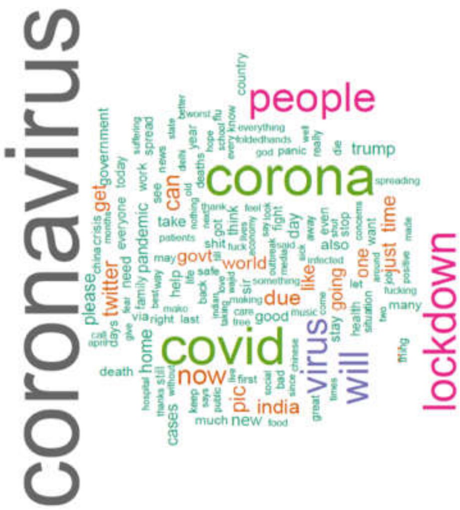

Figure 12. For the period of the lock down a word cloud representing the sentiment of twitter users can be found below.

5. Discussion

The aim of performing sentiment analysis is to help investigators, scientist, researchers and others with vested interest to understand the sentiments and emotions that surround an issue, these sentiment are very important to help with appropriate action. The sentiments behind covid-19 cannot be over emphasized, the political sentiments, and religious as well as cultural sentiments. As the Lockdown ebbs globally this timely sentimental analysis of live covid-19 tweets over a one week period revealed some insight as seen in the data and result obtained.

Before discussing some of the implications from our results, first we discovered through this study how easy it is to do a live sentiment analysis on twitter by simply using twitter API. Twitter has made their API available to all, using the twitter developer platform a user is able to deploy twitter API to their projects, this was our first discovery, the simplicity of the process really helped in our research. The live sentiment analysis taken over a 7 day period was one of our distinct approach, monitoring the daily fluctuations of the Covid-19 cases and other relevant data gave us some deep insight into what is going on not just on an overall basis but on a more specific and real time manner.

First from our 7 day visualization of daily Covid-19 tweets, we discovered that as at September 5th to 11th there is a general neutral sentiment by tweeters about the Covid-19 pandemic, this neutral sentiment is far from the past data of Covid-19 we collected, the most common words using NLP methods for past Covid-19 data was "fear", this was at the peak of the pandemic but as the pandemic ebbs down up to September where this study was made we can see a movement from negative sentiment of fear to neutral sentiments. This assumes two possibilities:

1. That people are getting used to the pandemic that there are no more afraid of the Covid-19 situation
2. That people have become more positive or hopeful than they were at the peak of the pandemic, this hope has been able to move the sentiments away from fear to a neutral standpoint. However, whatever is influencing that move away from fear is not enough it to the point of a total positive attitude. A mix of vaccination drive and continuous rise in cases can be attributed as a reason why it stays at a neutral sentiments and rightly reflects the mood of most people within this period as **correlated by other researchers[9]**.

We observed that despite an overall neutral feeling about the Covid-19 pandemic, this has not translated to a reduction or flattening in the daily covid-19 cases, in fact the daily global cases have continued to rise despite people generally not being afraid of it. Probably this shows a pattern of carelessness and irresponsibility that has **been observed in other studies[10],** since the advent of the vaccines and as people gradually get back to their normal activities. The neutral sentiment has contributed to rising Covid-19 cases. A plot of the change in total death cases on each day shows and a plot of the daily Covid-19 cases show a similar pattern, a rising number of cases and death daily with a sharp fall on the 10th of September 2021 for both the covid-19 cases as well as the death cases. From the 10th of September we also observe looking at our sentiment bar more specifically that there was an Increase in the positive sentiments. Implying that the drop in death cases and new Corona Virus cases generally influence the positivity of the sentiments of tweeters. While an increase in death and new case makes influence the neutrality and negative sentiment of the tweeters.

Finally our typical word cloud for these 7 days was almost the same, we discovered that the word "Death" was the most frequently used word as associated with the subject, this gives us how much death plays in affecting the sentiment of people whenever Covid-19 is mentioned, it reveals why there is a correlation between the death cases and the daily sentiments, and also that people associated Covid-19 more with death, despite a seeming reduction in the fear of the disease.

Our word cloud also identified new cases as another issue that is of most concern to the people tweeting, again this validates our data and inference on the importance new case as well as death plays in the overall sentiments of the people. Compared to word cloud of previous Covid19 tweets during the lock down period, we noticed that there has been a change in terms of the concerns of the people to other issues. Prior now as seen in the word cloud image above, topics like Corona, Covid, Lockdown, people dominated the word cloud but as at this period it is Death, New cases, vaccine, vaccinated, treatment, health etc. are the dominant topics on the word cloud. Comparing this we can deduce that there is a movement from uncertainty and confusion to a point where people are beginning to engage on what to do to come out of the pandemic.

6. Conclusion

In conclusion, our study established that as the world moved away from the lockdown, while the pandemic is totally not over and the number of cases still increasing globally(These are the situation of things as at the time of this study), that generally there has been a shift from negative sentiments towards a more neutral sentiment. This means that while people are not hopeful they are neither as cynical nor afraid as they were in the peak of the lock down and pandemic. We showed the influence of death cases and new cases

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 May 2022 doi:10.20944/preprints202205.0248.v1

on the general sentiment of the people. This happened to be an inverse relationship, the higher the death and new cases, the lower the positive sentiment, the lower the death and new cases the higher the positive sentiments. While previous studies dwelt on overall sentiments analyzed over a long period of time, our focus was a real time daily approach, this enabled us to pick up more intrinsic details that would be overlooked using broad methods.

From our result we also realized that vaccination is a key topic that has entered into the word cloud as compared to during the lock down when such a word was not there. It is safe to say that vaccination a new word dominant in the word cloud must have contributed towards the shift in sentiments towards a more neutral point. To move the sentiments to a positive point, we recommend that as a society we should look into methods of reducing death and new cases, as well as increasing vaccination, these are key factors that will influence a more positive sentiment.

References

1. Jackson LA, Anderson EJ, Rouphael NG, et al. An mRNA vaccine against SARS-CoV-2 — preliminary report. N Engl J Med 2020;383:1920-1931.
2. Ahmad, A.R.; Murad, H.R. The Impact of Social Media on Panic During the COVID-19 Pandemic in Iraqi Kurdistan: Online Questionnaire Study. J. Med. Internet Res. 2020, 22, e19556. [CrossRef] [PubMed]
3. Twitter, Twitter Usage Statistics—Internet Live Stats.Availableonline:https://www.internetlivestats.com/twitter-statistics/ (accessed on 19 September 2021).
4. Chakraborty, K.; Bhatia, S.; Bhattacharyya, S.; Platos, J.; Bag, R.; Hassanien, A.E. Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media. Appl. Soft Comput. 2020, 97, 106754.
5. Shahsavari, S.; Holur, P.; Tangherlini, T.R.; Roychowdhury, V. Conspiracy in the time of corona: Automatic detection of COVID-19 conspiracy theories in social media and the news. J. Comput. Soc. Sci. 2020, 3, 279–317.
6. Süral, I., Griffiths, M.D., Kircaburun, K. *et al.* Trait Emotional Intelligence and Problematic Social Media Use Among Adults: The Mediating Role of Social Media Use Motives. *Int J Ment Health Addiction* **17,** 336–345 (2019). https://doi.org/10.1007/s11469-018-0022-6
7. Talwar S, Dhir A, Singh D, Virk GS, Salo J. Sharing of fake news on social media: application of the honeycomb framework and the third-person effect hypothesis. *J Retail Consum Serv.* 2020;57:102197. doi: 10.1016/j.jretconser.2020.102197
8. Chintalapudi, N.; Battineni,G.; Amenta, F. Sentimental Analysisof COVID-19 Tweets Using Deep Learning Models. Infect. Dis. Rep. 2021,13,329–339.https://doi.org/10.3390/idr13020032
9. T. Vijay, A. Chawla, B. Dhanka and P. Karmakar, "Sentiment Analysis on COVID-19 Twitter Data," 2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE), 2020,pp.1-7,doi:10.1109/ICRAIE51050.2020.935831.
10. https://ourworldindata.org/grapher/daily-covid-cases-deaths?country=~OWID_WRL