

Twitter Sentiment Analysis during COVID19 Outbreak

Dr. Akash D Dubey

Assistant Professor

Jaipuria Institute of Management, Jaipur

drakashddubey@gmail.com

Abstract

On 11th March 2020, World Health Organization announced COVID19 outbreak as a pandemic. Starting from China, this virus has infected and killed thousands of people from Italy, Spain, USA, Iran and other European countries as well. While this pandemic has continued to affect the lives of millions, a number of countries have resorted to complete lockdown. During this lockdown, people have taken social networks to express their feelings and find a way to calm themselves down. In this research work, country wise sentiment analysis of the tweets has been done. This research work has taken into account the tweets from twelve countries. These tweets have been gathered from 11th March 2020 to 31st March 2020, and are related to COVID19 in some or the other way. This analysis has been done to analyse how the citizens of different countries are dealing with the situation. The tweets have been collected, pre-processed, and then used for text mining and sentiment analysis. The results of the study concludes that while majority of the people throughout the world are taking a positive and hopeful approach, there are instances of fear, sadness and disgust exhibited worldwide. However, four countries, France, Switzerland, Netherland and United States of America have shown signs of distrust and anger on a bigger scale as compared to remaining eight countries.

Keywords: COVID19, Pandemic, Corona Virus, Twitter, Sentiment Analysis

Introduction

The origin of COVID19 is said to be in the starting of December 2019, when several patients from Wuhan, Hubei Province reported severe respiratory infections. These patients had a background of working in the wholesale fish and seafood market, also known as wet markets [1]. In January 2020, the markets were completely closed down and disinfectants were used to sanitize them. On 7th January 2020, the researchers isolated a novel coronavirus which was referred as SARS-CoV-2 or 2019-nCoV. Initially the World Health Organization denied the possibilities of human-to-human transmission 2019-nCoV on 11th January 2020 [2]. However, the confirmed cases continued to soar and on 30th January 2020, World Health Organization declared this COVID19 a Public Health Emergency of International Concern (PHEIC) and an epidemic [3].

By the end of January, the novel coronavirus had already started spreading out to other countries steadily. The number of patients affected by this virus globally was 11,950 on 31st January, 69,197 on 15th February and 86,604 on 28th February. By the time when WHO finally decided to declare COVID19 on 11th March 2020, the number of patients had increased to 126,214, with a total of 4,628 casualties. However, there was an exponential increase in the number in the month of March. On 20th March, there were total 275,550 coronavirus affected patients, which meant a rise of more than 100% within 9 days. In next 11 days, the number of

COVID19 patients increased to 858,361, showing a rise of more than 211% and a total of 47,192 deaths were reported till 31st March 2020.

COVID19 has affected more than 166 countries till 31st March 2020. While the case fatality rate of the virus was reported to be 2.5% on 16th February 2020 [4], recent studies show that the CFR for COVID19 can range up to 9.26% [5]. The countries which have been severely affected by COVID19 includes USA (188,530 patients on 31st March), Italy (105,792 patients on 31st March), Spain (68,200 patients on 31st March), Germany (71,808 patients on 31st March), China (81,554 patients on 31st March), France (52,128 patients on 31st March), UK (23,226 patients on 31st March), Switzerland (16,605 patients on 31st March), Belgium (12,775 patients on 31st March), Netherland (12,595 Patients on 31st March) and Australia (4,763 patients on 31st March) [6].

World Health Organization has suggested that isolation and self-quarantine is one of the major ways to stop this pandemic to spread with such an alarming rate. China has witnessed benefits of the one of the largest lockdown at the start of this pandemic, where it locked down 20 provinces and regions. On March 18th, China reported no new local cases for the first time since this pandemic began. This has also encouraged the Chinese government to decide that the lockdown will be lifted on 8th April [7]. Following the suit, countries like Jordan, Argentina, Israel, Belgium and other countries have locked down their countries as well. Similarly, on 25th March, the Indian government took a major decision of locking down the whole nation for 21 days. This can easily be the biggest lockdown the world has ever seen, with 1.3 billion citizens of India being locked down for three weeks [8].

One of the most famous micro blogging site, Twitter has been one of the major ways for information sharing and self-documentation [9]. As the world is fighting with COVID19 since last two month and majority of the people are under lockdown, the importance of Twitter has increased more than ever. Even in the past, people have been using twitter to communicate, express and disseminate information related to the crisis, be it cyclones [10], ebola [11], floods [12] or Zika [13]. Twitter has been one of the platforms for millions to express their emotions regarding different issues.

This research study has been to identify the sentiments of the citizens of 12 different countries regarding COVID19 and identify what emotions people have been sharing from different parts of the world. The countries selected for the study are USA, Italy, Spain, Germany, China, France, UK, Switzerland, Belgium, Netherland and Australia. This study also included India since some experts believe that this nation with 1.3 billion population can be the next epicentre for COVID19 [14]. Hence it becomes trivial to include India along with the above mentioned 11 countries which are already facing the heat of COVID19.

Methodology

Launched in 2006, the microblogging platform Twitter became one of the most popular microblogging provider. Initially, the users were given 140 characters space to tweet but recently, this has been extended to 280 characters. With a total number of 330 million daily active users, Twitter has found a huge following throughout the world. People use twitter to share their thoughts, as well as to disseminate the information. The rapid sharing of the opinions of the user on Twitter has encouraged the researchers to determine the sentiments on almost everything, including sentiments towards products [15], movies [16], politics [17], digital technology [18] and natural calamities [19]

While there is no single accepted psychological theory of basic human emotions, most of the studies accept the theory that a simple positive-negative dichotomy cannot be used to decide

the human emotions as a whole. On the same lines, it is believed that the automatic sentiment analysis must also implement finely tuned algorithms to detail the human emotions. In their research work, Mohammad and Turney [20][21] have followed the same lines, where they not only identified positive and negative lexical items, they also analyzed the underlying emotions which has been defined by Plutchik's eight basic emotions model. The NRC Word-Emotion Association Lexicon contains 10,170 lexical items which not only analyze the positive and negative polarity but also detect the eight emotions defined by Plutchik [22].

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 1.0.1 [23] 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 [24]. 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 classifies the tweets on the basis of sentiments (positive and negative) and also categorizes them into 8 emotions (fear, joy, anticipation, anger, disgust, sadness, surprise, trust).

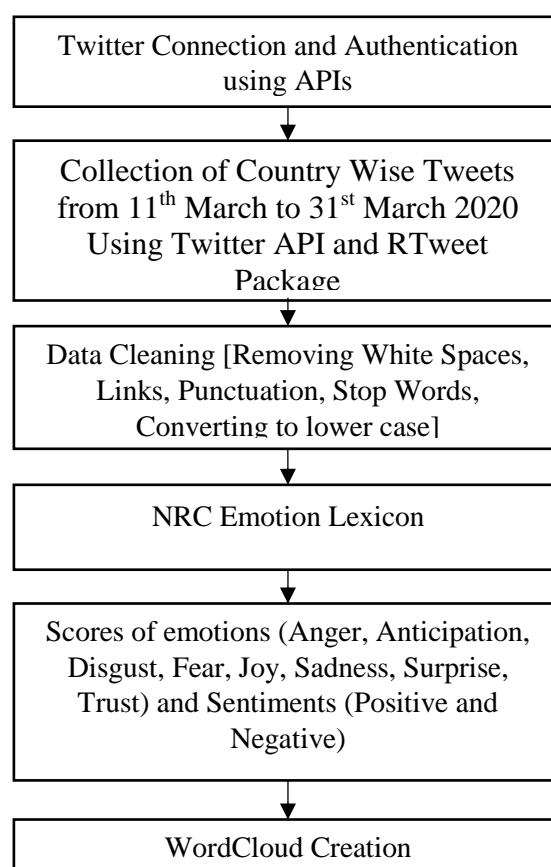


Figure 1 Flowchart for Sentiment Analysis of the tweets

The sentiment and emotion analysis for the tweets were being done as illustrated in figure 1. Overall, 50,000 tweets were collected from each country after every 4 days. For the collection, RTweet package in R was used. The keywords used for collecting the tweet were COVID19, COVID-19, CORONAVIRUS, CORONA, STAY HOME STAY SAFE and COVID19Pandemic. The retweets and replies were filtered out while collecting the tweets to

avoid duplication of the tweets. When the complete database was obtained, the data cleaning process was performed, where the white spaces, punctuation, stop words were removed and the tweets were converted to lower case. After the data cleaning, the NRC Emtoion lexicon was applied with the help of get_nrc_sentiment function to analyze the tweets. Once the scoring of the tweets was done on the basis of sentiments and emotions, corpus was created in order to develop the word cloud for each country.

Results and Discussion

The results of the research will discussed in two phases. In the first phase, the sentiments of the tweets from all the 12 countries will be discussed.

Figure 2 shows the sentiments of tweets that have been done by the citizens of the twelve countries for which the study was conducted. As it is evident from the figure the most positive sentiments were reflected from the tweets of Belgians where almost 63% of the population had positive sentiments and only 37% had negative sentiments. Belgium was followed by India where 60% of the citizens were tweeting with a positive attitude while 40% had negative perspective. The third most positive sentiments were reflected from Australians. On the other hand, Germany, France, Switzerland and USA were the countries where there was almost a balance between positive and negative sentiments. China, on the contrary, had more people who were tweeting with negative sentiments (~55%). This difference can be because China has reportedly seen the peak of the curve and is now following a downward trend, while other countries are still fighting to stop the spread and are optimistic about it.

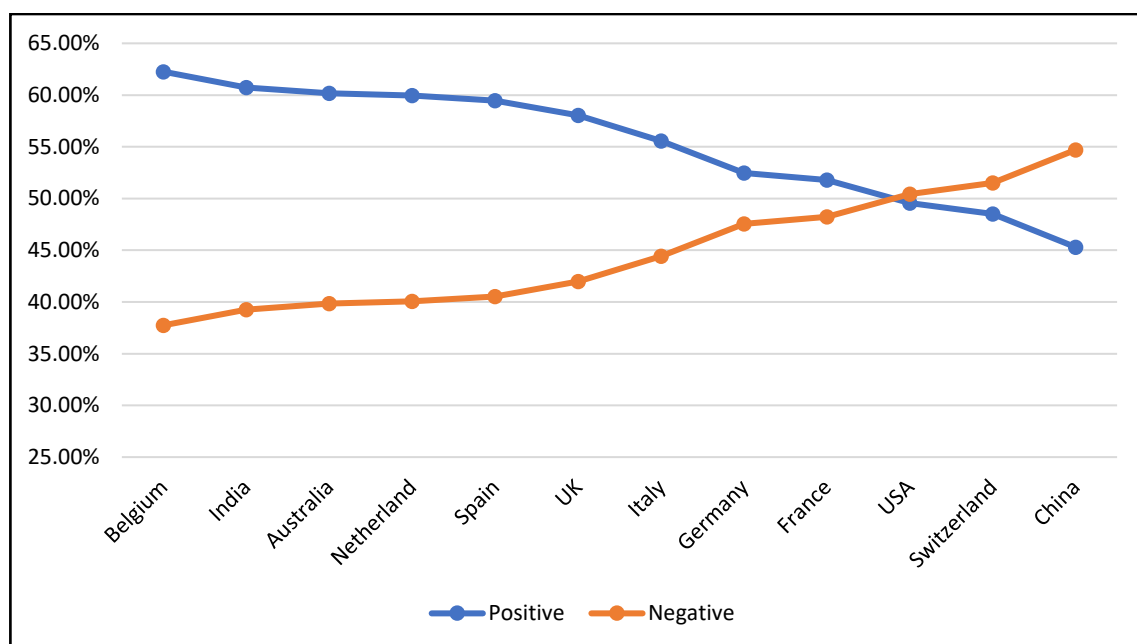


Figure 2 Sentiment Analysis of Tweets across different countries

In the second phase of the discussion, the emotions which are associated with the collected tweets were analyzed. In this process, it was observed that the USA, France and China had the highest number of tweets with anger. Switzerland had the highest number of tweets which had emotions of sadness and fear quotient. The country with most number of tweets associated with Trust and Surprise was Belgium while the anticipation quotient was highest in the tweets from Germany. There was also a good number of tweets with the Joy emotion which attributed to the various jokes and memes that may have been shared by the user amid lockdown.

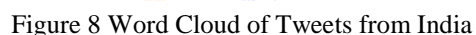
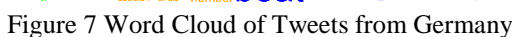
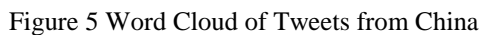
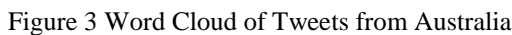
Table 1 Emotion Analysis of the tweets from each country

	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust
Australia	9.19%	16.74%	6.17%	19.33%	7.42%	13.71%	5.67%	21.75%
Belgium	9.88%	17.52%	5.31%	15.57%	8.74%	7.28%	9.80%	25.91%
China	10.67%	17.30%	4.35%	21.38%	6.16%	10.17%	5.56%	24.41%
France	18.20%	10.04%	5.22%	22.39%	4.31%	18.64%	5.80%	15.39%
Germany	7.88%	19.93%	4.20%	20.19%	6.75%	12.12%	7.22%	21.71%
India	9.69%	16.63%	4.62%	18.72%	12.45%	10.16%	6.53%	21.20%
Italy	7.31%	16.15%	4.63%	17.75%	11.25%	12.20%	8.23%	22.47%
Netherland	7.33%	13.27%	5.14%	23.04%	7.72%	19.69%	4.94%	18.87%
Spain	8.95%	16.01%	5.06%	18.66%	9.65%	12.85%	7.11%	21.70%
Switzerland	6.29%	9.98%	4.65%	28.08%	6.39%	24.47%	4.74%	15.40%
UK	8.47%	17.00%	5.12%	17.23%	11.12%	11.92%	7.28%	21.87%
USA	15.88%	9.30%	6.69%	17.23%	10.65%	12.28%	8.55%	19.42%

After this the tweets were organized into word clouds to analyze what words have been frequently used by the twitters users of different countries and also, what emotions were behind these words. As it can be seen from figure 2-13, words like Virus, Political, Hospitals, Pandemic, Die, Fight, Stay, Safe, Quarantine, Emergency and Death were very frequently used by the users of each country. China also got a considerable amount of mentions in these tweets, considering that the origin point of COVID19 was Wuhan, China.

In Australia, the words Flatten, Spread and Saying associated with anticipation were the most used words in the tweets. In Belgium, words like Political, Government and Pandemic were most used words with the emotions of disgust and fear respectively. China witnessed the usage of words like Pandemic and Epidemic associated with emotions of sadness and disgust. Similarly, people in France tweeted using the word Dire and with disgust emotions the most. Citizens of India used the word Fight and Stay Safe more frequently, and these words were associated with emotions of Anger and Joy. This trend in India may be attributed to the fact that India is still in early stages of COVID19 stages and has been locked down much earlier. People in Italy used the word Salute with the emotions of joy frequently. This was evident because the citizens were full of praises for the healthcare workers, doctors and police who have been serving them in this phase. Netherland had words like Ugly, Bad, Die and Pandemic trending with emotions of disgust and sadness. Netherland also had the words friends and share trending which were associated with emotions of joy. Spain followed the suit of Netherland, where people used the word Pandemic, Death and Fight more frequently. Citizens of Switzerland were using words like Political, Govt, Die, Pandemic, Hospitals very frequently which were associated with emotions of disgust, sadness and surprise. Similarly, people across UK used words like Pandemic, government, isolation which were used to emote the feeling of disgust and sadness.

However, across all the tweets analyzed from these twelve countries, there was a very good amount of mentions of a political personality. The name of US President, Donald Trump appeared consistently in the tweets across all the countries. These mentions were mostly associated with the emotion of surprise. It can be assumed that this emotion is consistent with the US President owing to his decision to not to lockdown the country even when the number of cases has gone way beyond 100,000. Bill Gates, the founder of Mircosoft also criticized the move and called for a 10 week shutdown across country to escape from COVID19 [25].



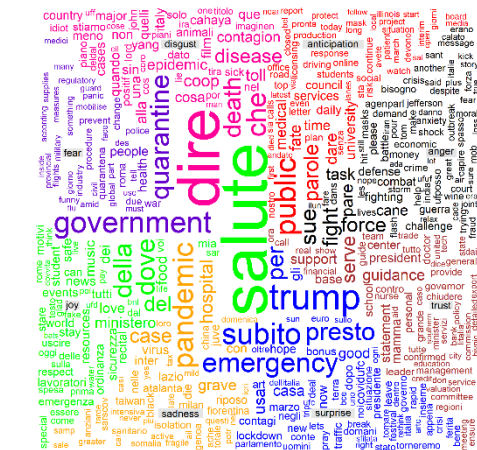


Figure 9 Word Cloud of Tweets from Italy

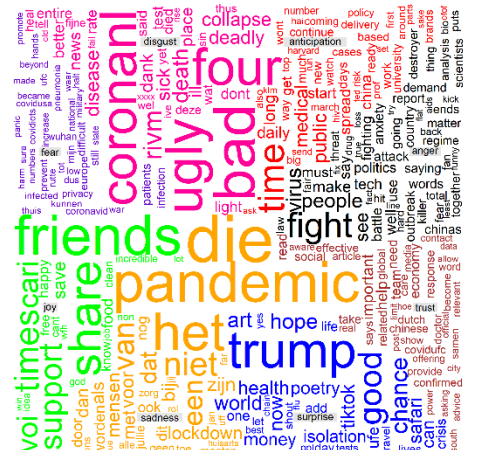


Figure 10 Word Cloud of Tweets from
Netherland



Figure 11 Word Cloud of Tweets from Spain



Figure 12 Word Cloud of Tweets from Switzerland

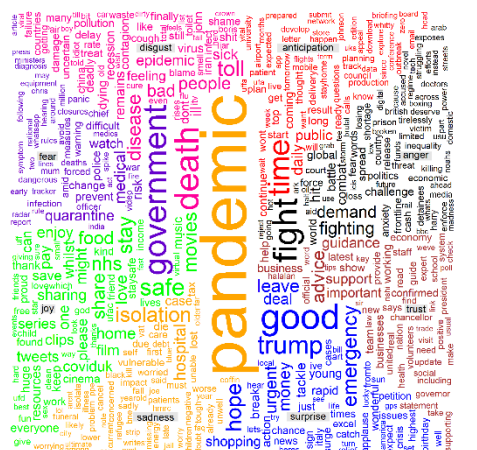


Figure 13 Word Cloud of Tweets from UK

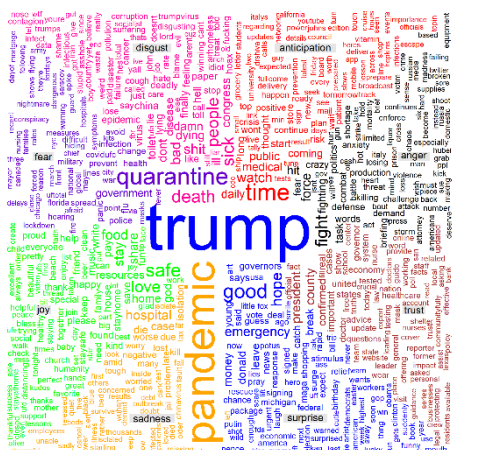


Figure 14 Word Cloud of Tweets from USA

Limitation, Conclusion and Future Scope

This research work aimed at analyzing the sentiments and emotions of the people during the pandemic COVID19. During the study, it was revealed that countries like Belgium, India and Australia were tweeting about COVID19 with a positive sentiment, people in China had negative sentiments about the same. Similarly, while analyzing the word clouds of different countries, it was concluded that people are tweeting words like Pandemic, Death, Quarantine, Hope, Stay Safe, Government, Political, Fight and Masks with different emotions. The name of the USA President, Donald Trump was amongst one of the most tweeted words not only in USA, but across all the twelve countries considered for the study.

The tweets which were collected for this study were in English language which might serve as a limitation for the study. Also, while NRC Lexicon used for this study analyzed the tweets for eight different emotions, it does not count the emotions of sarcasm and irony.

For the future works, this study can be used to analyze the changing emotions and sentiments of people from these countries and check whether there are major shifts in them over the period of time. It is expected that as the spread of this pandemic will increase, the sentiments and emotions in the tweets may change on the lines of what was seen in the case of China.

References

1. Huang, Chaolin, Yeming Wang, Xingwang Li, Lili Ren, Jianping Zhao, Yi Hu, Li Zhang et al. "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China." *The Lancet* 395, no. 10223 (2020): 497-506.
2. (WHO), World Health Organization. "Preliminary Investigations Conducted by the Chinese Authorities Have Found No Clear Evidence of Human-to-Human Transmission of the Novel #Coronavirus (2019-NCoV) Identified in #Wuhan, #ChinaCN. Pic.twitter.com/Fnl5P877VG." Twitter, Twitter, 14 Jan. 2020, twitter.com/WHO/status/1217043229427761152.
3. Jiang, Fang, Liehua Deng, Liangqing Zhang, Yin Cai, Chi Wai Cheung, and Zhengyuan Xia. "Review of the clinical characteristics of coronavirus disease 2019 (COVID-19)." *Journal of General Internal Medicine* (2020): 1-5.
4. "Chinese Center For Disease Control And Prevention". 2020. *Chinacdc.Cn*. <http://www.chinacdc.cn/en/>.
5. Khafaie, Morteza Abdullatif, and Fakher Rahim. "Cross-Country Comparison of Case Fatality Rates of COVID-19/SARS-COV-2." *Osong Public Health and Research Perspectives* 11, no. 2 (2020): 74.
6. "Coronavirus Update (Live): 1,410,095 Cases And 81,010 Deaths From COVID-19 Virus Pandemic - Worldometer". 2020. *Worldometers.Info*. <https://www.worldometers.info/coronavirus/>.
7. Nectar Gan, CNN. 2020. "China To Lift Lockdown On Wuhan, Ground Zero Of Coronavirus Pandemic". *CNN*. <https://edition.cnn.com/2020/03/24/asia/coronavirus-wuhan-lockdown-lifted-intl-hnk/index.html>.
8. "India's 1.3Bn Population Told To Stay At Home". 2020. *BBC News*. <https://www.bbc.com/news/world-asia-india-52024239>.
9. Liu, Ivy LB, Christy MK Cheung, and Matthew KO Lee. "Understanding Twitter Usage: What Drive People Continue to Tweet." *Pacis* 92 (2010): 928-939.
10. Soriano, Cheryll Ruth, Ma Divina Gracia Roldan, Charibeth Cheng, and Nathaniel Oco. "Social media and civic engagement during calamities: the case of Twitter use during typhoon Yolanda." *Philippine Political Science Journal* 37, no. 1 (2016): 6-25.
11. Van Lent, Liza GG, Hande Sungur, Florian A. Kunneman, Bob Van De Velde, and Enny Das. "Too far to care? Measuring public attention and fear for Ebola using Twitter." *Journal of medical Internet research* 19, no. 6 (2017): e193.
12. Nair, Meera R., G. R. Ramya, and P. Bagavathi Sivakumar. "Usage and analysis of Twitter during 2015 Chennai flood towards disaster management." *Procedia computer science* 115 (2017): 350-358.
13. Fu, King-Wa, Hai Liang, Nitin Saroha, Zion Tsz Ho Tse, Patrick Ip, and Isaac Chun-Hai Fung. "How people react to Zika virus outbreaks on Twitter? A computational content analysis." *American journal of infection control* 44, no. 12 (2016): 1700-1702.
14. "India Could Be Next Coronavirus Hotspot With 'Avalanche' Of Cases". 2020. *South China Morning Post*. <https://www.scmp.com/news/asia/south-asia/article/3075662/india-could-be-next-coronavirus-hotspot-avalanche-cases>.

15. 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.
16. 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.
17. M A, Ravikumar P. Survey: Twitter data Analysis using Opinion Mining. *International Journal of Computer Applications*. 2015;128(5):34-36.
18. 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.
19. 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.
20. Mohammad S, Turney P. CROWDSOURCING A WORD-EMOTION ASSOCIATION LEXICON. *Computational Intelligence*. 2012;29(3):436-465.
21. Mohammad S, Turney P. Nrc emotion lexicon [Internet]. 2013 [cited 2020 Apr 3]. Available from: <http://www.saifmohammad.com/WebDocs/NRCemotionlexicon.pdf>
22. Plutchik R. The Nature of Emotions. *American Scientist*. 2001;89(4):344.
23. Jockers M. Introduction to the Syuzhet Package [Internet]. Cran.r-project.org. 2016 [cited 2020 Apr 3]. Available from: <https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html>
24. 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/>
25. DeCambre M. Bill Gates on Trump call for quick end to lockdown: It's tough to tell people 'keep going to restaurants, go buy new houses, ignore that pile of bodies over in the corner' [Internet]. MarketWatch. 2020 [cited 2020 Apr 3]. Available from: <https://www.marketwatch.com/story/bill-gates-on-trump-call-for-quick-end-to-lockdown-its-tough-to-tell-people-keep-going-to-restaurants-go-buy-new-houses-ignore-that-pile-of-bodies-over-in-the-corner-2020-03-24>