2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS) | 978-0-7381-1180-3/20/$31.00 ©2020 IEEE | DOI: 10.1109/SNAMS52053.2020.9336548

Study the Impact of COVID-19 on Twitter Users with respect to Social Isolation

Simranpreet Kaur*, Pallavi Kaul, Pooya Moradian Zadeh
School of Computer Science
University of Windsor
Windsor, Canada
{kaur1a1,kaulp,pooya}@uwindsor.ca

***Abstract*—The COVID-19 pandemic is a major public health issue that has affected millions of people in many different manners. This research proposes a novel approach to study and monitor this pandemic's impact on Twitter users concerning Social Isolation. Various sets of analyses is conducted to identify potential correlations between tweets related to social isolation, people's movement trends, and their emotional status. In this research, multiple datasets are used, ranging from our own collected dataset to a huge COVID-19 twitter dataset with over 140 million tweets and from the Google mobility report to the World Health Organization dataset. A couple of cloud-based services are used to analyze this data and extract patterns and insights from them. We believe the results of this research can be used as a public health indicator to anticipate the possibility of social isolation and design the health policies accordingly.**

***Index Terms*—Social Network Analysis, Sentiment Analysis, Emotion Analysis, Social Isolation, Social Media, COVID-19**

I. Introduction

The rapid spread of the coronavirus worldwide has had an immense impact on almost every aspect of human life. Due to this virus's widespread nature, many companies and businesses had to suspend their supply chains or limit their services [1]. The majority of indoor and outdoor events, including sports events (e.g., Olympics 2020), concerts, and social events, have been cancelled or postponed In many countries, social gatherings and travels have been limited, and many people and businesses have been directly or indirectly affected by the devastating effects of this pandemic [2]. Social isolation is one of these adverse effects, which can be defined as a condition where an individual has wholly or partially no contact with the society [3]. Feelings of loneliness, negative self-esteem, and fear are some of the factors observed in social isolation. It is strongly associated with depression, anxiety, and dementia and leads to a lower quality of life [4]. During the COVID-19 pandemic, many people suffered from feelings of social isolation, loneliness, depression, and social anxiety. There are some warning signs and reports which discuss the role of this ongoing pandemic on suicide rate [5].

In this research, we present a framework to study the impact of COVID-19 on Twitter users regarding social isolation. It is expected to see some correlations between the number of times that social isolation related keywords are used in tweets during different stages of the pandemic and people's movement trends, as social media can be considered a mirror of society [6]. Therefore, our research work focuses on analyzing data from various COVID-19 related datasets to explore the impact of the pandemic on these tweets.

Four different datasets are used in this research: A dataset that we have collected from Twitter related to COVID-19 from February to July 2020, the COVID-19 dataset with over 140 million unique tweets [7], which consists of all the hashtags found in daily tweets related to the COVID-19 pandemic from January 2020 until August 2020, a dataset derived from the Google Mobility reports [8] which provides aggregated and anonymous insight about the movement trends over time, based on geographical locations, in different places such as grocery and pharmacy stores, transit stations and residential. The last dataset is derived from the World Health Organization (WHO) [9] reports that contains mortality rate and the number of confirmed cases for the selected countries.

IBM cloud services (Tone analyzer and Watson Personality Insight) [10] are used to conduct prescriptive analysis based on these datasets. The results are visualized and discussed to show the patterns and behavioral changes among Twitter users regarding social isolation.

The rest of the paper is structured as follows: Related works are briefly reviewed in section II, and the architecture of our proposed framework is introduced in section III. The experimental setup and the details of its implementation are discussed in section IV. The results are reported and discussed in section V, and finally, there is a conclusion section.

II. Literature Review

In this section, we briefly review some of the recent works related to Twitter Data Analysis.

Xue et al. [11] performed sentiment analysis and latent topic modeling on tweets during the initial stages of the pandemic situation. According to their results, the sentiment of fear was most commonly observed. Sentiment analysis confirmed Positive sentiments such as trust, and negative emotions such as anger in those tweets.

Merchant et al. [12] highlighted the importance of social media in spreading awareness about the consequences and precautionary measures that need to be taken to handle the ongoing pandemic situation. They mentioned that the spread of misinformation and rumors became very common in this pandemic. They also pointed out that during this time of social isolation and lockdowns, the only mode of communication left is social media, providing insights worldwide. In [13], the authors monitored the extent of false information in Twitter data during the pandemic. They began by looking for tweets consisting of commonly used hashtags and keywords connected to COVID-19 and found that only some of them belonged to verified Twitter accounts. In contrast, the rest contained false information or a kind of information that cannot be verified. They concluded that informal or unverified accounts had greater chances of containing misinformation than professional accounts.

Guntuku et al. [14] developed a dashboard to analyze the tweets and track the mental health changes during a pandemic such as COVID-19. Their dashboard shows trends in various disciplines, such as variations in the level of anxiety, trending topics related to COVID-19, top symptoms mentioned in the tweets, and others alike. As a result, they recorded some of the skin-related disorders due to COVID-19, even before they were included in the symptoms list provided by the Centers for Disease Control. However, the main limitation was that their research was confined to just one dataset.

Chen et al. [15] developed a dataset containing tweets related to the COVID-19 pandemic. They collected their data in diverse languages using Tweepy and Twitter's streaming API and looked for prominent accounts active during the pandemic. The dataset already has over 120 million tweets, out of which 60% were in English.

Lwin et al. [16] aimed at examining the trending emotions depicted by users in their Twitter posts during the current pandemic situation of COVID-19. Four primary emotions considered in their research were fear, sadness, anger, and joy. They found that shortages linked to the supplies, including test kits, N95 masks, and various other amenities, became the main reason for spreading fear among the masses. They observed that sadness became prevalent in the social distancing period, and people felt happy only when someone reported recovery or wished for their fellow human beings' good health. In [17], the authors investigated the emotional and behavioural changes in the tweets during the pandemic. Zhang et al. [18] proposed a transformer-based deep learning model to classify the tweets and identify users showing symptoms of depression.They constructed a channel and aggregated past tweets of the individual, to monitor the depression levels. They mentioned that there was a considerable rise in depression as more and more users talked about COVID-19.

Although much research was conducted during the COVID-19 pandemic, very few focused on its impact on social isolation. Therefore, in this research, we focus on studying this issue using multiple sources of data.

III. Proposed Framework

As shown in Fig. 1, our proposed framework consists of four main stages: data collection and extraction; data cleaning and pre-processing; emotion and sentiment analysis; and prescriptive analysis and visualization of the collected datasets The first stage comprises collecting the tweets using a set of keywords related to coronavirus pandemic from Twitter using Twitter Search API. These keywords are selected based on the trending topics during the search process in February, May, and June. The motive for considering these months is since in February, COVID-19 was in its initial stage, and not many countries were affected by it. However, by the end of May, it already left its impact in most countries irrespective of their development. In contrast, by the mid of June, some states could recover or extenuate from its effect. Tweets collected possess specific attributes like tweet text, location, and timestamp stored in our database for the next stage. This dataset is used to identify the emotional and sentiment trends and patterns among the tweets.

In the second stage, the collected twitter dataset is cleaned and pre-processed by removing the URLs, emojis, duplicate and irrelevant tweets. The process is done automatically and is verified manually. Besides, we use another dataset obtained from the Zenodo data repository [7] consisting of all hashtags that have appeared in 142,360,288 unique tweets related to COVID-19 from February to July 2020. Here we call it the hashtag dataset, which is already cleaned and is ready for pre-processing. We used this dataset to measure and explore the change and evolution in usage of social isolation related keywords among the twitter usage.

As a part of pre-processing, the locations of the tweets are identified using their geographical information. Some tweets did not have any location attached in their geo-tag; thus, the profile's location is used as their current location (if found). Since the tweets posted are in diverse languages, therefore these are translated to the English language using an automation script that accesses Google Translate service. This data is used for emotion and sentiment analysis. Besides, to get some insight into people's overall status on a specific day, we create another data set. For this purpose, just the tweets' text collected each day is extracted, merged, and stored in a separate key-value dataset where the key is the date, and the value is all the tweets text collected on that day. In addition to performing date-wise analysis, tweets were segregated based on their countries, in which case, the country acts as a key, and value is the text of the tweet. Meanwhile, in this stage, the hashtags dataset is being analyzed, and all the hashtags related to social isolation will be extracted along with their frequency, dates, and locations.

In stage three, we use IBM Watson analytics services to analyze our data. IBM Watson Tone Analyzer is used to extract sentiment and emotions from our collected tweets. As a result, tweets are classified as Positive, Negative, or Neutral. For the emotion analysis, our focus is mainly on just three categories of emotions, which are Anger, Sadness, or Fear.

IBM Watson Personality Insights [10] is another cloud service that we use in this framework to provide insight into our collected tweets. This service is designed to provide insight into users' personalities based on the text given as input. Some of the categories of characters focused on by IBM are Openness, Melancholy, Agreeableness, Altruism, and various others. For our research, we considered majorly three person-

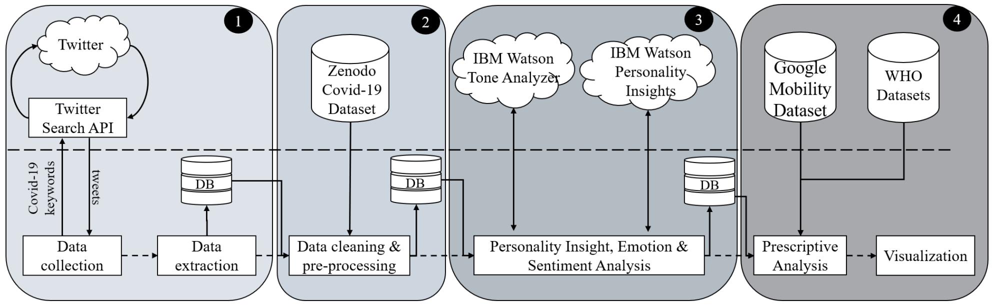

Fig. 1. The proposed architecture of our framework

alities: 'Melancholy,' 'Sympathy,' and 'Susceptible to stress' and their personality percentile. In this service, melancholy has been defined as a tendency to experience feelings of sadness or loneliness. As mentioned before, all the tweets' text separated by date or country are sent to this service to get an insight into people's overall status regarding the tendency to feel lonely and sad, sympathetic, and stress. For validation, each label received via automation is manually reviewed to confirm its correctness and modified accordingly.

The prescriptive analysis is done to deduce specific correlations with the data from different datasets in the last stage. Various types of visualizations are performed to study the users' emotional changes related to Social Isolation. These analyzes include exploring the number of changes in usage of social isolation related keywords in our hashtag dataset and its correlation with the emotional changes. We also use two more datasets, the Google Mobility dataset and WHO dataset, to have a deeper insight into the trends and patterns among Twitter users' emotional behaviors during the pandemic.

IV. Experimental Setup

We used Python programming language and Twitter Standard Search API to stream tweets based on the selected keywords and create our tweets dataset. The keywords used for contriving this dataset were "coronavirus", "coronavirusec", "coronavirusoutbreak", "COVID", "COVID19", "COVID-19". The data was collected for three time periods from February 04 to 29, May 21 to 28, and June 15 to 21. Further, the Python code was used to parse the response to extract the required elements like tweet timestamp, geo-tag, and the tweet text. As a result, 22,005 tweets were extracted and stored in a dataset for cleaning and pre-processing.

Through the data collection process, some of the twitter users' locations were found to be unreal. Therefore, to determine the tweets' exact source, they were classified based on their geographical locations, such as country. Due to the worldwide availability of Twitter [19], the tweets obtained were multi-lingual. Since the analysis of this study required the tweets to be in English language, a Java automation script, was developed using Selenium Web Driver that accesses Google Translate service to translate these tweets. For pre-processing, URL entities, duplicate and irrelevant tweets, and emoticons were removed from the original tweets list.

Following the pre-processing phase, the tweets were categorized into three categories: Positive, Negative, and Neutral. This process was implemented using TextBlob [20] in Python. It takes tweets as input and returns the text's polarity in terms of sentiment score, ranging from -1 to +1. Manual checking of the sentiments was done to validate the results.

In addition, IBM Watson Tone Analyzer was used to detect the emotion depicted by the user based on the tweet text. A list of seven kinds of emotions, including Analytical, Anger, Confident, Fear, Joy, Sadness, and Tentative, was retrieved along with their score using the python framework and was stored in the database for further processing. Each tweet doesn't need to depict just one kind of emotion; therefore, we considered the ones whose corresponding score was either higher than 0.55 or expressing just a single emotion. Similar to sentiment analysis, the results of the emotional analysis were also verified manually to ensure its correctness. Tweets showing Anger, Fear, and Sadness emotions facilitated more in comparison to other emotions in identifying the extent of social isolation. Therefore, only those emotions were held under consideration.

The second dataset used for this research is the hashtags dataset, which has been obtained from a large COVID-19 Twitter dataset in [7]. This dataset is updated regularly, but we have used their data from early February to the end of July. The dataset contains clean hashtags used in the tweets extracted from COVID-19 related keywords, along with their frequency and date of collection. To identify the hashtags related to social isolation and their rate of occurrence, we have created a bag of words of 20 keywords related to social isolation such as alone, lonely, anxiety, fear, isolation, quarantine, mental health, and depression. We searched and extracted them from the dataset.

Another dataset used in this research is Community Mobility Reports provided by Google [8]. This dataset tracked the movement and amount of activity taking place during COVID-19. The significant areas focused were mobility in parks, grocery and pharmacies, transit stations, workplaces, residential and, retail and recreation. In addition to the datasets mentioned above, another source of data used in this research is from the World Health Organization (WHO). It consists of the daily mortality count and the number of confirmed cases of COVID-19, country-wise. However, only some of the highly affected countries were considered in this research for analysis.

V. Results and Discussions

This section provides the analysis from the visualizations highlighting various correlations found in the data obtained from different data sources. Fig. 2 visualizes the changes in the usage of hashtags related to social isolation in the tweets obtained from February to July. It clearly shows a significant increase in the use of those hashtags during the pandemic, especially in March and April.

In addition, the data related to various mobility activities in different countries has been shown in Fig.3 which shows a close relationship with the usage of isolation related hashtags in the tweets. For example, mobility in Canada in terms of transit stations, retail and recreation, and workplaces began decreasing in its percentage value from March 13, onwards. Groceries and mobility in parks also began to decline slowly and steadily. As far as hashtags related to social isolation are concerned, their levels began to increase from the same period. Meanwhile, on February 23 and 24, mobility followed a little decreasing trend being above the reference line. At a similar time, there was a slight increase in the social isolation related to hashtags, as shown in Fig. 2. It has also been observed that from March 12 to 13, mobility in terms of groceries and residential increased while the rest of the mobilities began sharply decreasing. During the same time interval, the level of hashtags used suddenly increased to nearly two-folds of its previous value. March 19 onwards, the mobility data remained consistently low which is in direct relation to the number of increasing social isolation hashtags used in the tweets. In April, the data remained consistent both in terms of mobility and hashtags. May 18 shows a sharp decline in mobility in all fields except for parks and residential. At the same time, depression levels followed an increasing trend on the very same day. Mobility in terms of parks began increasing from May 15 onwards, resulting in a slight decrease in tweets related to isolation.

In France also, when the mobility in parks followed an decreasing trend, the rising trend in the hashtags used related to social isolation are observed. Meanwhile, the results shows slight increase in the number of the hashtags used, during the time that there was a sharp decline in mobility activities.

In Italy, February 23 to 25, are found to correlate with the hashtags used directly. Hashtags on February 23 are followed with a slight increase on the following day, which was then followed by a drop on February 25. A similar trend for mobility in groceries is observed in the same period. All kinds of mobility began declining from March 9 and remained to drop till March 15, after which it slightly increased on March 16 and remained constant after that. The hashtags linked to social isolation also increased abruptly on March 15 and kept growing till March 17. Mobility in groceries and recreation followed a sharp downward trend on May 10. There was a slight decrease in the levels of isolation on the same day.

For Spain, a continuous decreasing curve in mobility is observed starting from March 12, in contrast to an increase in hashtags linked to isolation. From March 22 onwards, all the mobility except for residential increased slightly, which shows the downward trend observed from the same period. After which, the bar graph for hashtags remained consistent. For example, on April 9, a sharp downward pattern in mobility is observed, which closely correlates to the increase observed in the levels of hashtags linked to social isolation. Besides, from May 14, a slightly increasing mobility trend in parks, workplaces, and transit stations is observed. During the same period, levels in isolation began decreasing. In another observation, the mobility in parks increased slightly from June 16, whereas there was a decline of approximately 200 hashtags on the same day.

March 15 onwards, the USA's mobility graph descended well below the reference line, which correlates with the slight increase first, and constant increase until March 22 hashtags related to social isolation. According to the data, on April 12, almost all kinds of mobility decreased except for residential. This is found to be in a close correlation with the increasing hashtags linked to social isolation. In addition, the mobility in parks increased from May 5, and on the same day, the hashtags slightly decreased from its previous value and followed by a negligible increase. On July 4, there was a slight decrease in workplaces' mobility, recreation, and transit stations. Simultaneously, there was a slight increase in the hashtags linked to social isolation from July 4 to 5. This is in slight contrast to Canada's levels on July 1, where there was a huge increase in the mobility of parks.

The continuous nature of mobility in India even though mostly staying below the reference line, did not solidly correlate with the social isolation related hashtags. However, one-point worth focusing on was that of mobility in workplaces in April. Whenever there was a peak observed in the workplace mobility, at the same time, there was an increase in the number of hashtags corresponding to its previous value. Also, a total contrasting trend is observed in May and June. Wherever the mobility in workplaces assumed its peak, the hashtags count decreases from their original value.

Fig.4 shows the number of tweets depicting melancholy, susceptibility to stress, and sympathy. February shows an inconsistent nature considering all three categories. Tweets with sympathy attribute assumed higher levels when compared to that of Melancholy and Susceptible to Stress. It is interesting to note that the graph for melancholy in May 1 decreased till May 23 and then began increasing till May 28, thus forming an increasing peak. In contrast, the susceptible to stress graph shows a wavy trend except May 21, where it possesses the highest level. The tweets in June, symbolizing melancholy is highest on June 16, with the lowest being on the very next day, while the susceptible to stress bar remained consistent throughout the given period.

The bar graph in Fig.5 shows the number of tweets categorized as 'Anger,' 'Fear,' and 'Sadness' along with the number of confirmed cases and mortality rate in France, Italy and USA. In France, the number of confirmed cases in May shows close relation with the number of tweets showing Anger, Fear, and Sadness emotions. It is also worth mentioning that the number

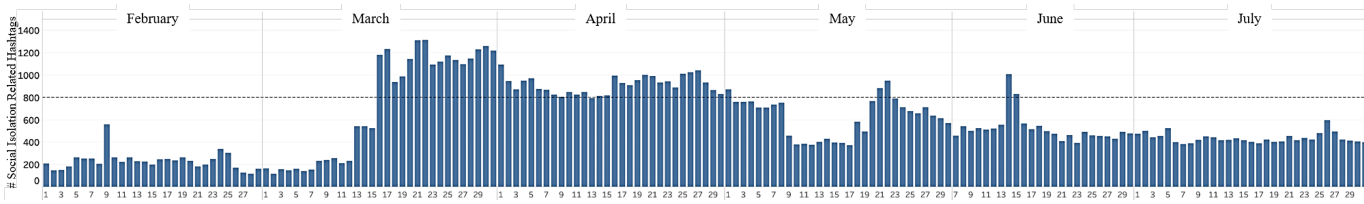

Fig. 2. Hashtags linked to Social Isolation

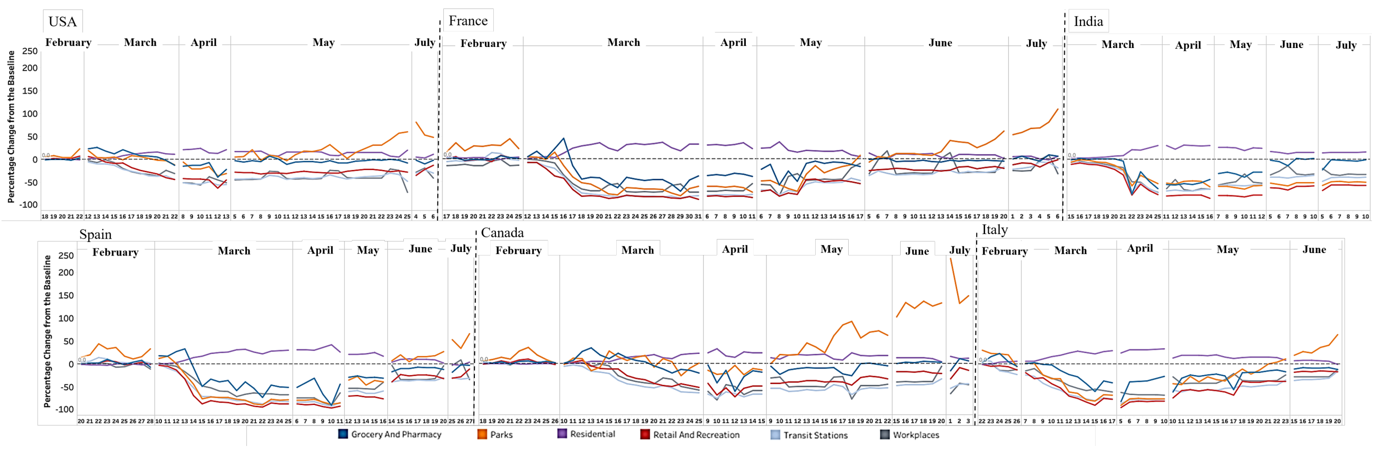

Fig. 3. Mobility in different countries

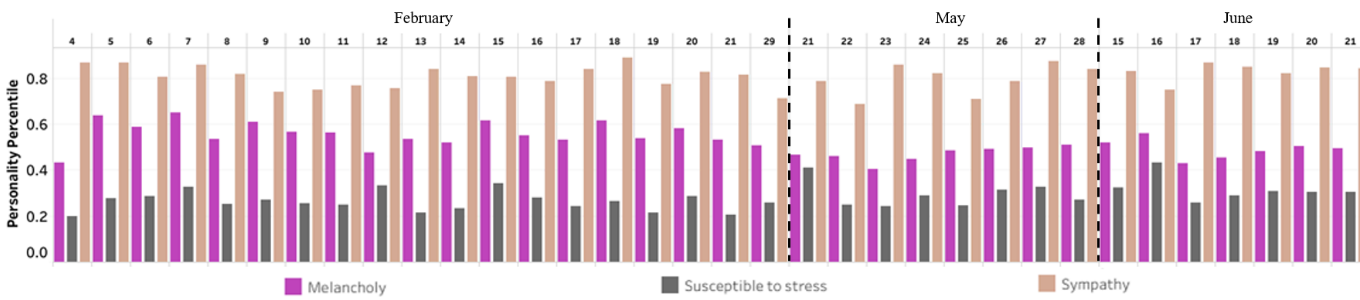

Fig. 4. Tweets distribution for Melancholy, Susceptible to Stress and Sympathy

of deaths in June depicts similar trends compared to the level of tweets. Whereas, the data for confirmed cases shows exactly contrasting trends in comparison to the tweets. For Italy, the number of confirmed cases followed a sharp incline from February 20, followed by an approximately 60% rise in the number of tweets relating to Anger, Fear, and Sadness. The line graph for deaths and confirmed cases in Italy remains consistent with the tweets in May and June.

The graphs in the second row of Fig.5 show the number of tweets categorized as 'Negative' in Canada, India, and Spain In Canada, the number of confirmed cases in June shows a similar trend towards the number of negative tweets except June 18 and 20. In May, the line graph for India's confirmed case count shows a similar trend to the number of negative tweets ranging from May 21. There is also a slight correlation when confirmed cases are concerned in that same month. In May, the number of deaths in Spain shows a uniformity with the number of Negative tweets obtained.

VI. Conclusion

This research aims to propose a framework to study and examine the impact of COVID-19 on the behavior of Twitter users in regards to social isolation. Four different datasets were used to analyze and extract the trends and patterns during the pandemic. A set of analyses was conducted using IBM Watson analytics services, ranging from sentiment and emotional analysis to personality insight and human mobility analysis. Visualization of collected data was done either in combination or independently to figure out some real-time inferences. The results show a close relationship between the level of people's physical activities and the usage of social isolation related keywords on social media. Simultaneously, we have observed a meaningful relationship between the changes in tweets' rate consisting of social isolation associated keywords and the number of death or confirmed cases during

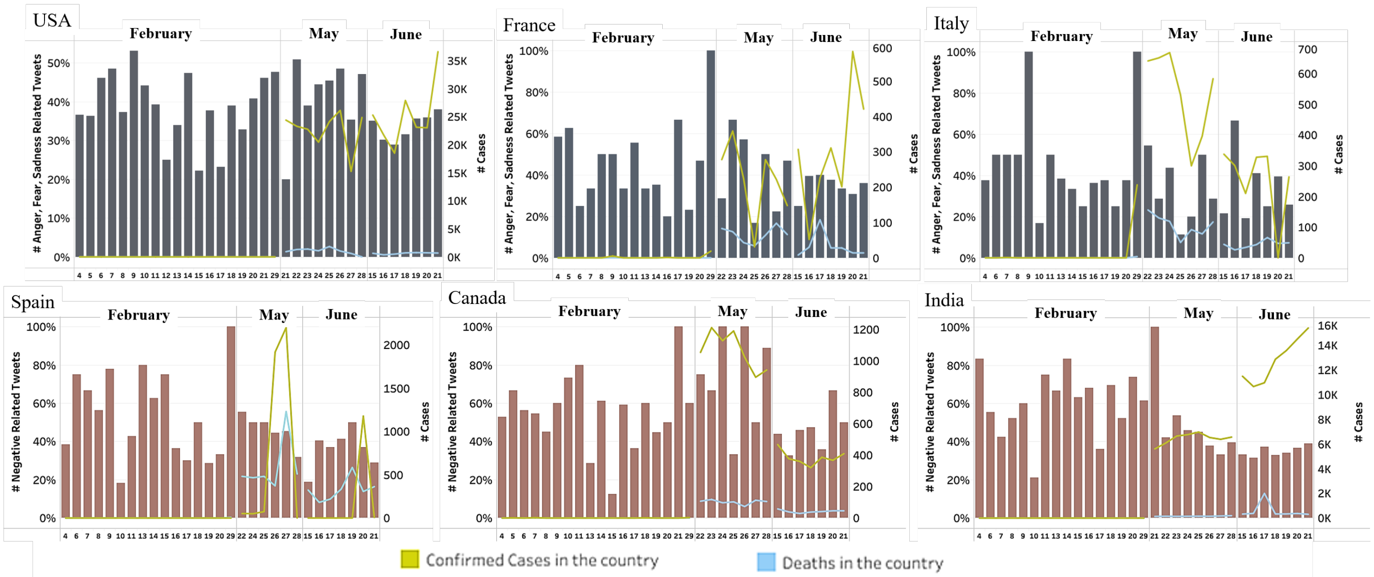

Fig. 5. Tweets distribution for Anger, Fear and Sadness Emotions (Row 1), and Negative Sentiment (Row2) for each interval with respect to WHO data

the pandemic. In the future, we will expand this research to include more data sources from different social network platforms.

REFERENCES

[1] A. W. Bartik, M. Bertrand, Z. Cullen, E. L. Glaeser, M. Luca, and C. Stanton, “The impact of covid-19 on small business outcomes and expectations,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 30, pp. 17 656–17 666, 2020.

[2] T. Ritter and C. L. Pedersen, “Analyzing the impact of the coronavirus crisis on business models,” *Industrial Marketing Management*, 2020.

[3] D. Khullar, “How social isolation is killing us,” *New York Times*, vol. 22, 2016.

[4] A. H. Jansson, N. M. Savikko, and K. H. Pitkälä, “Training professionals to implement a group model for alleviating loneliness among older people–10-year follow-up study,” *Educational Gerontology*, vol. 44, no. 2-3, pp. 119–127, 2018.

[5] L. Sher, “The impact of the covid-19 pandemic on suicide rates,” *QJM: An International Journal of Medicine*, 2020.

[6] E. D. A. Manzoor, *Social Media as Mirror of Society*, 01 2017.

[7] J. M. Banda, R. Tekumalla, G. Wang, J. Yu, T. Liu, Y. Ding, K. Artemova, E. Tutubalina, and G. Chowell, “A large-scale COVID-19 Twitter chatter dataset for open scientific research - an international collaboration,” Aug. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3977558

[8] “COVID-19 Community Mobility Report,” 2020. [Online]. Available: https://www.google.com/covid19/mobility?hl=en

[9] WHO, “Coronavirus disease (covid-19) situation reports,” 2020. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/

[10] “IBM Watson.” [Online]. Available: https://www.ibm.com/watson

[11] J. Xue, J. Chen, C. Chen, C. Zheng, S. Li, and T. Zhu, “Public discourse and sentiment during the covid 19 pandemic: Using latent dirichlet allocation for topic modeling on twitter,” *PloS one*, vol. 15, no. 9, p. e0239441, 2020.

[12] R. M. Merchant and N. Lurie, “Social media and emergency preparedness in response to novel coronavirus,” *Jama*, 2020.

[13] R. Kouzy, J. Abi Jaoude, A. Kraitem, M. B. El Alam, B. Karam, E. Adib, J. Zarka, C. Traboulsi, E. W. Akl, and K. Baddour, “Coronavirus goes viral: quantifying the covid-19 misinformation epidemic on twitter,” *Cureus*, vol. 12, no. 3, 2020.

[14] S. C. Guntuku, G. Sherman, D. C. Stokes, A. K. Agarwal, E. Seltzer, R. M. Merchant, and L. H. Ungar, “Tracking mental health and symptom mentions on twitter during covid-19,” *Journal of general internal medicine*, pp. 1–3, 2020.

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

[16] M. O. Lwin, J. Lu, A. Sheldenkar, P. J. Schulz, W. Shin, R. Gupta, and Y. Yang, “Global sentiments surrounding the covid-19 pandemic on twitter: analysis of twitter trends,” *JMIR public health and surveillance*, vol. 6, no. 2, p. e19447, 2020.

[17] S. Kaur, P. Kaul, and P. M. Zadeh, “Monitoring the dynamics of emotions during covid-19 using twitter data,” *Procedia Computer Science*, vol. 177, pp. 423 – 430, 2020.

[18] Y. Zhang, H. Lyu, Y. Liu, X. Zhang, Y. Wang, and J. Luo, “Monitoring depression trend on twitter during the covid-19 pandemic,” *arXiv preprint arXiv:2007.00228*, 2020.

[19] L. Sinnenberg, A. M. Buttenheim, K. Padrez, C. Mancheno, L. Ungar, and R. M. Merchant, “Twitter as a tool for health research: a systematic review,” *American journal of public health*, vol. 107, no. 1, pp. e1–e8, 2017.

[20] “Simplified Text Processing.” [Online]. Available: https://textblob.readthedocs.io/en/dev/