International Journal of Computer Engineering in Research Trends
Multidisciplinary, Open Access, Peer-Reviewed and fully refereed
Research Paper | Volume-9, Issue-6 ,2022 Regular Edition | E-ISSN: 2349-7084

A Real-time Visualization of Global Sentiment Analysis on Declaration of Pandemic

Lavanya, A. [1], Waqas Ali [2], Dr. Jaime Lloret [3], Vidya Sagar, S. D. [4], Chivukula Bharadwaj [5]

[1,3] *Universidad Politécnica De Valencia, Valencia, Spain*
[2] *School of Information Engineering, Yangzhou University, Yangzhou 225009, China*
[4] *Kuvempu University, Shimoga. Karnataka, India*

**Corresponding Author: phani.lav@gmail.com*

Available online at: http://www.ijcert.org

Received: 06/06/2022, Revised: 20/06/2022, Accepted: 22/06/2022, Published: 24/06/2022

Abstract: - The paper's objective is to carry out a real-time visualization of pandemic sentiment at the very first instance. The paper shows multilevel visualization of sentiment analysis conducted on the covid19 dataset acquired from Twitter. The visualization tools used for real-time data are Google data studio, Python matplotlib, Carto, and Tableau. On Mar 11, 2020, Covid19 was declared a global pandemic, and stage wise lockdown protocols were implemented. The covid19 virus has spread worldwide and consumed millions of people. The impact of the virus is affected not only on the physical body but also on mental health and results in increased distress, depression, anxiety, fear, and panic simultaneously. The data was downloaded using twitter's official API on Mar 11, 2020. Vader sentiment analysis is performed on 3,27,717 tweets downloaded from 200 Megacities globally. The study achieved 50.95% negative and 58.72% positive sentiment and neutral values ranging between 0 to 1 polarity.

Keywords: Covid19; Geo-Spatial; Temporal Data; Pandemic; Sentiment; Visualization;

1. Introduction

The COVID-19 pandemic endangers the general population's physical well-being and mental health, leading to heightened and sustained feelings of uncertainty, alienation and grief, and disturbance of social and economic structures [1][2]. Emerging data on public mental sentiments suggest the signs and traces of posttraumatic stress disorder (PTSD) and depression are widespread in the general population at the early stage of this pandemic[3][4]. Another danger to the mental well-being of the country and its people is the introduction of national quarantine steps to curb the dissemination of COVID-19. While quarantine can be a successful measure in public health [5] it has a substantial physical, social, and psychological impact. The paper focuses on acquiring and visualizing the real-time sentiment analysis from Twitter streaming data from Mar 11, 2020, when covid19 was declared a global pandemic. The data collection has about 200 geo-location, including China, the source of the virus outbreak, to understand the sentiment and real-time better when there is no previous evidence or reference data set available for Artificial Intelligence and machine learning approach. The paper evaluates the emotions using VADAR sentiment analysis, which is best suitable for real-time data analysis when no previous reference is available. The real-time data visualization is done using data visualization tools like Google Data Studio, Tableau, and Carto for spatial and temporal analysis of the data, and the matplotlib for projecting the compliment cumulative distribution range sentiments from positive-negative, neutral, and compound.

The paper aims to focus on ready-to-use visualization tools for addressing real-world sentiment when historical training data and learning modules are unavailable.

The paper is divided as follows: Section 1 introduces the ground situation of the pandemic outbreak; section II discusses an extensive background and literature about the spread of the virus towards pandemic declaration and the need for sentiment analysis and visualization tools; section III discusses the data downloaded from Twitter, the keywords used for searching, and the stats of acquired raw data. Section IV covers the results and illustrates the outcome from selective visualization tools at a multi-dimensions level exposing the geo-spatial and temporal data. The last section concludes the paper with a marginal positive sentiment about the pandemic.

2. Background

In 1918-20 about 100 years ago, a global epidemic broke out called the Spanish Flu. The pandemic lasted almost two years and consumed about 1.5% of the Global Population at that time [6]. After a century, another pandemic outbreak is called Novel Corona Virus. The population who saw the Spanish Flu has very few records; back then, the technical reach was bleak to none; thus, impacted mental health and preventive thoughts are likely to be unregistered. The world health organization has setup up a few protocols for stages of severity to understand the phases of the outbreak, and based on that epidemic or pandemic is declared. The phases as mentioned in the figure below:

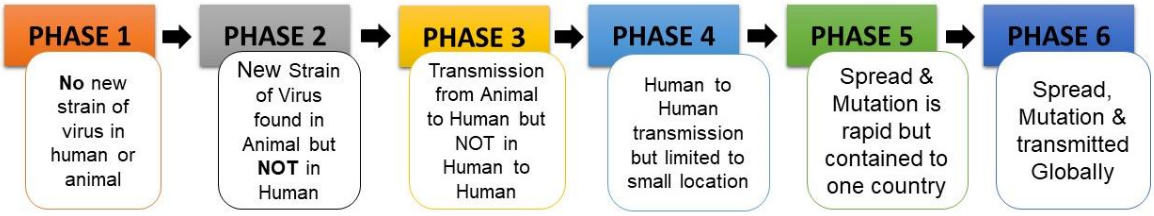

Figure 1: World Health Organization Epidemic Phases

The 21st Century Pandemic is new, and none of the alive population has ever experienced the feeling of a pandemic. What are the things to be anticipated? On Mar 11, 2020 world health organization declared the outbreak of the novel coronavirus as a global pandemic after spreading across 114 countries and more than 4000 people dead[7][8]. On Jan 9th, 2020, WHO announced an epidemic in China spreading a mysterious virus. Jan 20, 2020, CDC suggested screening at three major airports in the USA. On Jan 21, 2020, the first case of covid19 was identified in the USA on the same day a Chinese scientist confirmed human-to-human transmission. Jan 23, 2020, Wuhan and some parts of China were put under quarantine to contain the spread of the virus. On Jan 31, 2020, WHO declared a global emergency, stage 4, per the WHO guidelines.

With the crisis said, international air restrictions started on Feb 02, on Feb 03 USA declared a health emergency, more than 200 deaths were officially noted, and 9800 cases were tested positive. By Feb 10, the death toll surpassed the SARS outbreak from 2003, where 908 deaths were reported, and in the case of Covid-19, already 774 deaths were reported in one month, with rising numbers every day. On Feb 25, CDC anticipated a pandemic. On Mar 06, a cruise ship passenger, more than 21, tested positive, and finally, on Mar 11, WHO, along with CDC, declared COVID-19 as a pandemic officially. With so many deaths and positive tests, by the end of 2020, a total of 83,832,334 people were infected by covid19, and more than 1,824,590 people had died. The top ten countries affected strongly by the covid19 outbreak are the USA, India, Brazil, Russia, UK, France, Spain, Italy, Turkey, and Germany. Each country has seen more than 2 million deaths yearly[9]. The pandemic raises the need for mental health facilities. Bereavement, alienation, income depletion, and anxiety cause mental well-being or worsen chronic conditions. Many people have increased alcohol and other drug consumption, insomnia, and anxiety. Meanwhile, COVID-19 can induce neurological and psychiatric problems, including delirium, hysteria, and stroke [10] . People with pre-existing psychiatric, neurological, or drug use conditions are often more vulnerable to infection with SARS-CoV-2 — they may be at greater risk of severe consequences and death [11] .

The growing interest in sentiment analysis, especially in Twitter data, is a leading waste area of research[12] . The domain has no label details in the general sense classification method for using the target. There are two structured and unstructured data analysis methods. One needs previous data for training, and unstructured sentiment analysis does not require any training data and has one of the best approaches for real-time sentiment analysis [13]. Moreover, this system measures each word frequency in a tweet [14] . The current dataset predicted the polarity of emotions reflected in opinions, acceptance, and distress about the pandemic. Traditional classification algorithms can train sentiment classifiers from manually labeled text data. Still, the labeling work needs previous domain knowledge and time-consuming [15]. Several studies show that the output is inferior if a trained classifier is extended explicitly to other realms. The work shows the accuracy of numerous algorithms for different tweet numbers, including Naive Bayes, Multi-nominal NB, Linear SVC, Bernoulli NB classifier, Logistic Regression, and SGD classifier. Results found in unstructured sentiment analysis are more effective than other methods, particularly in the case of psychological sentiment analysis [16][17][18].

3. Data Information

Twitter is a common micro-blogging site where users generate short messages called tweets that convey various topics. Over the past decade, Twitter has become a popular social networking application, and thus there has been much curiosity about how to efficiently gather data from the site[19][20]. The participation levels of users on pubic-related topics, catastrophes, and natural calamities are adequate to carry forward research based on the tweets from our previous research[21].

Twitter's two methods of downloading data:

1. Use the REST API for historical info, contacts, or custom user timeline.
2. Use Streaming API to download real-time data.

Streaming API was used to download the tweets using the keywords #cvoid19, #corona, and #lockdown. The data collection is done for the first 24 hours of the pandemic declaration by the World health organization. The tweets were collected from about 200 megacities worldwide, and 17.15% of tweets were location enabled. Table 1 & Graph 1 display the total number of tweets acquired. Post sanitization, the tweet started on half 11th & ends by Mar 12, 2020, covering 24hrs of time in all the time zones. Steps involved in data acquisition:

- Step 1; the First phase is registering for a Twitter Developer account and getting the Twitter API token and keys. The token and critical are the most important for using Tweepy Package for Streaming API.
- Step 2: Import Tweepy in Python, set up authentication, and stream listener with API keys. Here the Twitter authentication details are used when registered and submitted.
- Step 3: Develop class StreamListener. In the same file, a new class named StreamListener inherits the StreamListener class of tweepy and overrides the status and error methods to customize the configuration.
- Step 4: Initialize filter stream. Finally, the stream is launched by defining search keywords "Covid19", "Coronavirus", and "Lockdown".
- Step 5: the data saving and processing stage where the needed information from Twitter data, including metadata, is stored in CSV for future usage.
- Step 6: As the streaming API searches for keywords and hashtags, there is a high chance of duplicate data download. Data preprocessing is done by removing duplication and sanitizing the data as needed.
- Step 7: The data is supposed to have geo locations enabled; thus, the data with "sweet place" present only those tweets are considered. Therefore, only about 17% of data is used for future analysis. Not everyone has geo location-enabled, and not all tweets show geo-locations.

Table 1: below are the stats of the total number of tweets acquired and the number of tweets remaining post sanitations.

Table 1: Data Information

Total Tweets	Sanitized Tweets	Location	Time
19,10,191	3,27,717	200	24 hrs.

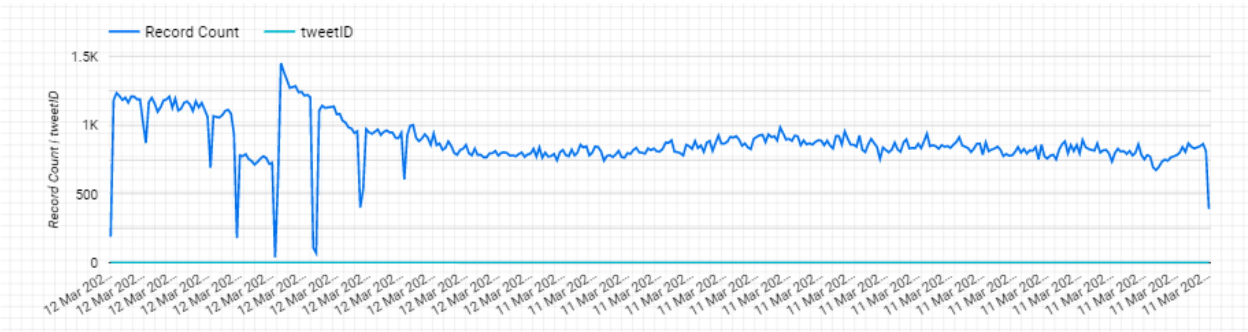

Figure 2: Time Series - Tweet Counts

 DOI: https://doi.org/10.22362/ijcert/2022/v9/i06/v9i0602

Figure 2 shows the growth of tweets from the time of the announcement of a pandemic. For some time, a consistent number of tweets are seen and later show fluctuations, sharp falls, and rises. The times series visualization is plotted in Google Data studio.

3.1 Data Visualization Tools

Since the rise in Twitter and big data, data visualization has picked up the pace in 2014. Data visualization is nothing but showing the data in the form of graphs, charts, and plots. The term big data is self-explanatory; thus, to understand the nature of the data, it is impossible only to read and understand it as it is. There is a need for a good representation of the data to understand the data better. There have been several codding and non-coding tools for the data visualization process. In real-time data analysis, where the nature of data is unknown except for the data format, programming ready-use visualization tools come in handy[22][23]. Few such devices used in this real-time data visualization are explained as:

A. Google data studio: Data Studio makes the critical data available and usable. Data Studio performs data authentication, access permissions, and structure for use in measurements and data visualization independent of the data source. Several sources are to import data from channels such as Analytics, Google Advertising, Google BigQuery, Campaign Manager 360, MySQL, and more[24] [25].

B. Tableau: Tableau is a popular and fastest-growing platform used in business intelligence. It helps simplify raw data in an understandable format. Tableau helps generate evidence that experts appreciate at any stage of an enterprise. It allows non-technical users to build personalized dashboards. With the Tableau platform, data analysis is rapid, and the visualizations generated are in dashboards and worksheets[26][27].

C. CARTO is a tool that transforms spatial data into successful distribution routes and improves spatial-temporal analysis with geographical time-series animation to understand the spread over time [28][29] .

D. Python Matplotlib: Matplotlib is a robust repository for Python's static, animated and immersive visualization. This paper uses matplotlib to construct the compliment cumulative distribution function of the sentiment to have a comparative study on the distribution of emotions [30][31].

Figure 3 illustrates the geo-spatial spread of the source of tweets. The visualization is done in Tableau. The geographical map feature makes it easy to apply filters to separate the colors or range of color grading from light to dark to display less concentration. The map also shows the size of the marker in variation based on the number of tweets acquired from the location. The map illustrates high concentration in far eastern countries, central and south of Europe, and the east coast of the USA as a reaction to the pandemic declaration.

Figure 3: Geographical Distribution of Tweets across 200 Mega Cities

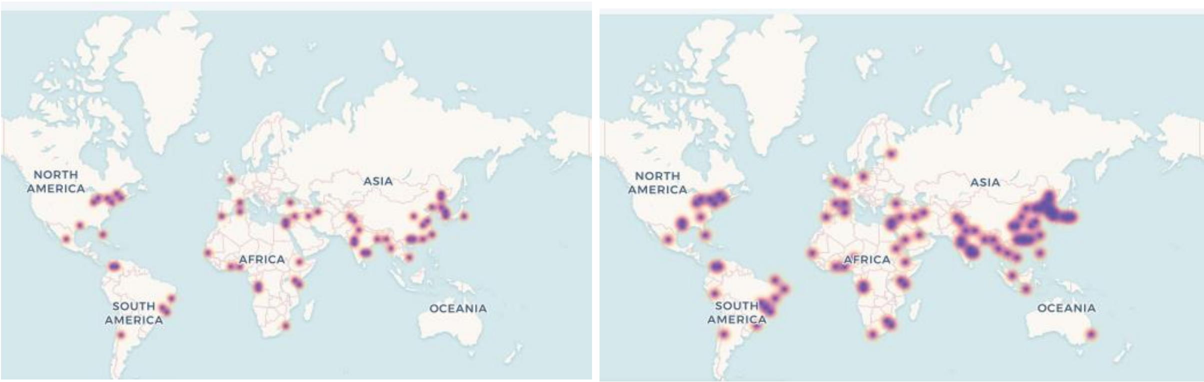

Figure 4: Increase of tweets: Spatial-Temporal Visualization

Figure 4 the map visualization is the display of geo-spatial over a temporal distribution; it was constructed using CARTO for understanding the time series and geographical spread of tweet contribution. The map is a two-part illustration, with Map A showing a few hours later the outbreak announcement and Map B illustrating the end of 24hrs news announcements of a pandemic. The heat maps are well understood as time passes; more contributors have posted their thoughts, views, and opinions on the WHO's declaration on the pandemic. Map B shows a high concentration around South East Asia, wide Spread Europe, Africa and the East Coast of the USA, and some central locations in South America.

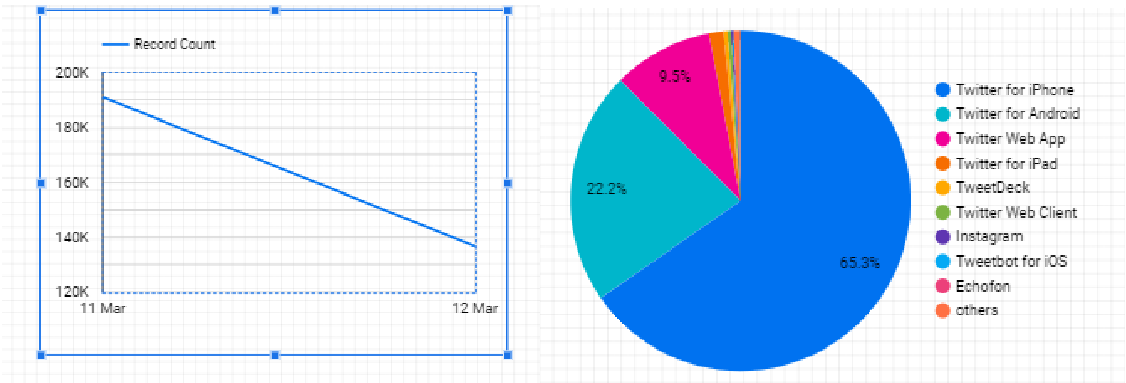

Figure 5: (A & B) Tweet Counts: Date & Source of Tweets

The beauty of data visualization is that so much information is hidden inside the primary data, and identifying and displaying it is a task that becomes easy with us of tools. The above figure 5 shows the number of tweets acquired on the 11th and 12th. The figure is quite explanatory. The highest contribution was Mar 11, and by the end of 24hrs on the 12th, the number of contributors had reduced. The other part of Image 1: is the type of devices used to post. Twitter is well-known for having actual human beings, and automated bots are contributors. For branding and promotional purposes, automated bots are created that keep posting at a particular preprogrammed interval. But in the case of a real-time situation where automated bots need pre-planning and programming to set up based on the keywords trending. It is understood that in case of unexpected and real-time scenarios, the tweet contributors, about 65% are users of iPhone, 22% are android, 9% are direct Twitter web users, and there are Twitter Ipad users. Then there are Twitterdeck users for multiple timeline users. "TweetSource" gives a picture of the source of the tweets for a better understanding of actual humans and automated bots. The visualization was conducted on google data studio.

4. Data Computation

Sentiment analysis is a text analysis that senses polarity (e.g., a positive or negative opinion) in the text, an entire document of conversation, a single sentence, or a paragraph. Sentiment analysis assesses a speaker writer's mood and thoughts, evaluations, behaviors, and emotions dependent on computational subjectivity treatment in a text. VADER (Valence Aware Dictionary for Sentiment Reasoning) is a paradigm used in text sentiment analysis that is adaptive to emotional polarity (positive/negative) and intensity (strength). It is included in the NLTK package and can be applied to unlabeled text files. VADER's sentimental research focuses on a dictionary translating lexical traits to subjective intensities known as sentiment ratings. A text's feeling score can be obtained by summarizing each word's strength in the text.

The overall Sentiment Analysis

IF:
SC = 1: (SC= Score Count: Range 0 to 1)

The overall polarity of the tweets is computed for Positive

IF:
PV > 0.5(Positive Value)
THEN:
TP = +2 (Tweet Polarity)
ELSE:
PV < 0.5
THEN:
TP = +1

IF:
SC = -1

The overall Polarity of Tweets is computed for Negative

IF:
NV > 0.5 (NV= Negative Value)
THEN:
TP = -2 (Tweet Polarity)
ELSE:
NV < 0.5
THEN:
TP = -1

IF
SC = 0
THEN:
TP = 0

The value of polarity provides the tweet's overall polarity of sentiment. The polarity value is set between -2 (highly negative) and +2 (highly positive). Depending on the positive value, positive tweets are classified as highly positive or positive; negative tweets are classified as highly negative or negative, depending on the negative value; negative tweets are classified as neutral in other cases.

5. Discussion

This segment discusses the findings of a Twitter sentiment analysis using VADER sentiment analysis instruments. As the VADER Sentiment Analyzer received, the below section displays the sentiment score of each tweet as positive, negative, neutral, or compound.

Table 2: Data Sentiments: VADAR Sentiments

SELECTIVE TWEETS	NEGATIVE	POSITIVE	NEUTRAL	COMPOUND
"You took a cheap flight to Italy and died from coronavirus?"	0.13	0.076	0.794	-0.3612
"Flu Symptoms, Spring Allergies, Corona Virus Symptoms all meeting each other"	0.14	0	0.86	-0.3818
"My wife and I get coronavirus. We go to Disneyland and ride California screaming. The park finds out and quarantines us."	0	0	1	0
"Mumbai is reporting its 1st two cases of coronavirus. Be safe, everyone Who declares it a global pandemic. Be safe, Mumbaikar."	0	0.195	0.805	0.7003
"Coronavirus leaving the world crippled sucks like, don't get me wrong, everyone having to stay home from work and school."	0.079	0.16	0.761	0.3724
"Corona will not touch you. Say amen!"	0	0	1	0

The table above displays the classification of tweets with the polarity of Positive, Negative, Neutral, and Compound. Even though all the tweets are related to covid19, every tweet has its own emotion and flavor to extract emotion.

"You took a cheap flight to Italy and died from coronavirus?" A rhetorical type of question with a score less than 0.5; thus, it is a negative inclination.

"Flu Symptoms, Spring Allergies, Corona Virus Symptoms all meeting each other" there is nothing positive in the symptoms discussed in the tweet; thus, it is part negative and part neutral.

The following tweets do not discuss any emotion, and it is one of the first tweets like that; practically on11th March, there have been hardly any positive cases in the USA, and thus it falls in the Neutral category. Sentiment analysis analyses the echo or sound of the keywords, and there needs to be an emotion in the given message. English is a fluid language and can be presented in several ways; thus, it is a limitation it hit with unstructured real-time sentiment analysis. Despite such a setback, Vader's sentiment is tested and proved to be 85% accurate. The data size puts this into the big data category, and post sanitization, the analysis process is conducted, achieving the following results.

Table 3: Sentiment Matrix

Sentiment		Null Values	Greater than 0.5	Less than 0.5
POSITIVE		0	Highly Positive (+2)	Positive (+1)
Tweet Count	3,27,717	1,92,459	376	1,34,882
Percentage		58.72	0.12	41.16
NEGATIVE		0	Highly Negative (-2)	Negative (-1)
Tweet Count	3,27,717	1,60,739	1,129	1,65,849
Percentage		49.04	0.35	50.61
TOTAL		Neutral	Positive	Negative
		7.76	41.28	50.96

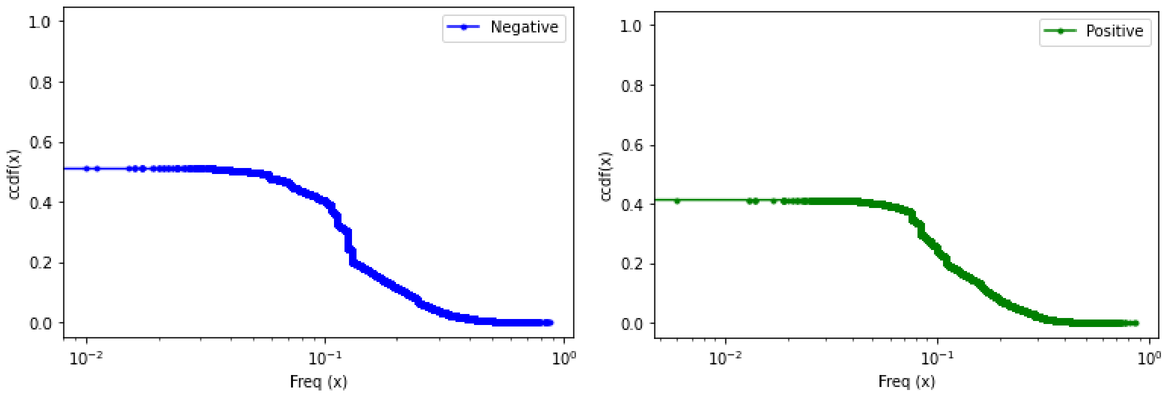

Figure 6: (A & B) CCDF of Negative & Positive Sentiment)

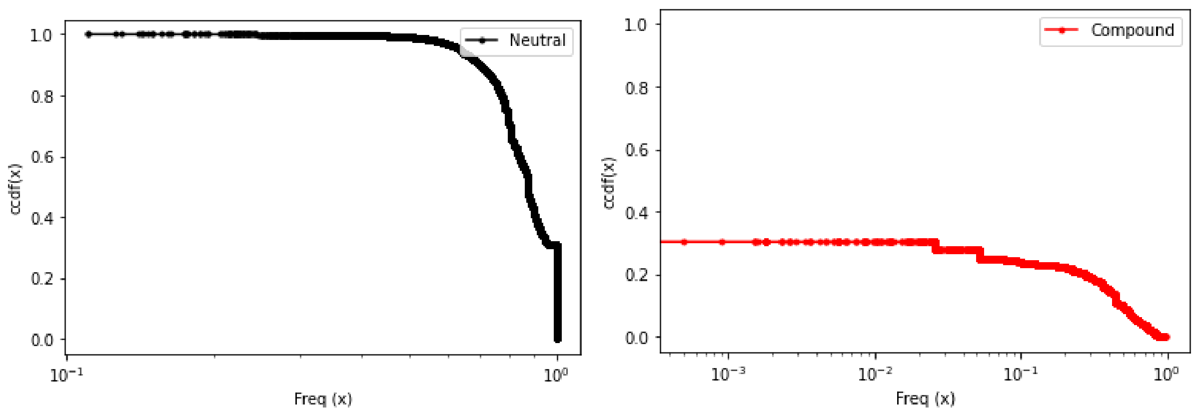

Figure 7: Graph 3: CCDF of Neutral & Compound Polarities

Table 4: Normalized Compound Distribution

Polarity	Distribution Percentage
Positive	30.29%
Negative	38.62%
Neutral	31.09%

Table 3 & 4, Figure 6 & 7 discusses the cumulative polarity and distribution of all the tweets from a total of about 58% tweets fall into the category of null for positive, 41.16 % were positive but not highly positive, and only 0.12 % were highly positive tweets. About 1.29% are highly damaging for the negative polarity, and 50.96% are negative sounded tweets. Overall sentiment distribution was 50.96% negative sentiment, and 41.28 % of the sentiment was positive, leaving 7.76% in the neutral category. The compound scores are the normalized positive, negative, and neutral scores. There is nothing good to sound about deaths and illness worldwide, but the positivity is towards implementing lockdown protocol for the public's safety.

Figure 8: (A & B) Spread of Negative Sentiment: Spatial-Temporal Visualization

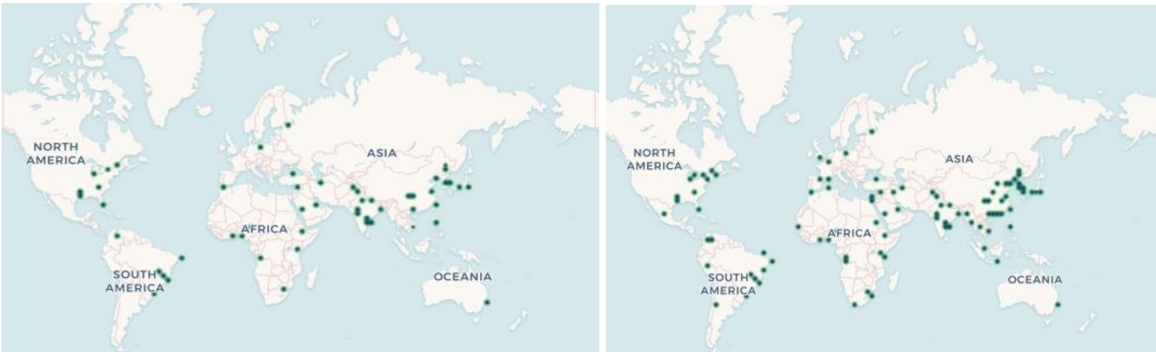

Figure 9: (A & B) Spread of Positive Sentiment: Spatial-Temporal Visualization

Figure 10: (A & B) Micro Level Focused Difference of Polarity

Figures 8 & 9 display the spread of negative and positive sentiment across the geographical location based on the locations from where the data was acquired. It shows two map plots, each one in 12hrs of news outbreak and the other one by the end of 24hrs of pandemic news and news about the lockdown protocols, quarantine protocols, etc. in Maps A of both figures 8 & 9, the first 12 hrs where significant data contribution is seen in that major negatively is seen southeast Asian countries and eastern coast of America. By the end of 24 hrs, the negativity has heavily concentrated in the southeast Asian countries, mainly in India, China, and its neighboring countries. The positivity is also more concentrated in South Asian countries when compared to other countries. Table 4 explains the normalized cumulative polarity distribution, and the difference between the negative and positive polarity is 8-9%; thus, the visualization does not show a significant difference. A focused micro-level visualization is shown in Figure 5 (A &B). The cumulative compound polarity difference of locational sentiment can be observed with a minor difference in the map plot.

6. Conclusion

Several methods for assessing emotions exist. Some are commercial platforms like Meaning Cloud, Get Sentiment, or Watson Natural Language Understanding. Sentiment analysis libraries are also available in standard machine learning applications such as Rapid Miner or Weka, expanding to a widespread sentiment analysis lexicon library. Vader sentiment analysis for real-time data and set of tools for real-time distress assessing visualization was achieved adequately by the set of visualization tools like google data studio, Tableau, Carto, and finally, Python-based plotting packages like matplotlib. Understanding distress or any other sentiment in real-time allows the authorities and concerned person to take necessary public mental health well-being steps.

References

[1] A Timeline of COVID-19 Developments in 2020. (2021, Jan 02). (ajmc.com) Retrieved Feb 24, 2021, from https://www.ajmc.com/view/a-timeline-of-covid19-developments-in-2020

[2] Apricio, M., & Costa, C. J. (2015). Data visualization. Communication design quarterly review, 3(1), 7-11.

[3] Barrett, P., Hunter, J., & Miller, J. T. (2005). Matplotlib--A Portable Python Plotting Package. Astronomical data analysis software and systems XIV, 347, 91.

[4] Chppell, B. (2020, Maarch 11). Coronavirus: COVID-19 Is Now Officially A Pandemic, WHO Says. (npr.org) Retrieved Feb 24, 2021, from https://www.npr.org/sections/goatsandsoda/2020/03/11/814474930/coronavirus-covid-19-is-now-officially-a-pandemic-who-says

[5] Day, T., Park, A., Madras, N., Gumel, A., & Wu, J. (2006). When is quarantine a useful control strategy for emerging infectious diseases? American journal of epidemiology, 163(5), 479-485.

[6] Elbagir, S., & Yang, J. (2019). Twitter sentiment analysis using natural language toolkit and VADER sentiment. Proceedings of the International MultiConference of Engineers and Computer Scientists, 122, 16.

[7] Elflein, J. (2021, feb 24). Covid-119 cases worldwide as of feb 24, 2021. (statista.com) Retrieved feb 24, 2021, from https://www.statista.com/statistics/1043366/novel-coronavirus-2019ncov-cases-worldwide-by-country/

[8] Fernando, D. (2018). Visualisasi Data Menggunakan Google Data Studio. Prosiding Seminar Nasional Rekayasa Teknologi Informasi| SNARTISI.

[9] Friendly, M. (2008). A brief history of data visualization. Handbook of data visualization, 15-56.

[10] Ghebreysus, T. A. (2020). Addressing mental health needs: an integral part of COVID‐ 19 response. World Psychiatry, 19(2), 129.

[11] Hutto, C., & Gibert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the International AAAI Conference on Web and Social Media, 8(1).

[12] Hyland, P., Shevlin, M., McBride, O., Murphy, J., & Karatzias, T. (2020). Anxiety and depression in the Republic of Ireland during the COVID‐ 19 pandemic. Acta Psychiatrica Scandinavica, 142(3), 249-256.

[13] Islam, M. A., Barna, S. D., Raihan, H., Khan, M. N., & Hossain, M. T. (2020). Depression and anxiety among university students during the COVID-19 pandemic. Bangladesh: A web-based cross-sectional survey PloS one, 15(8), e0238162.

[14] Kim, A. E., Hansen, H. M., Murphy, J., Richard, A. K., Duke, J., & Allen, J. A. (2013). Methodological considerations in analyzing Twitter data. Journal of the National Cancer Institute Monographs, 2013(47), 140-146.

[15] Lavanya, A., Panwar, D., Jaime, L., & et. al. (2022). Event-Based Multi-Model Classification to Assess the User Participation Levels on Twitter. In N. Thakur, & B. D. Parameshachari (Eds.), Human-Computer Interaction and Beyond: Advances Towards Smart and Interconnected Environments-II (pp. 76-120 (45)). Bentham Books.

[16] Mane, S. B., Sawant, Y., Kazi, S., & Shinde, V. (2014). Real time sentiment analysis of twitter data using hadoop. International Journal of Computer Science and Information Technologies, 5(3), 3098-3100.

[17] McGinty, E. E., Presskreischer, R., Han, H., & Barry, C. L. (2020). Psychological distress and

loneliness reported by US adults in 2018 and April 2020. Jama, 324(1), 93-94.
[18] Mucchetti, M. (2020). Google Data Studio. BigQuery for Data Warehousing, 401-416.
[19] Murphy, S. A. (2013). Data visualization and rapid analytics: Applying tableau desktop to support library decision-making. Journal of Web Librarianship, 7(4), 465-476.
[20] Ozdin, S., & Bayrak, O. S. (2020). Levels and predictors of anxiety, depression and health anxiety during COVID-19 pandemic. Turkish society: The importance of gender International Journal of Social Psychiatry, 66(5), 504-511.
[21] Reece, A. G., Reagan, A. J., Lix, K. L., & Dodds, P. S. (2017). Forecasting the onset and course of mental illness with Twitter data. Scientific reports, 7(1), 1-11.
[22] Roser, M. (2020, March 04). The Spanish flu (1918-20): The global impact of the largest influenza pandemic in history. Retrieved from ourworldindata.org: https://ourworldindata.org/spanish-flu-largest-influenza-pandemic-in-history
[23] Sedoc, J., Buechel, S., Nachmany, Y., Buffone, A., & Ungar, L. (2019). Learning word ratings for empathy and distress from document-level user responses. arXiv preprint arXiv:1912.01079.
[24] Talbot, J., Charron, V., & Konkle, A. (2021). Feeling the Void: Lack of Support for Isolation and Sleep Difficulties in Pregnant Women during the COVID-19 Pandemic Revealed by Twitter Data Analysis. International Journal of Environmental Research and Public Health, 18(2), 393.
[25] Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. J. Lang. Soc. Psychol, 29(1), 24-54.
[26] Telea, A. C. (2014). Data visualization: principles and practice. CRC Press.
[27] VanderPlas. (2016). Python data science handbook: Essential tools for working with data. O'Reilly Media, Inc.
[28] Wagh, B., Shinde, J. V., & Kale, P. A. (2018). A Twitter sentiment analysis using NLTK and machine learning techniques. International Journal of Emerging Research in Management and Technology, 6(12), 37-44.
[29] Wang, C., Pan, R., Wan, X., & Tan, Y. (2020). A longitudinal study on the mental health of general population during the COVID-19 epidemic in China. Brain, behavior, and immunity, 87, 40-48.
[30] Wesley, R., Eldridge, M., & Terlecki, P. T. (2001). An analytic data engine for visualization in tableau. Proceedings of the 2011 ACM SIGMOD International Conference on Management of data, 1185-1194.
[31] Zastrow, M. (2015). Data visualization: Science on the map. Nature News, 519(7541), 119.