Public Sentiment Analysis on Twitter Data during COVID-19 Outbreak

Mohammad Abu Kausar[1], Arockiasamy Soosaimanickam[2], Mohammad Nasar[3]

Department of Information Systems, University of Nizwa, Sultanate of Oman[1, 2]

Department of Computing & Informatics, Mazoon College, Sultanate of Oman[3]

Abstract—The COVID-19 pandemic, is also known as the coronavirus pandemic, is an ongoing serious global problem all over the world. The outbreak first came to light in December 2019 in Wuhan, China. This was declared pandemic by the World Health Organization on 11th March 2020. COVID-19 virus infected on people and killed hundreds of thousands of people in the United States, Brazil, Russia, India and several other countries. Since this pandemic continues to affect millions of lives, and a number of countries have resorted to either partial or full lockdown. People took social media platforms to share their emotions, and opinions during this lockdown to find a way to relax and calm down. In this research work, sentiment analysis on the tweets of people from top ten infected countries has been conducted. The experiments have been conducted on the collected data related to the tweets of people from top ten infected countries with the addition of one more country chosen from Gulf region, i.e. Sultanate of Oman. A dataset of more than 50,000 tweets with hashtags like #covid-19, #COVID19, #CORONAVIRUS, #CORONA, #StayHomeStaySafe, #Stay Home, #Covid_19, #CovidPandemic, #covid19, #Corona Virus, #Lockdown, #Qurantine, #quarantine, #Coronavirus Outbreak, #COVID etc. posted between June 21, 2020 till July 20, 2020 was considered in this research. Based on the tweets posted in English a sentiment analysis was performed. This research was conducted to understand how people from different infected countries cope with the situation. The tweets were collected, pre-processed and then text mining algorithms used and finally sentiment analysis have been done and presented with the results. The purpose of this research paper to know about the sentiments of people from COVID-19 infected countries.

Keywords—*COVID-19; corona virus; corona; pandemic; social media; sentiment analysis; Twitter*

I. INTRODUCTION

Coronavirus disease (COVID-19) first was identified in December 2019 in Wuhan, China and has spread throughout the world covering every region. In 3-4 months this epidemic has disturbed the whole world. The world has witnessed many pandemic periods, but this pandemic today arouses severe economic problems on a country-scale as well as at micro scale. Individuals may experience psychotic symptoms due to pandemic and nations may suffer economic recession due to people with traveling restriction, and all activities pertaining to economics have been closed and social distancing has been imposed. Twenty one million people around the world were reported positive for COVID-19 by mid-August 2020 and nearly 773,072 were dead [1]. Now COVID-19 is increasing at very fast rate, especially in countries like USA and India. COVID-19 has affected more than 215 countries till 18th August 2020. The top 10 countries which have been severely affected by COVID-19 as on date 17th August 2020, includes USA (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) [1].

Diverse use of social networking sites, like Twitter, speeds up the process of sharing information and having views on community events and health crises [2-5]. COVID-19 has been one of Twitter's trending areas throughout January 2020 and it has continued to be debated so far. Since more countries have adopted quarantine measures, people have increasingly relied on various social media sites to get news and expressing their opinion. Twitter data is useful in exposing public debates and feelings about exciting issues and real knowledge of emerging pandemics. In the ongoing COVID-19 pandemic, several government agencies around the world use Twitter as one of the key means of contact to frequently exchange policy updates and news related to COVID-19 with the general public [6]. Increasing numbers of studies have been collected from Twitter data since the COVID-19 outbreak to understand the general public's reactions and conversations related to COVID-19 [7-12]. For example, Abd-Alrazaq and colleagues use Tweets collected between 2nd February and 15th March 2020 to follow topic modeling and sentiment analysis to understand key topics and feelings around COVID-19 [7]. Doctors and individuals mentally more influenced by the epidemics are the most likely to speak about it on social networks like Twitter, which have become significant in our day to day lives.

The Twitter messages created via Twitter are named as Tweets. These data are available in public domain. It can thus be taken as raw data primarily for the extraction of opinions, for the analysis of customer fulfillment and for different rating policy schemes and, ultimately, a study of sentiment has been conducted. Even the online purchases nowadays take place on the basis of people's opinions about different products. For its positivity, advertisers and buying teams need to spend more time evaluating the consumer experience.

This research study has been conducted to identify the sentiments of the people of eleven different infected countries with COVID19 and identifies what emotions people have been sharing from different parts of the world. The countries selected for the study are top 10 [1] infected countries plus one more country from Gulf region, i.e. Oman. The countries selected for the study are USA, Brazil, India, Russia, South Africa, Peru, Mexico, Chile, Spain, UK and Oman.

II. PROBLEM STATEMENT

This research aims to capture, process and evaluate people's feelings within the certain timeframe on the tweets posted on twitter. The study would therefore concentrate on the following questions: i. Collect the tweets through Twitter API using RTweet package in R programming was used. The Hashtag used for collecting the tweet were #covid-19, #COVID19, #CORONAVIRUS, #CORONA, #StayHomeStaySafe, #Stay Home, #StayHomeSaveLives , #Covid_19, #CovidPandemic, #covid19, #CoronaVirus, #Lockdown, #Qurantine, #qurantine, #CoronavirusOutbreak and #COVID. ii. Preprocess the tweets by data cleaning (removing white spaces, links, punctuations, stop words, tokenization, retweet). iii. Calculate the sentiment using syuzhet package and analyze the result.

The tweets posted in English have been considered for a sentiment analysis to understand how people from different infected countries have responded during this pandemic situation to cope with it. The collected tweets will be used, pre-processed and applied with text mining algorithms for performing the sentiment analysis.

III. RELATED WORK

Several researchers have been working on sentiment analysis on different social media data particularly on Twitter, few main contributions that help to discover user attitudes or sentiments in various cases when pandemic happening around the world. This section covers a number of the important papers, which was used as reference.

Researchers analyzed Twitter data for real-time projections of influenza spread and other communicable outbreaks [31]. Researchers measured the emerging risk in an outbreak of influenza in 2009 by analyzing tweet keywords and measuring the incidence of disease in real time and the efforts to prevent disease [32]. Throughout the 2014 outbreak of the Ebola virus, Twitter users shared important health information from media outlets with peak Twitter activities within 24 hours of the news events [33]. [13] Investigated the feelings concerning coronavirus COVID-19, thereby examining the feelings of various people about the pandemic. For this reason, the twitter API used to obtain useful corona virus tweets, and then analyzed based on positive, negative, and neutral emotions with the help of machine learning techniques. Additionally, authors used NLTK library for pre-processing of fetched tweets and the Textblob dataset has been used to evaluate the tweets, after that the exciting results indicates positive, negative, neutral feelings throughout various visualizations.

In [14], the researcher examined and visualized the effect of COVID-19 on the World by implementing certain machine learning methods and algorithms in sentiment analysis on the twitter dataset to recognize positive and negative views people across the globe. It shows that there has been stronger justification for the implementation of Naive Bayes' machine learning approach. In [15], the author drew up a list of COVID-19-related hash tags to looking for specific tweets during 14 days period from 14 to 28 January 2020. Tweets are collected via the API and stored in plain text form. Keywords associated with the level are identified and evaluated for instance, strategies for infection control; vaccination and racial discrimination were also analyzed. At last, the analysis on sentiment data to determine the emotional valence and predominant emotion of each tweet. Ultimately, over time, tweets are analyzed to identify with related topics using an unsupervised method of machine learning. [16] Illustrate observations into the development of anxiety-feeling over time as COVID-19 hit the highest levels in the United States, using textual descriptive analytics assisted by appropriate textual visualization. The author provides a conceptual insight of two important classification methods for machine learning throughout the field of sentiment analysis insights and compares their effectiveness in the classification of varying lengths of Coronavirus Tweets. Authors observe 91 percent classification accuracy for short Tweets using the Naïve Bayes process.

In [17], the authors proposed an effective platform to gather, store, manage, mine, and other activities, called MISNIS (Intelligent Mining of the Influence of Public Social Networks in Society). This program helps non-technical users to quickly mine data and has one of the highest levels of success in Portuguese language tweet collection. [18] Studied the emotional changes using Twitter posts. The understanding of the feeling associated with the text being evaluated [19] is some of the primary insights that can be gained from textual analytics. Twitter social media site includes knowledge which is rich and significant and used as a forum for expressing emotion among its users. Due to the immense number of views, [20] described about twitter is a multi-domain, which covers a wide variety of topics including: education, politics, which goods. One way of analyzing Twitter's large number of views is to apply sentiment analysis. Analysis of sentiment is an application of natural language processing, computer linguistics and text interpretation which classifies text into a division. [21] Sentimental analysis has several applications, for example in businesses, for reviews on products that allow businesses to understand feedback from users and social media reviews to analyze customer reviews. Opinion and sentimental mining were well investigated in this regard, and all the alternative techniques and research areas were discussed. In [22], author addresses Tree kernel and feature-based models used in twitter for sentimental analysis. [23] Reveals the seven (7) years of sentimental review of twitter. Since tweets on Twitter are a particular text not like a regular text, several other works tackle this concern, such as the work on short and concise texts. In [24], author evaluated the data with a large quantity of tweets that were taken as big data and therefore listed the words, sentences or whole records. Authors used the linear method to estimate tweet divisions. This analytical approach did better result and the accuracy was 85.23%. The tree-structured multi-linear principal component analysis (TMPCA) [41] proposed for text classification is a novel data processing technique. To facilitate the machine learning task that follows, the TMPCA can effectively decrease the size of the entire sentence data. In [42], the author proposed a new multi-modal attention (one for text and one for image) Unsupervised neural machine translation model that is trained under an auto-encoding and cycleconsistency paradigm. Tree-structured Multi-linear PCA (TM-PCA) reduces the size of input sequences and sentences, rendering the classification of sentences simple and fast. For text data classification, TM-

PCA with SVM has demonstrated better performance than recurrent neural network [43]. Neural networks may use information from all input sequences to predict each particular output element that is suitable [44].

IV. RESEARCH METHOD

Sentiment analysis in the micro-blogging domain is a relatively recent research field and a fair amount of relevant prior work on user reviews, papers, web posts, articles and general phrase level sentiment analysis has been done. These variations from twitter mostly due to the limit of 280 characters per tweet, which requires the user to express compressed opinion in very short text. Twitter is a platform for microblogging and social networking launched in March 2006. With 330 million active monthly users Twitter is the most popular and reliable social networking platform. Twitter has encouraged the researchers to determine the sentiments on almost everything, including sentiments towards public health information [25], digital technology [29], products [26], natural calamities [30], politics [28], movies [27] etc.

Between 21st June 2020 and 20th July 2020 we created a list of hashtags associated to COVID-19 to check for appropriate tweets. We retrieved the tweets using the advanced programming interface (API) of Twitter's search application and stored them as a CSV format. We carried out a sentiment analysis using tweet text to classify the emotional valence (positive, negative or neutral) of each tweet[34] and prevailing emotions (anger, disgust, fear, happiness, sadness, or surprise)[35]. Eventually, we did topic modeling using an unsupervised method of machine learning to classify and evaluate relevant topics over time within the tweet corpus [36].

A. Data Collection

From 21st June to 20th July 2020, about 1,305,000 tweets were collected from each infected country as shown in Fig. 1. For the tweet collection, RTweet package in R programming was used. The Hashtag used for collecting the tweet were #covid-19, #COVID19, #CORONAVIRUS, #CORONA, #StayHomeStaySafe, #StayHome, #StayHomeSaveLives , #Covid_19, #CovidPandemic, #covid19, #CoronaVirus, #Lockdown, #Qurantine, #qurantine, #CoronavirusOutbreak and #COVID; and the collected Tweets saved in CSV file. The retweets and replies were filtered out while collecting the tweets to avoid duplication of the tweets. As the complete database was obtained, the data cleaning process has been performed, where the white spaces, punctuation, stop words were removed. After the data cleaning process, the NRC Emtoion lexicon was applied with the help of get_nrc_sentiment function to analyze the tweets.

B. Sentiment Analysis

It is about measuring people's feelings, i.e. thoughts about a specific context such as product reviews etc. Sentiment analysis is the process whereby a portion of letters is positive, negative or neutral. A sentiment analysis system for text analysis incorporates natural language processing (NLP) and machine learning techniques to assign weighted feeling scores

within a sentence or phrase to entities, topics, themes and categories. It is believed that the automatic sentiment analysis must also implement finely tuned algorithms to detail the human emotions. Mohammad and Turney [37] not just recognized the positive and negative lexical things, they likewise examined the basic feelings which has been characterized by Plutchik's eight fundamental feelings model. The NRC Word-Emotion Association Lexicon contains 10,170 lexical things which break down the positive and negative extremity as well as recognize the eight feelings characterized by Plutchik [38].

The 10,170 lexical items of the NRC include 1,587 most frequently used nouns, verbs, adverbs and adjectives, 640 words defined by Ekman subset from WordNet Affect Lexicon and 8,132 terms from General Inquirer. Syuzhet package version 1.0.4 [39] has implemented the NRC via open access with the method "src" and is freely available for language R. The syuzhet package has been improved over the years, following several issues raised by the researchers [40]. However, it was also stated that irony and sarcasm are two very complex emotions and are conveyed more on the basis of spoken texts rather than texts such as speeches.

The Syuzhet package in R has been used to compare four sentiment analysis algorithms: syuzhet (default), bing, afinn, and nrc. First, transform the clean text of a tweet into vectors to analyze tweets. A vector is a fundamental data structure in R that contains elements of the same kind, where words and phrases are the elements. When R recognizes a tweet as a variable, the sentiment analysis algorithms will independently evaluate and rate each word and expression. Here the vector function has been used to transform tweets into a vectors and forms the set of words and phrases into a new frame of data.
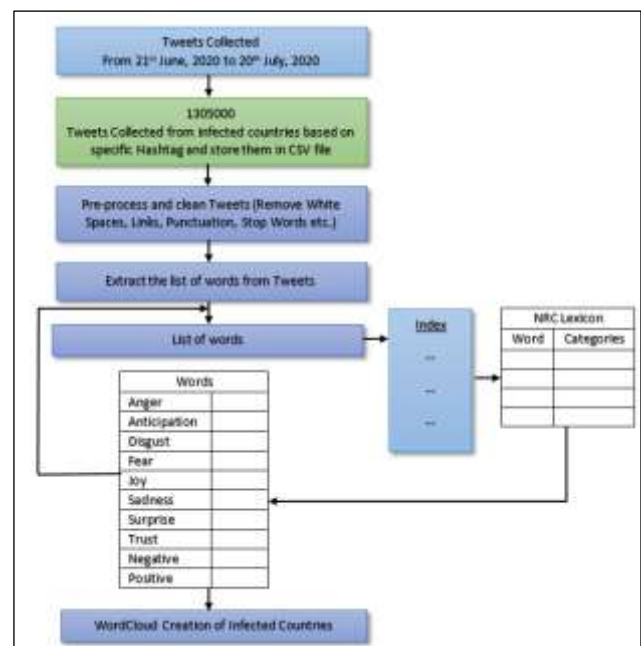

Fig. 1. Flowchart for Sentiment Analysis of the Tweets.

At the Nebraska Literary Lab, the Syuzhet lexicon was created and the name "Syuzhet" comes from the Russian formalists Victor Shklovsky and Vladimir Propp, who split the narrative into two parts, the "fabula" and the "syuzhet". The Syuzhet algorithm is used to evaluate literary works and focuses on how the text elements are constructed and assigns a sentiment score using fractions for each word ranging from -1 to 1. We used the get-sentiment function on the words-df data frame to extract the sentiment values on each tweet and bring the values into a new variable. On the basis of feelings (positive and negative), the syuzhet package classifies the tweets and categorizes them into 8 emotions (fear, joy, anticipation, anger, disgust, sadness, surprise, trust).

V. EXPERIMANTAL SETUP

R programming was used to collect the tweets through Twitter API using RTweet package. Many packages have functions for text classification and also for sentiment analysis in R programming. TM, tidytext, wordcloud, dplyr, syuzhet are some of the packages used. Applications for text mining use tm package in R. Tidytext is used to modify unstructured text data in such a way that it can be analyzed. The Wordcloud package has features that are used to build nice word clouds. Dplyr is a data manipulation grammar that offers a consistent collection of verbs to help you overcome the most common problems in data manipulation. From the syuzhet package, sentiment dictionaries, sentiment derived plots and feelings can be extracted. Datasets from the local library are imported. The collected dataset will be assigned to a corpus variable that can be used in R for preprocessing.

Prior to further processing with the text, all text must be preprocessed. To delete unused document type entry datasets, some text preprocessing methods were used. In text preprocessing, there are several techniques available, only some of them model have been used in this work. Special characters like @, #, / ... have no value adding to the sentiments of the review. The Term Document Matrix is a type of data that is often used in R programming. This is used primarily in the input to obtain word frequencies. The corpus variable can be type cast into a matrix of the text name. We also generate word cloud for the selected data set for better visualization. In the cloud, not all the words are shown, but words with more frequencies are identified and shown as a word cloud.

A. Result and Discussion

A total of 1,305,000 tweets from 11 infected countries were collected during the study period. The results of the research are discussed in two parts.

In the first part, the sentiments of the tweets from all the 11 infected countries were discussed. Fig. 2, shows the sentiments of tweets by the people of the eleven infected countries for which the study was conducted. It is evident from the Fig. 3 that the tweets from almost infected countries had positive sentiments. In countries, like USA and Chile had almost a balance between positive and negative sentiments, whereas, Brazil had 55% positive sentiments and 46% negative sentiments. India was followed by Peru, Spain and UK where 55% of the peoples were tweeting with a positive attitude while

45% with negative attitude. In Russia 51% of the peoples were tweeting with a positive attitude while 49% with negative attitude. In South Africa, Mexico and Oman has 54%, 49% and 57% of the peoples were tweeting with a positive attitude while 46%, 51% and 43% with negative attitude. In Oman maximum people expressed positive attitude since the recovery rate is high.

In the second phase, the emotions associated with the collected tweets were analyzed. In this process, it was observed that almost all countries had the highest number of tweets with more trust, since the recovery rate is high in almost all infected countries. Almost all countries had the highest number of tweets with fear initially due to more number of people infected with COVID-19. The country with most number of tweets associated with Fear and Sadness was India while the anticipation quotient was highest in the tweets from Spain. There were also a good number of tweets with the emotion of Trust which attributed to the invention of medicine is going on different countries. Sentiments are extracted as outputs from functions of the syuzhet package from confidential data and the sentiment approaches categorize the text with its corresponding sentiment meaning. A line chart in Fig. 3 shows the feelings examined from the classified texts.
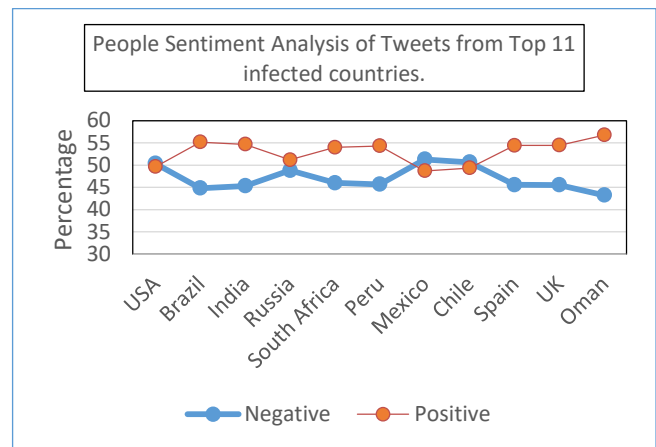

Fig. 2. Sentiment Analysis of Tweets among different Infected Countries.
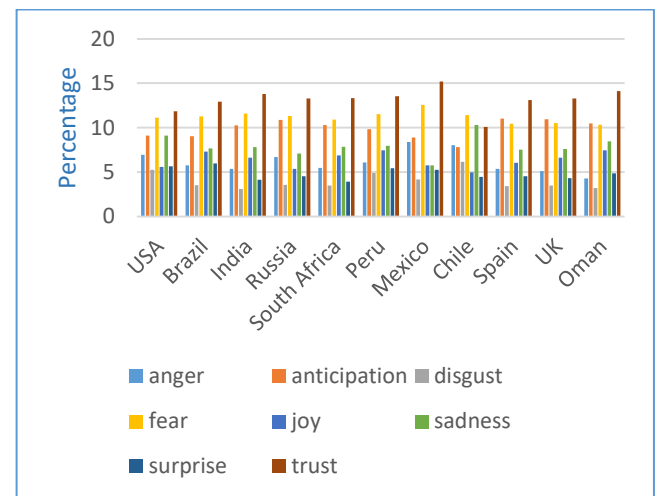

Fig. 3. Emotion Analysis of the Tweets from Infected Country.

After this, the tweets were organized into word clouds to analyze what words have been frequently used by the twitter users of different countries and also what emotions were behind these words. As it can be seen from Fig. 4 to Fig. 14, words like Pandemic, COVID, Virus, Hospitals, Health, Fight, Stay, Safe, Help, Emergency and Death were very frequently used by the users of almost from all countries.

In USA, the words such as trump, Pandemic and Death associated with Surprise, Sadness and Disgust, were the sentiments mostly used in the tweets. In Brazil, words like Pandemic, COVID, Fight, Death and Pandemic were mostly used with the emotions of sadness, anger and disgust respectively. People of India and South Africa used the words in tweets such as Pandemic, Hospital, Death and Fight associated with emotions of sadness, disgust and anger. People in Russia tweeted using the words Pandemic, trump and protest with emotions of sadness, surprise and disgust. People in Peru used the words such as pandemic, stay home and distance with the emotions of sadness and joy. This was evident because the people were now aware of COVID-19 precautions. Mexico had used the words such as covid, help and health with emotions of joy. In Spain, mostly people used the words in tweets such as pandemic, covid and death with emotions sadness, fear and disgust. Citizens of UK were using words like Pandemic, Government and death very frequently which were associated with emotions of sadness, fear and disgust. Similarly, people across Oman used words like Pandemic, covid and death which were used to emote the feeling of sadness and disgust.

However, across all the tweets analysed from eleven infected countries, there was a very good amount of mentions of a political personality. The name of US President, Donald Trump appeared consistently in many tweets across all countries. These mentions were mostly associated with the emotion of surprise.
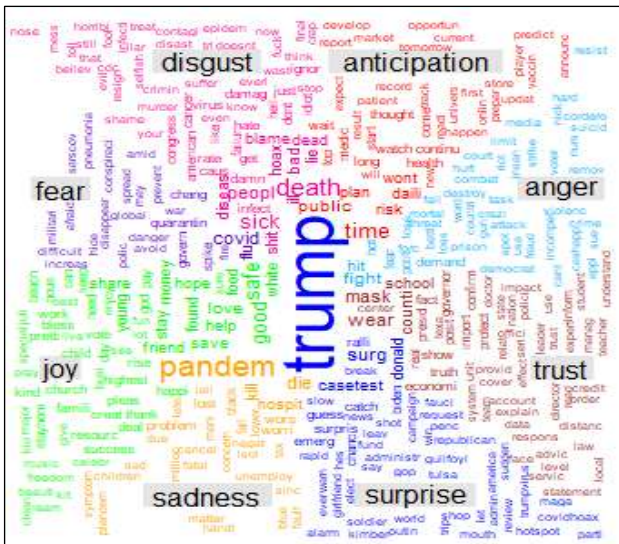

Fig. 4. Word Cloud Tweets from USA.

Fig. 5. Word Cloud Tweets from Russia.
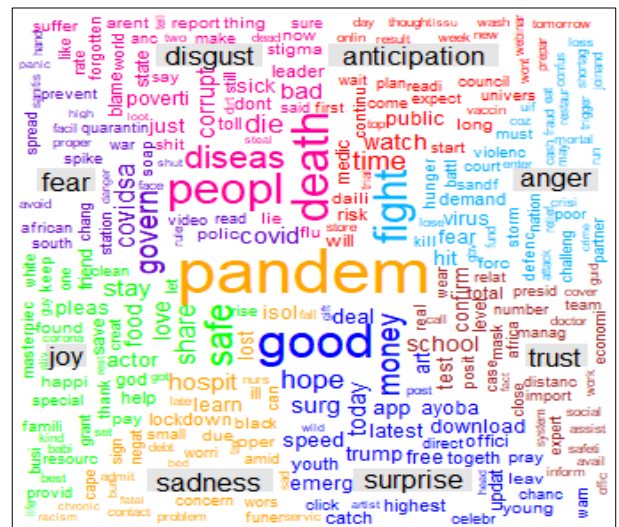

Fig. 6. Word Cloud Tweets from Brazil.

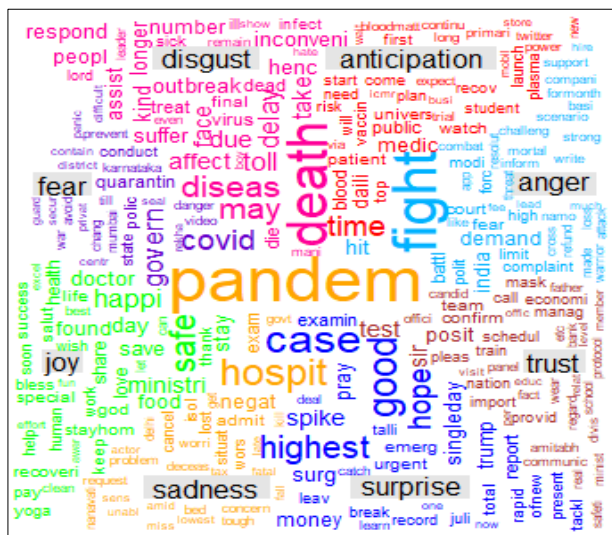Fig. 7. Word Cloud Tweets from South Africa.
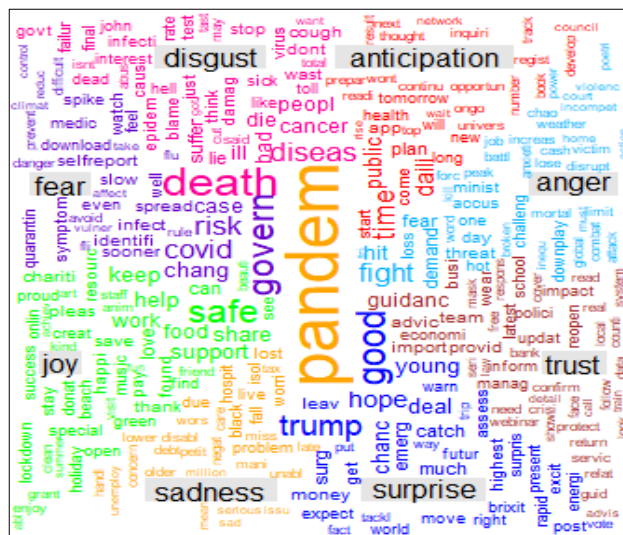
Fig. 8. Word Cloud Tweets from India.

Fig. 11. Word Cloud Tweets from UK.

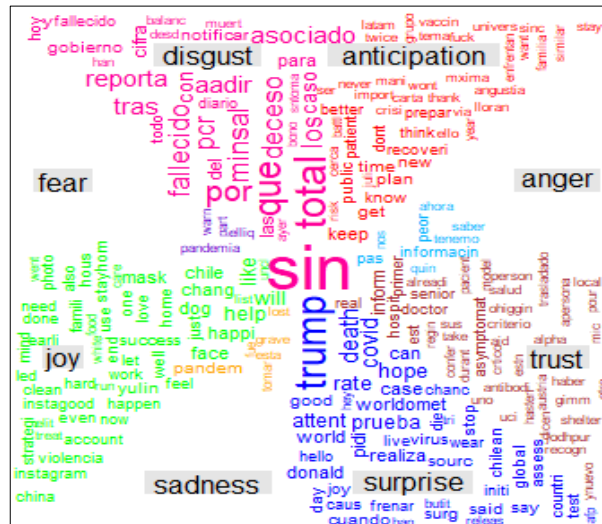Fig. 9. Word Cloud Tweets from Peru.

Fig. 12. Word Cloud Tweets from Chile.

Fig. 10. Word Cloud Tweets from Mexico.

Fig. 13. Word Cloud Tweets from Spain.

Fig. 14. Word Cloud Tweets from Oman.

B. Limitations

It is important to mention that there were some limitations to this study. First, we used a comprehensive hashtag list, which was limited by the knowledge of the authors about trending hashtags and imagination. We have overlooked alternate words or incorrect spellings and implemented with biased selection in the tweets that we have reviewed. Secondly, we addressed tweets in English; thus, our findings may not be generalized one compared to some countries where English is not the main language.

VI. Conclusion

This research work aimed at analyzing the sentiments and emotions of the people during the pandemic COVID-19 have been successfully conducted. To maintain the credibility of data and also the ease of extracting tweets of users, the Twitter platform has been chosen for the study. To have reasonable and good useful data for the study, the twitter users of top 10 infected countries along with one country chosen from the Gulf region for this work. During the study, it was observed that almost all countries were tweeting about COVID19 with a positive sentiments, since all those people become habitual to COVID-19 and also the recovery rate has been improved over the time. Similarly, while analysing the word clouds of different countries, it was concluded that people are tweeting words like Pandemic, COVID, Virus, Hospitals, Health, Fight, Stay, Safe, Help, Emergency, Death and Masks with different emotions. This study provided a good analysis on sentiments and mind sets of people on Covid-19 and enabled us to understand that almost the same level of thinking of people all over the world. This research can be used for future works to examine the shifting emotions and feelings of individuals from these nations and to check if there are noticeable changes over time in them.

References

[1] https://www.worldometers.info/coronavirus/, accessed on July 11, 2020.

[2] Rosenberg, H., Syed, S., & Rezaie, S. (2020). 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, 1-4.

[3] Scanfeld, D., Scanfeld, V., & Larson, E. L. (2010). Dissemination of health information through social networks: Twitter and antibiotics. American journal of infection control, 38(3), 182-188.

[4] Guy, M., Earle, P., Ostrum, C., Gruchalla, K., & Horvath, S. (2010, May). Integration and dissemination of citizen reported and seismically derived earthquake information via social network technologies. In International symposium on intelligent data analysis (pp. 42-53). Springer, Berlin, Heidelberg.

[5] Cheong, M., & Lee, V. C. (2011). A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter. Information Systems Frontiers, 13(1), 45-59.

[6] Rufai, S. R., & Bunce, C. (2020). World leaders' usage of Twitter in response to the COVID-19 pandemic: a content analysis. Journal of Public Health.

[7] Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. (2020). Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. Journal of medical Internet research, 22(4), e19016.

[8] Ahmed W, Vidal-Alaball J, Downing J, López Seguí, F. COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. Journal of Medical Internet Research 2020;22(5):e19458. doi:10.2196/19458.

[9] Alvarez-Risco A, Mejia CR, Delgado-Zegarra J, Del-Aguila-Arcentales S, Arce-Esquivel AA, Valladares-Garrido MJ, . . . Yáñez, JA. The Peru Approach against the COVID-19 Infodemic: Insights and Strategies. The American Journal of Tropical Medicine and Hygiene 2020 June 04:1-4. doi:10.4269/ajtmh.20-0536.

[10] Budhwani H, Sun R. 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 2020;22(5):e19301.

[11] Chen E, Lerman K, Ferrara E. Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. JMIR Public Health and Surveillance 2020;6(2):e19273. doi:10.2196/19273.

[12] Mackey T, Purushothaman V, Li J, Shah N, Nali M, Bardier C, . . . Cuomo R. 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 2020;6(2): e19509. doi:10.2196/19509.

[13] C.Kaur and A. Sharma, "Twitter Sentiment Analysis on Coronavirus using Textblob," EasyChair2516 -2314, 2020.

[14] M. Ra, B. Ab, and S. Kc, "COVID 19 Outbreak: Tweet based Analysis and Visualization towards the Influence of Coronavirus in the World.".

[15] R. J. Medford, S. N. Saleh, A. Sumarsono, T. M. Perl, and C. U. J. m. Lehmann, "An" Infodemic": Leveraging High Volume Twitter Data to Understand Public Sentiment for the COVID 19 Outbreak," 2020.

[16] Samuel, J., Ali, G. G., Rahman, M. M., Esawi, E., & Samuel, Y. (2020). COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification. Information, 11(6), 314. doi:10.3390/info11060314.

[17] Carvalho, J.P.; Rosa, H.; Brogueira, G.; Batista, F. MISNIS: An intelligent platform for twitter topic mining. Expert Syst. Appl. 2017, 89, 374–388.

[18] De Choudhury, M.; Counts, S.; Horvitz, E. Predicting Postpartum Changes in Emotion and Behavior via Social Media. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France, 27 April–2 May 2013; pp. 3267–3276.

[19] Samuel, J.; Holowczak, R.; Benbunan-Fich, R.; Levine, I. Automating Discovery of Dominance in Synchronous Computer-Mediated Communication. In Proceedings of the 2014 47th IEEE Hawaii International Conference on System Sciences, Waikoloa, HI, USA, 6–9 January 2014; pp. 1804–1812.

[20] ansen, B.J., Zhang, M., Sobel, K., Chowdury, A., 2009. Twitter power: Tweets as electronic word of mouth. Journal of the American societyfor information science and technology 60, 2169–2188.

[21] Wilson, T., Wiebe, J., & Hoffmann, P. Recognizing contextual polarity in phraselevel sentiment analysis. In Proceedings of the conference on human language technology and empirical methods in natural language processing (pp. 347-354). Association for Computational Linguistics(2005, October).

[22] Liu, B. Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167(2012).

[23] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. Sentiment analysis of twitter data. In Proceedings of the workshop on languages in social media (pp. 30- 38). Association for Computational Linguistics(2011, June).

[24] Sulthana, A. R., Jaithunbi, A. K., & Ramesh, L. S. (2018). Sentiment analysis in twitter data using data analytic techniques for predictive modelling. Journal of Physics: Conference Series, 1000, 012130. doi:10.1088/1742-6596/1000/1/012130.

[25] Scanfeld, D., Scanfeld, V., & Larson, E. L. (2010). Dissemination of health information through social networks: Twitter and antibiotics. American journal of infection control, 38(3), 182-188.

[26] Jagdale, Rajkumar S., Vishal S. Shirsat, and Sachin N. Deshmukh. "Sentiment analysis on product reviews using machine learning techniques." In Cognitive Informatics and Soft Computing, pp. 639-647. Springer, Singapore, 2019.

[27] Uma Ramya V, Thirupathi Rao K. Sentiment Analysis of Movie Review using Machine Learning Techniques. International Journal of Engineering & Technology. 2018;7(2.7):676.

[28] M A, Ravikumar P. Survey: Twitter data Analysis using Opinion Mining. International Journal of Computer Applications. 2015;128(5):34-36.

[29] Maindola, Pallavi, Neetu Singhal, and Akash D. Dubey. "Sentiment Analysis of Digital Wallets and UPI Systems in India Post Demonetization Using IBM Watson." In 2018 International Conference on Computer Communication and Informatics (ICCCI), pp. 1-6. IEEE, 2018.

[30] Wang Y, Taylor J. Coupling sentiment and human mobility in natural disasters: a Twitter-based study of the 2014 South Napa Earthquake. Natural Hazards. 2018;92(2):907-925.

[31] Chorianopoulos K, Talvis K. Flutrack.org: Open-source and linked data for epidemiology. Health Informatics J 2016; 22:962–974.

[32] Signorini A, Segre AM, Polgreen PM. The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic. PLoS ONE 2011; 6:e19467.

[33] Househ M. Communicating Ebola through social media and electronic news media outlets: A cross-sectional study. Health Informatics J 2016; 22:470–478.

[34] Matthew L. Jockers. Syuzhet: Extract Sentiment and Plot Arcs from Text. 2015. Available at: https://github.com/mjockers/syuzhet. Accessed 30 January 2020.

[35] Colneric N, Demsar J. Emotion Recognition on Twitter: Comparative Study and Training a Unison Model. IEEE Trans Affective Comput 2019; :1–1.

[36] David M. Blei, Andrew Y. Ng, Michael I. Jordan. Latent dirichlet allocation. The Journal of Machine Learning Research 2003; 3:993–1022.

[37] Mohammad S, Turney P. Crowdsourcing A Word-Emotion Association Lexicon. Computational Intelligence. 2012;29(3):436-465.

[38] Plutchik R. The Nature of Emotions. American Scientist. 2001;89(4):344.

[39] Jockers M. Introduction to the Syuzhet Package [Internet]. Cran.r-project.org. 2017. Available at: https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html.

[40] Problems with the Syuzhet Package [Internet]. Anglophile in Academia: Annie Swafford's Blog. 2015 [cited 2020 Apr 3]. Available from: https://annieswafford.wordpress.com/2015/03/02/syuzhet/.

[41] Su, Yuanhang, Ruiyuan Lin, and C-C. Jay Kuo. "Tree-structured multi-stage principal component analysis (TMPCA): Theory and applications." Expert Systems with Applications 118 (2019): 355-364.

[42] Su, Yuanhang, et al. "Unsupervised multi-modal neural machine translation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.

[43] Su, Yuanhang, Yuzhong Huang, and C-C. Jay Kuo. "Efficient text classification using tree-structured multi-linear principal component analysis." 2018 24th International Conference on Pattern Recognition (ICPR). IEEE, 2018.

[44] Su, Yuanhang, and C-C. Jay Kuo. "On extended long short-term memory and dependent bidirectional recurrent neural network." Neurocomputing 356 (2019): 151-161.