COVID-19 TWITTER SENTIMENTS ACROSS THE UNITED STATES IN AUGUST 2020

Umesh R. Hodeghatta[1] , Ph.D and Sanath V. Haritsa[2]

[1]Northeastern University, Boston, MA, USA;
[2]NU-Sigma U^2 Analytics Lab, India

ABSTRACT

COVID-19 has drastically affected the entire nation. This study involved collecting tweets and analyzing the COVID tweets for August 2020. The aim was to understand whether people have expressed sentiments related to COVID-19 across all the states of the United States and find any correlation between the sentiment tweets and the number of actual cases reported. Around 400000 COVID-19 Twitter data was collected for August 2020 from the primary Twitter database. A simple NLP-based unigram sentiment analyser, a novel approach different from the traditional machine learning approach, was adopted to identify twitter sentiments. The results indicate that tweets related to COVID demonstrate the two types of sentiments, one related to the deaths and the other about the COVID symptoms.

Furthermore, the results show that the sentiments for each category vary from State to State. For example, states of New York, California, Texas are higher tweets sentiments regarding expressing death sentiment, and states of New York, California, Nevada, are higher regarding sentiments of expressing COVID-19 symptoms with an accuracy of 83%. As a part of the research, a new sentiment scorecard was created to provide a sentiment score based on the sentiments of the tweets expressed to the actual reported death cases. The sentiment scores for the 'symptoms' class are higher for Maryland, New Jersey, and Oregon, whereas sentiment scores for the 'death' class are higher for Virginia, Delaware, and Hawaii. These sentiment scores indicate that the Twitter users of these states are actively tweeting about symptoms and deaths even though the actual reported cases are less in these states. The analysis results also found no or little correlation between the COVID Tweets and the number of COVID death cases reported across all the states.

KEYWORDS

COVID 19, Twitter Behaviour, Twitter Analytics, Sentiment Analysis, Big Data Analytics, Business Analytics.

1. INTRODUCTION

Social media has changed the way people communicate and propagate information. Social media promotes individuals to connect with their friends and families and share pictures, videos, and their thoughts about the day-to-day activities, including experiences of using a product, services, latest news, or any other important information within their social network [5]. Over the last decade, social media also played an essential role in creating awareness and knowledge about public events, science & technology, world politics, and even public health.

On December 31, 2019, an outbreak of a new virus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), coronavirus disease 2019 (COVID-2019), was announced by the Wuhan Municipal Health Commission. As of October 25, 2020, the virus resulted in total

confirmed cases of 42.6 Million and more than 1.15 million deaths. According to the available statistics, 80% of COVID-19 infections are mild, 15% are severe, and 5% are critical (Coronavirus Resource Center, 2020). Within weeks of the announcement of COVID-19, information started spreading on social media platforms. People started talking about the diseases, precautions to be taken care of, number of cases, number of deaths, symptoms, isolations, quarantine, travel information, and other relevant data about the diseases, which created awareness among people also created panic. Further, COVID-19 itself caused several fake claims about the disease, as well as spreading racism instead of making use of social media for the good cause [6, 2]. Therefore, social media played an important role during the health crisis communications and changed the tourists risks perception [11].

Researchers have considered Twitter as an informational media and a popular microblogging platform among many social network platforms because of its ease and instantaneous reach to millions. Messages can be written to the point, crisp, and within a limited number of words and convey your points across to millions almost immediately [5]. Twitter is a micro-blogging site founded in 2006. It allows people to post their thoughts in a text from using just 140 characters. These posts are popularly known as tweets, which may include texts, and URLs. Today, Twitter has more than 600 million users [8]. Every three days, a billion tweets are being sent which reaches millions of people [8]. People use Twitter to track news, find out what others are talking about, the latest in politics, technology, events around your city, the latest phone or gadget, jobs in your area, and find out what people are talking about the latest movies. Twitter allows sharing the latest information, news, and ideas and solicits suggestions or ideas instantaneously across the globe unlike ever before. Some users may be active listeners, and some may be actively participating and exchanging information. Tweets may also contain people's opinions, views, or experiences and these tweets are available to the public [7]. Though the authenticity and meaning of the message may sometimes create confusion and conflicts due to the messages' unfriendly and short characters, it is still a popular medium.

Sentiment analysis is the process of detecting sentiments expressed in a given text [12]. The sentiments can be found in customer feedback, reviews, or critiques in different forums. The sentiment analysis aims to determine an author's attitude concerning a given topic. Under conventional circumstances, it is challenging to find out why a consumer did not buy a product, but with the help of sentiment analysis tools, it becomes easier to find out the reasons and logic behind a customer not purchasing the product [12, 3]. Apart from product and marketing, sentiment analysis is useful in areas such as politics, sociology, and psychology.

In this paper, we analyze how people used Twitter, as a social media, to convey messages related to COVID during the pandemic time. The tweets are extracted across all states within the United States during August 2020. Analysis was carried out to see whether these messages carry any sentiments related to COVID, such as symptoms, deaths, or just information. Further, a correlation with the actual number of cases was also carried out. The following sections describe the research questions, research methodology, data collection, and data analysis.

2. RESEARCH QUESTIONS

This research aimed to study how people use Twitter to communicate about COVID-19 during the pandemic time. In this study, "COVID-19" was chosen as a search topic of Twitter message, and the study was carried out to answer the following:

1. How frequently people are communicating and tweeting COVID-19 information.
2. Type of behavioral sentiments and messages users are tweeting.
3. Relation between the tweets and number of actual death cases across all the states.

4. Predicting COVID-19 sentiments.

3. METHODOLOGY

Based on the limitations of the twitter.com free open access to the database, nearly 400,000 tweets were collected for all the 52 USA states. Not all the 400,000 tweets had COVID related tweets. The flow structure and the sentiment analyser model is as shown in Figure 1.
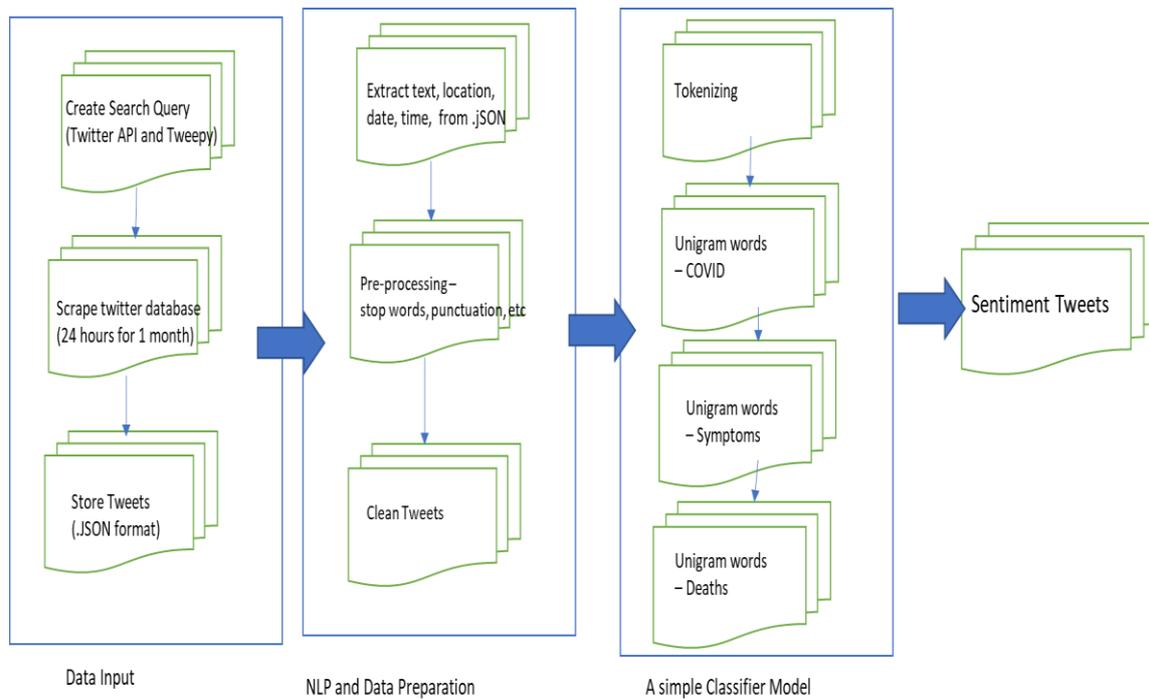

Figure 1. Architecture Framework of COVID Tweets sentiment analyser

3.1. Data Source (Data Input)

In this research, the primary data was directly from the Twitter database (http://www.Twitter.com). The Twitter database was accessed using Twitter search API. We selected the top three most populous cities in each State of the US and collected tweets every day for one month (of August 2020). Table 1 provides the details of states and cities used for search API geo-locations to collect the tweets.

Table 1. Twitter users cities and states used in the study

State,Federal District,or	Most populous	2nd most populous	3rd most populous	State,Federal District	Most populous	2nd most populous
Alabama	Birmingham	Montgomery (198,218)	Huntsville (197,318)	Missouri	Kansas City	Saint Louis (302,838)
Alaska	Anchorage	Juneau (32,113)	Fairbanks (31,516)	Montana	Billings	Missoula (74,428)
American Samoa	Tafuna	Nu'uuli (3,955)	Pago Pago (3,656)	Nebraska	Omaha	Lincoln (287,401)
Arizona	Phoenix	Tucson (545,975)	Mesa (508,958)	Nevada	Las Vegas	Henderson (310,390)
Arkansas	Little Rock	Fort Smith (87,845)	Fayetteville (86,751)	New Hampshire	Manchester	Nashua (89,246)
California	Los Angeles	San Diego (1,425,976)	San Jose (1,030,119)	New Jersey	Newark	Jersey City (265,549)
Colorado	Denver	Colorado Springs (472,688)	Aurora (374,114)	New Mexico	Albuquerque	Las Cruces (102,926)
Connecticut	Bridgeport	New Haven (130,418)	Stamford (129,775)	New York	New York City	Buffalo (256,304)
Delaware	Wilmington	Dover (38,079)	Newark (33,673)	North Carolina	Charlotte	Raleigh (469,298)
District of Columbia	Washington			North Dakota	Fargo	Bismarck (73,112)
Florida	Jacksonville	Miami (470,914)	Tampa (392,890)	Northern Mariana Isl	Saipan2	Tinian (3,136)2
Georgia	Atlanta	Augusta (196,939)	Columbus (194,160)	Ohio	Columbus	Cleveland (383,793)
Guam	Dededo	Yigo (20,539)	Tamuning (19,685)	Oklahoma	Oklahoma City	Tulsa (400,669)
Hawaii	Honolulu1	East Honolulu (49,914)1	Pearl City (47,698)1	Oregon	Portland	Salem (173,442)
Idaho	Boise	Meridian (106,894)	Nampa (96,252)	Pennsylvania	Philadelphia	Pittsburgh (301,048)
Illinois	Chicago	Aurora (199,602)	Naperville (148,304)	Puerto Rico	San Juan	BayamÃ³n (170,480)
Indiana	Indianapolis	Fort Wayne (267,633)	Evansville (117,963)	Rhode Island	Providence	Cranston (81,274)
Iowa	Des Moines	Cedar Rapids (133,174)	Davenport (102,085)	South Carolina	Charleston	Columbia (133,451)
Kansas	Wichita	Overland Park (192,536)	Kansas City (152,958)	South Dakota	Sioux Falls	Rapid City (75,443)
Kentucky	Louisville	Lexington (323,780)	Bowling Green (68,401)	Tennessee	Nashville	Memphis (650,618)
Louisiana	New Orleans	Baton Rouge (221,599)	Shreveport (188,987)	Texas	Houston	San Antonio (1,532,233)
Maine	Portland	Lewiston (35,944)	Bangor (31,997)	Utah	Salt Lake City	West Valley City (136,401
Maryland	Baltimore	Frederick (72,146)	Gaithersburg (68,289)	Vermont	Burlington	South Burlington (19,486
Massachusetts	Boston	Worcester (185,877)	Springfield (155,032)	Virgin Islands (U.S.)	Charlotte Amalie3	Sion Farm (13,003)3
Michigan	Detroit	Grand Rapids (200,217)	Warren (134,587)	Virginia	Virginia Beach	Norfolk (244,076)
Minnesota	Minneapolis	Saint Paul (307,695)	Rochester (116,961)	Washington	Seattle	Spokane (219,190)
Mississippi	Jackson	Gulfport (71,870)	Southaven (54,944)	West Virginia	Charleston	Huntington (46,048)
Missouri	Kansas City	Saint Louis (302,838)	Springfield (168,122)	Wisconsin	Milwaukee	Madison (258,054)
Montana	Billings	Missoula (74,428)	Great Falls (58,701)	Wyoming	Cheyenne	Casper (57,461)

We developed the Twitter crawler application using Python and Java, and Tweepy library to collect tweets from the Twitter database. The Twitter API enables to access the Twitter database and interact with tweets from users. Twitter data allows developers, researchers, and others to study the twitter conversation using the search API, which gives access to the Twitter database for the last seven days. We used Tweepy library, an open source wrapper library that enables your program to communicate with Twitter platform via OAuth authentication protocol. Tweepy supports accessing Twitter via OAuth. It has an in-built method to handle OAuth. Tweepy provides access to the well documented Twitter API. With tweepy, it is possible to get any object and use any method that the official Twitter API offers.

Twitter REST API allows you to retrieve tweets and related information from Twitter. The Twitter standard REST APIs utilize a technique called 'cursoring' to paginate large result sets. Cursoring separates results into pages and provides a means to move backwards and forwards through these pages. Tweepy has a Cursor method provided which takes care of pagination for us. Method API.search is called within the cursor method to mine for tweets based on search parameters. The Cursor coupled with wait_on_rate_limit ensures we can get as many tweets as we need by specifying the count request parameter.

The search API criteria used to collect the tweets are as follows:

1) Keyword- Matches a keyword within the body of a Tweet. Multiple keywords can be specified by using OR operator and placing the words/phrases within parentheses () and wrapping them in double quotes "".
2) Geocode- Search for tweets made from particular locations around the globe by providing the following information [latitude, longitude, radius(mi)]. In order to obtain latitude, longitude coordinates for multiple cities in an efficient manner, Geopy library was used.

3) Geopy is a Python client for geocoding web services. Geopy makes it easy to locate the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources.

4) Language- Only tweets made exclusively in the English language were retrieved for this project.

5) Tweet mode- Extended tweet mode was used in order to capture complete tweets. The character limit was 140 for publishing tweets at the start, but on September 26, 2017, Twitter started testing 280 characters for certain languages.

6) Filter: retweets- In order to obtain only the original tweets and leave out the retweets.

Twitter has a very stringent rate limit that prohibits requests from being made continuously. Tweepy provides a way to overcome this limit by passing wait_on_rate_limit=True into the api object. This removes the need for any complicated automation scripts being used to make sure the tweet gathering process goes on uninterrupted. Since there is a limitation to the number of tweets that can be retrieved from the Twitter database, a scheduler was introduced in our application. The scheduler process calls the crawler process automatically to retrieve the tweets over a period of time. Both the search keywords and the location are passed on as parameters to the crawler. In the Twitter search API's location 'geocode' parameter, the value is specified by the latitude, the longitude, and the radius. Search API returns the tweets by users located within a given radius of that particular latitude and longitude. For example, to retrieve tweets from Bangalore, the 'geocode' parameter is: 12.9715987, 77.5945627, 300mi. The crawler connects to the Twitter database, Twitter authenticates the user, and then pulls the relevant data based on the parameters. Finally, the tweets were dumped into a .json file for further analysis.

3.2. Data Preparation and NLP

After collecting tweets, the next step was to use python, numpy, pandas, NLTK, sklearn, ggplot(), and other libraries to process the tweets and perform sentiment analysis to extract the sentiments of the tweets.

In Twitter, users express their views in one or two sentences only. Hence most of the tweets may not contain full words and many tweets may not carry any meaning. Also, Twitter users are likely to make spelling and grammatical errors. Therefore, extracting information from Twitter messages is challenging, and a lot of time was spent cleaning the tweets before further processing and analysis.

Pre-processing was performed using python NLTK library. Using NLTK built-in libraries and regular expression techniques, we performed pre-processing of tweets. Following are some of the pre-processing techniques that was performed for cleaning tweets before feeding them into the analytics model and sentiment classifier:

(a) Converting to lower case: This has been the very first step in the pre-processing techniques. The tweets were first converted to lower case before applying other methods.

(b) Punctuation: Since the classifier is based on 'bag-of-words' techniques, removing punctuation brings more sanity to input. This involved identifying and removing coma, full stop, semicolon, colon, question marks, exclamation marks, any combination of punctuations, and removing more than one punctuation mark.

(c) Hash Tags and URLs: Many micro-blog messages contain URLs and hash tags. People use the hash-tag symbol # before a relevant keyword or phrase (no spaces) in their tweet to show more easily in Twitter Search. Clicking on a hash-tagged word in any message shows all other Tweets marked with that keyword. Hash-tags can occur anywhere in the tweet – at the beginning, middle, or end. Hash-tagged words are very popular in trending topics. However,

URLs and hash-tags do not contribute to the actual meaning of the message. Hence, both hash-tags and URLs were removed from the tweets.

(d) User name: The raw tweets contain the user name with @ sign in the beginning to identify which user has tweeted that particular message. Since the user name does not contribute to this analysis, the pre-processing technique required the removal of the user name as well as @ symbol.

(e) Stop words: Common words such as 'a', 'an', 'the' etc., do not provide useful information in classifying documents, and thus, it is worthwhile to remove these words. It is commonly known in information retrieval, as 'stop words'. The application used a file which contained these 'stop words' and every tweet was searched against this file to remove the stop words.

(f) Emoticons: In order to convey emotions, such as 'happy', 'sad', 'angry', etc., in the text messages, it is very common for people to use symbols called emoticons. There are 30 emoticons including□, :D, □, =], =[, =(, etc. Since the analysis is based on bag-of-words, all the emoticons were removed from the tweets.

The crawler application programs for the experiments were ran on a quad-core Windows 7 operating systems machine. Tweets from different locations stored as a .json file database. The tweets are further processed and cleaned before the sentiment analysis.

After pre-processing, the tweets were meaningful to understand the pattern and find sentiments associated with the tweets.

3.3. Sentiment Analyser Model

None of the tweets retrieved from the Twitter database had any labels to perform any supervised machine learning sentiment or classification method. Therefore, as an initial step, we developed a unique but powerful NLP unigram method to look for the specific terms and categorize and label the sentiments.

After the initial cleanup of the tweets, using unigram approach to classify tweets to find out the sentimets. For this study, we identified three sentiments. Table 2 describes the sentiments and respective unigram terms to categorize the sentiments. For example, 'Symptoms' sentiment category is the twitter messages with terms such as 'isolation, quarantine, symptoms' and the COVID virus terms. Similarly, the 'Death' sentiment category is the twitter messages with terms such as 'death, dying, died' and the COVID virus terms.

Table 2. Sentiment Category

Sentiment Category	Unigram Terms
COVID virus	COIVD, Virus, corona, coronavirus
Symptoms	Isolation, Quarantine, Symptoms
Death	Died, death, dying.

3.4. Secondary Data Source

Secondary data comprises of the number of COVID cases from John Hopkin's University. The secondary data sources used for this research study are listed in Table 3.

Table 3. Dataset used for the case study

Data	Time Span	Source
Tweets	01 August, 2020 to 31st August, 2020	**http://www.twitter.com**
COVID Reported Cases	01 August, 2020 to 31st August, 2020	**https://coronavirus.jhu.edu/data**

3.5. Sentiment Score

We also developed a new score card to score the sentiments and behaviour of Twitter by comparing with the actual cases reported. The sentiment score is a ratio based on the total number of actual cases reported (data taken from John Hopkins University) and total number of tweets tweeted by users. Sentiment score provides a picture of how people are expressing sentiments in Twitter as a social media. Higher score indicates higher involvement, and lower score indicates lower involvement.

4. RESULTS AND DISCUSSION

Figure 2 shows the tweets distribution across different states of United States. New York state had the highest number ~ 40% of tweets compared with all other states, followed by New Jersey, Maryland and Washington DC. However, when compared with the number of tweets per unit population of each State, the Z-test which indicates whether the value is higher than the mean value showed only 5 states above the mean value. District of Columbia, followed by Delaware, New Jersey, Maryland, and New York exhibited the highest tweets when compared to the mean value (Table 4). New York, which ranked 1st in the total tweets, ranked 5th compared to the number of tweets per unit population. Of the 52 states in US, only 5 states had the number of tweets per unit population above the mean value. Further the tweets were more in the northern states when compared with the southern States.
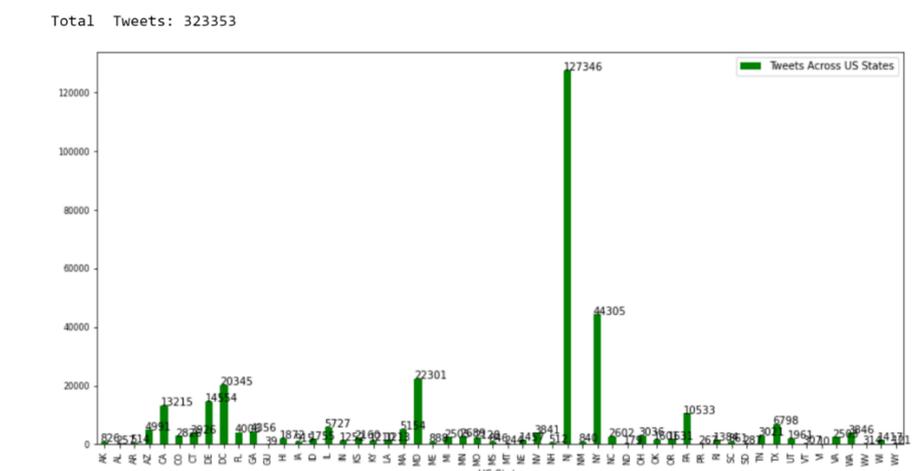

Figure 2. Tweets across different US States

Figure 3, shows that the sentiment analysis of the 'Symptoms' category. This graph shows the people who tweeted when they showed symptoms of COVID-19 or were isolated and quarantined or quarantined due to COVID-19 virus symptoms.
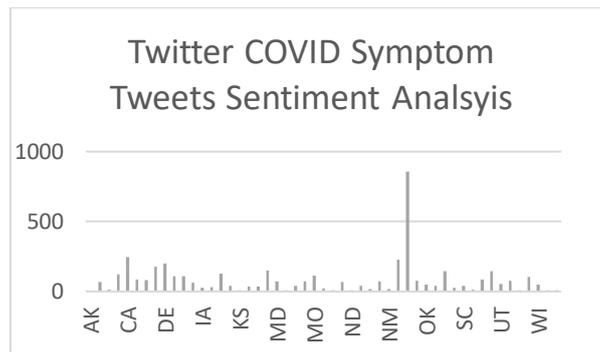
Figure 3. Twitter COVID Symptoms Sentiment Analysis

As we can see from the graph and Table 4, lists the highest number of people who tweeted about their symptoms during August 2020. Top 5 states are from New York, California, Nevada, Washighton DC, Massachusetts.

Table 4. Top 10 Twitter Symptoms Sentiment Tweets State

State	COVIDTweets	SymptomTweets
NY	27426	858
CA	10818	244
NV	3268	226
DE	3023	200
DC	6689	176
MA	4515	152
PA	5408	147
TX	6798	147
IL	4774	127
AZ	4579	122

Figure 4 shows the sentiment score, States of Maryland, New Jersey, Oregon shows the higher scores that means there is a higher correlation of the number of cases reported and people who are tweeting to bring the awareness to the community or to inform their friends and family about their situation.
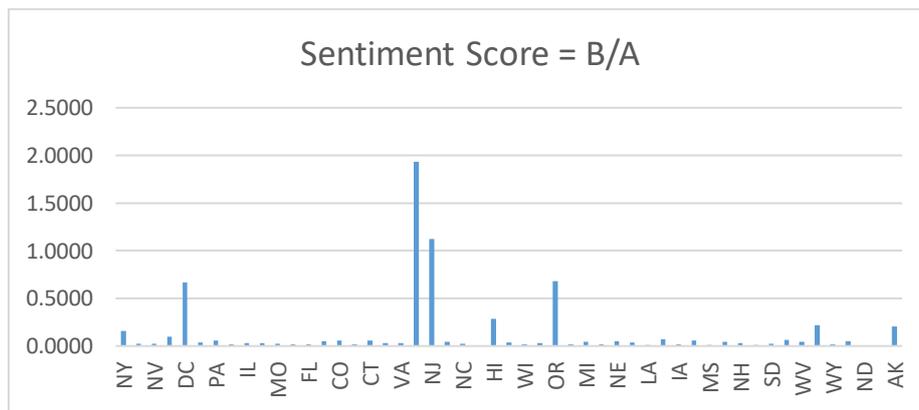

Figure 4. Twitter COVID-19 Symptoms Sentiment Score

Table 4 lists the sentiment scores. States of Maryland, New Jersey, Oregon shows the higher scores compare to other States. This means there is a higher correlation of the number of cases reported and people who are tweeting.

Table 5. Twitter COVID-19 Symptoms Sentiment Score

State	COVIDTweets	SymptomTweets	Reported Cases(John Hopkins)	Sentiment Score = B/A
MD	20265	73	2907358	1.9350
NJ	20784	71	5428251	1.1208
OR	4321	41	668251	0.6815
DC	6689	176	381652	0.6661
HI	1596	65	136543	0.2870
ME	464	8	121956	0.2207
AK	272	3	122403	0.2015
NY	27426	858	5663267	0.1548
DE	3023	200	472774	0.0966
ID	1260	29	790549	0.0692

Similar analysis was performed for the case of death sentiment. This sentiment explains the number of twitter users expressed COVID death sentiment on twitter. These messages contain the news sentiment of someone they know in the family or friends died due to COVID.

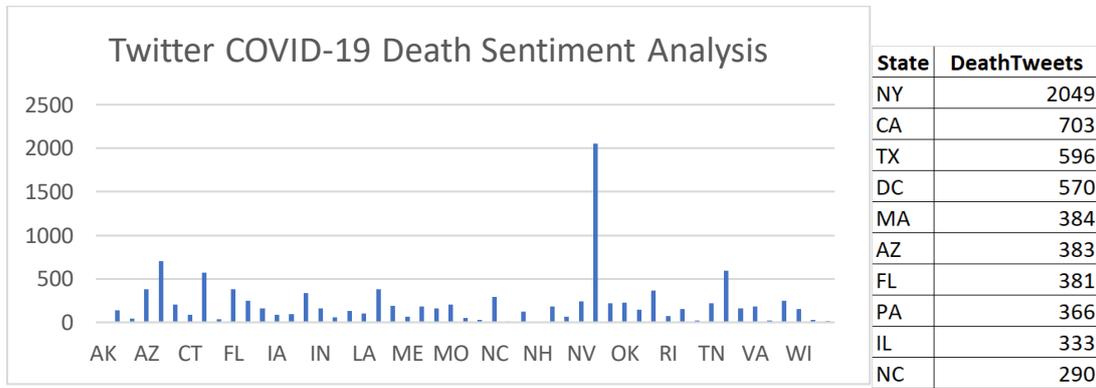| State	DeathTweets
NY	2049
CA	703
TX	596
DC	570
MA	384
AZ	383
FL	381
PA	366
IL	333
NC	290

Figure 5. Twitter COVID Death Sentiment Analysis

Figure 5 and Table 6 shows that the states of New York, California, Texas tops the list of tweets related to death sentiment.

Table 6. Twitter COVID-19 Symptoms Sentiment Score

State	Sentiment Score = B/A	DeathTweets
NY	1.4218	2049
CA	0.5146	703
TX	0.3599	596
DC	4.5465	570
MA	0.2072	384
AZ	0.4238	383
FL	0.2791	381
PA	0.3698	366
IL	0.2962	333
NC	0.4042	290

We also developed a similar score card to score the death sentiments. In this case, we took the reported deaths data from John Hopkins University and the total number of tweets tweeted. Sentiment score provides a picture of how people are expressing sentiments in Twitter social media. Higher score indicates higher involvement and lower score indicates lower involvement.
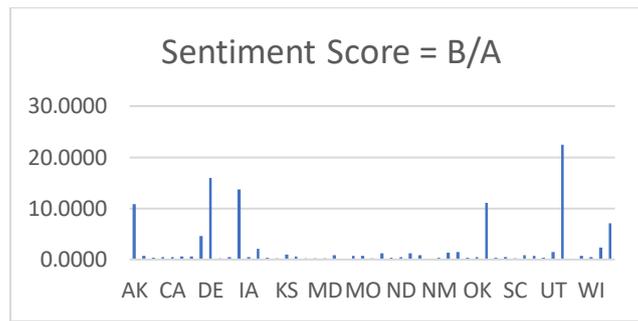
Figure 6. Twitter COVID-19 Symptoms Sentiment Score

Figure 6 and Table 7 shows the sentiment score for the twitter death sentiments, States of Varginia, Delaware, Hawai, Oregon, Arkansa shows the higher scores that means there is a higher correlation of the number of death cases reported and people who are tweeting to inform their friends and family about their loss.

Table 7. Twitter COVID-19 Death Sentiment Score

State	Sentiment Score = B/A	DeathTweets	ReportedDeaths(JOHN)
VA	22.3913	182	1678
DE	16.0045	33	17303
HI	13.6568	157	1188
OR	11.0805	149	11309
AK	10.8672	8	851
WY	7.1485	17	929
DC	4.5465	570	17265
WV	2.3443	30	4636
ID	2.1154	94	7984
UT	1.5030	160	10579

When the total cases were compared with the total COVID tweets, little or no correlation was found and when the number of reported deaths to the total COVID death tweets were compared there was a correlation of 0.40 as shown in figure 7.
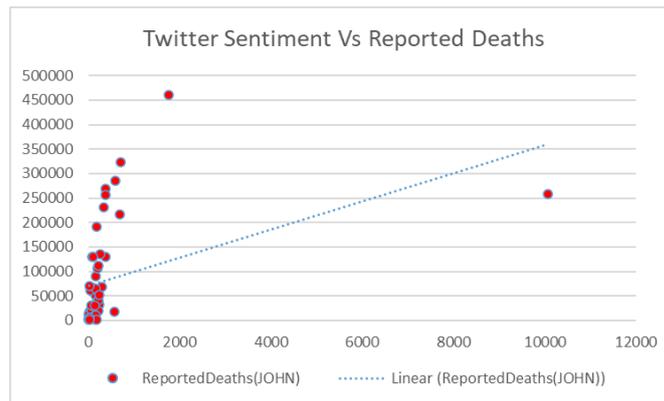

Figure 7. Correlation of Twitter Sentiments to the Reported Cases

4.1. Performance of the Sentiment Analyser Model

Though this is not a machine learning prediction or classifier, we still tested our unique sentiment analyser performance on the sample test data. The following truth Table 8 provides our testing performance with an accuracy of 83%.

Table 8. Performance of the Unigram Sentiment Analyser

Predicted/Predicted	Symptoms	Deaths
Symptoms	85	15
Deaths	20	80

Table 9 summarizes our experiments and tweets sentiment statistics.

Table 9. Tweets Summary

Description	TweetsCount	Probability
TotalTweets	394457	1
COVIDTweets	207792	0.52678
SymptomTweets	4728	0.022754
DeathTweets	22252	0.107088

5. FUTURE WORK

The method used in this work is not the traditional machine learning classification or deep learning technique (Naïve bayes or Neural Network or NER). These conventional supervised machine learning techniques require training data with the classes labeled. Such data is not readily available unless someone manually labels the sentiment classes by looking at every tweet. This process is laborious and time consuming. If the sample data is small, this is possible. However, when we have lots of data, then this process can take lots of time. As a result, we spend more time on data preparation than solving research problems and finding some patterns in the data.

Hence, we used some of the classical NLP and exploratory data analysis model techniques to identify the sentiments using keywords to classify the sentiments. This work has several limitations. All these limitations are part of future work. Here are some limitations and future work under consideration:

1. We have considered only two classes of sentiments. Tweets may carry more categories, and this has to be explored.
2. Our tweets data collection during August 2021 month is based on several limitations such as language, region, cities, etc. Therefore, it just represents a small sample, not the entire Twitter population.
3. Applying supervised machine learning technique to predict twitter sentiments on the same or larger population of tweets. With our current method, data is already classified and this labeled data will be used as training data to create the new AI model.
4. Twitter has lots of slangs, spelling mistakes, grammatical errors. We can still improve data processing methods beyond what we have achieved in this work to enhance the model's efficiency.

6. CONCLUSION

In conclusion, tweets related to COVID demonstrates the two sentiments, one related to the deaths and the other related to the symptoms across all the USA states. The sentiments for each category vary from State to State. States of New York, California, Texas are higher tweets sentiments regarding expressing death sentiment, and states of New York, California, Nevada, are higher regarding sentiments of expressing COVID-19 symptoms. Our unique sentiment analyzer performed this prediction with an accuracy of 83%, and it can be further improved by overcoming certain limitations mentioned.

When we analysed the tweets sentiments expressed with actual reported cases and deaths, the sentiment scores are higher for the states of Maryland, New Jersey and Oregon for the Symptoms sentiment where as for the death sentiments, the states of Virginia, Delaware, and Hawai, showed higher sentiment scores indicating that the twitter users of these states are actively tweeting about symptoms and deaths even though the reported cases are less.

Twitter messages also indicate that people have tweeted more about the deaths than actual symptoms or isolations. The majority of the tweets are just related to COVID-19 and informational.

We also found no or little correlation between the Tweets and the number of cases across all the states. This is merely a case study and does not reflect the overall twitter behavior across the US states during the pandemic. Although some states had the higher number of tweets, it is in no way representative of the tweets over the entire period of the pandemic. Finally, the case study was carried out about the Twitter behaviour in the US for August 2020 only and is not a reflection of the entire pandemic.

REFERENCES

1. Coronavirus Resource Center, August 2020, https://coronavirus.jhu.edu/map.html
2. Depoux, A., Martin, S., Karafillakis, E., Preet, R., Wilder-Smith, A., & Larson, H., "The pandemic of social media panic travels faster than the COVID-19 outbreak", *Journal of Travel Medicine*, Vol. 27, No.3, 2020
3. Kadam, Abhay B., and Sachin R. Atre. "Negative impact of social media panic during the COVID-19 outbreak in India.", *Journal of travel medicine, Vol.* 27. No.3, 2020.
4. Lee, D., Kim, H. S., & Kim, J. K., "The impact of online brand community type on consumer's community engagement behaviors: Consumer-created vs. marketer-created online brand community in online social-networking web sites", *Cyberpsychology, Behavior, and Social Networking*, Vol. 14 No. 1-2, pp. 59-63, 2011.
5. Liu, I.L., Cheung, C.M. and Lee, M.K., "Understanding Twitter usage: What drive people continue to tweet", *Pacific ASia Conference on Information Systems.* Taipei, Taiwan, 2010
6. Radwan, E. and Radwan A., "The spread of pandemic of Social Media Panic during the COVID-19 outbreak", *European Journal of Environment and Public Health, Vol. 4, No.* 2, pp 20-26, 2020.
7. Smith, A. N., Fischer, E., and Yongjian, C., "How does Brand-Related User-Generated Content Differ across YouTube, Facebook, and Twitter?", *Journal of Interactive Marketing,* Vol. 26, No. 2, pp. 102-113, 2912.
8. Twitter Data source, Retrieved from Twitter, *http://www.twitter.com*
9. Ye, Q., Zhang, Z. and Law, R., "Sentiment Classification of Online Reviews to Travel Destinations by Supervised Machine Learning Approaches", *Expert Systems with Applications*, Vol.36, No.3, pp.6527-6535, 2009.
10. Yu M, Li Z, Yu Z, He J, Zhou J., "Communication related health crisis on social media: a case of COVID-19 outbreak", *Current issues in tourism*, Vol.16, pp1-7, April 2020.

11. Zhao, Dejin, and Mary Beth Rosson, "How and why people Twitter: the role that micro-blogging plays in informal communication at work.", In *Proceedings of the ACM 2009 international conference on Supporting group work*, pp. 243-252. 2009.

12. Prabowo, R. and Thelwall, M., "Sentiment analysis: A combined approach", *Journal of Informetrics*, Vol. 3, No. 2, pp.143-157, 2009.

AUTHORS

Prof. Umesh R. Hodeghatta, Ph.D

Dr. Umesh Hodeghatta Rao is an Engineer, a Scientist, and an Educator. He is a faculty member at Northeastern University, MA, USA, specializing in Analytics, AI, Machine Learning, Deep Learning, Natural Language Processing (NLP), Big Data Analytics and Cyber Security. He has more than 25 years of work experience in technical and senior management positions at AT&T Bell Laboratories, Cisco Systems, McAfee, and Wipro. He was also a faculty member at Kent State University, Kent, Ohio, and Xavier Institute of Management, Bhubaneswar, India. He has his master's degree in Electrical and Computer Engineering (ECE) from Oklahoma State University, USA and Ph.D. from the Indian Institute of Technology (IIT), Kharagpur. His research interest is applying AI Machine Learning to strengthen an organization's information security based on his expertise on Information Security and Machine Learning. As a Chief Data Scientist, he is helping business leaders to make decisions and recommendations linked to the organization's strategy and financial goals, reflecting an awareness of external dynamics based on data driven approach.

Dr. Hodeghatta has published many journal articles in international journals and conference proceedings. In addition, he has authored books titled "Business Analytics Using R: A Practical Approach" and "The InfoSec Handbook: An Introduction to Information Security" published by Springer Apress, USA. Dr. Hodeghatta has contributed his services to many professional organizations and regulatory bodies.

Sanath V Haritsa

Despite having a bachelor's degree in mechanical engineering, my keen interest for statistics and programming helped me change my career path. I have been studying machine learning and data science for the last 2 years through online as well as offline courses. Currently working for Nichesoft and Yokogawa Technology Solutions has helped me gain experience in fields of computer vision, performance optimization and fault detection in industries.