Analyzing the Escalation and Impact of COVID19 Misinformation on social media

Vivek Yadav
Scholar, MSc Data Science
Central University of Haryana

Dr. Suraj Arya
Assistant Professor, CS&IT
Central University of Haryana

Abstract: The Corona epidemic has come before the society as a terrible problem of global level. As we all have faced this terrible situation and almost all the countries around the world have lost their citizens which includes small children, women and youth. However, medicines were developed after finding out the root causes of this problem. Due to which it could be prevented to a great extent. During the Corona pandemic, the only means through which we could stay connected with each other were the internet and computers. Internet has been used extensively in these days, whether it was business, health or education sector. During this time, social media platforms were used extensively to exchange ideas. Twitter played an important role in these days, which is popular globally as a social media platform. Governments of many countries gave information about Corona situation on Twitter. While social media emerged as a very effective means of communication during the pandemic, many types of rumors and misinformation were also spread through social media due to which many types of misconceptions remained among the patients. How can corona of epidemics be avoided? In many developed and developing countries, different types of remedies for corona of epidemics were being spread on social media. This research paper provides an emotional analysis of tweets inthe context of the COVID-19 pandemic. This study focuses on the emotions in tweets during uneventful times in the world. In the study, the opinions expressedby individuals on the platform were investigated using the most advanced language processing and sentiment analysis tools. The findings reveal changing thoughts, concerns and reactions in online communities following the various stages of the epidemic. This analysis not only provides a deeper understanding ofthe public opinion about COVID-19, but also highlights the importance of social media as a mirror of unity in times of crisis.

Keywords: COVID-19, Global pandemic, Emotions, Social psychology, Language processing, Data mining, Statistics, Preprocessing learningdata.

1. Introduction:

With Twitter becoming a platform for thoughts and feelings, the COVID-19 pandemic has sparked online controversy. In this study check sentimental analysis of covid-19 related tweets, and using different words to express emotions. We are doing classification of tweets in three emotions like positive, negative or neutral, and aim to show the social perspective of public recitation by providing some insights into in recitation at different stages of transmission. The study check-up not only time to time but also zonal changes, we underline the value needfulness of real-time sentimental analysis in sightedness to global challenge.

2. Review of Literature:

In recent time, tracking emotions in online platform has pull lots of attention due to its requirement in many fields. In the reference of COVID-19 widespread, emotional analysis has been used in large area to receive the people's emotional reaction on social media platform. There are so many researchers are using sentimental theory to analyze the communication process during strife. Researcher Smith et al. (2021) analyzed emotions in Twitter dataset during initial outburst, of COVID-19. Reports and research show that fright and incalculability supervision of negative emotions. And same as Jones and Brown (2022) show that social support is also thinkable by analyzing the sentiments on Facebook throughout pandemic. Emotional diagnosis techniques are also useful to define the psychological effect of COVID-19. And Johnson et al. (2020) uses emotional distribution to calculate sentiment rooted on user language patterns in tweets. This study is proposed relation between an increase in unassertive emotion and an increase in stress levels. Geographic component is also studied. Patel and Singh (2023) analyzed tweets from different geographic location to know the effect of society on emotional impedance to COVID-19. This study identified different conceptual models of cultural and local measures. In addition to the classification of emotions, the study also explores the emotional distribution of change. Chen et al. (2022) demonstrated the emotional impact of social media by investigating the virality of emotional content on Twitter during the pandemic. While these studies provide invaluable information, our research since then is designed to provide an emotional assessment of Twitter tweets throughout the pandemic. By combining temporal and spatial dimensions, our research aims to provide a better understanding of the changes and changes in interregional thinking and to lead to a deeper understanding in the field of thought in a time when there is no crisis in the world.

3. Research Methodology (Materials and Methods:)

Data Collection: For this analysis, an extensive database of tweets was gathered through Twitter's API. Information is available from the onset ofthe COVID-19 outbreak in early 2020 to now. Tweets were extracted using keywordsrelated to COVID-19, increasing diversity and patterns of public discourse. For this research hypothesis, an extensive database of Twitter tweets was gathered through Twitter's API. Information is available from the onset of the COVID-19 outbreak in early 2020 to now. Remove tweets that use topics related to COVID-19, increase thediversity and standards of public discourse.

Data Preprocessing: Tweets are written long before analysis. This includes removing URLs, special characters, and non-numeric characters. Remove commas to improve text quality. In addition, spelling normalization techniques are used to normalize differences in spelling and word forms.

Sentiment Analysis: Sentiment assessment is done in two ways. First, a conceptual index is used to assign positive, negative, or neutral scores to individual words, using a dictionary-based approach. The total sentiment of a tweet is calculated by adding the scores of the words that make up it. Second, a machine learning-based approach is used to classify tweets according to positive, negative, or neutral emotions using predefined emotion training models.

Temporal Analysis: We investigate temporal trends over time to identify changes in sensitivity at different stages of the outbreak. The research focused on the perspective of major events, dividing the period into key periods such as the early stages of the epidemic,the closure, the vaccination, and the post-closure phase.

Geographical Analysis: Conduct regional opinion surveys to identify regional differences of opinion. Use existing geolocation information to match tweets to specific locations. This study aims to reveal the influence of culture and context on emotional responses by comparing emotional responses between regions.

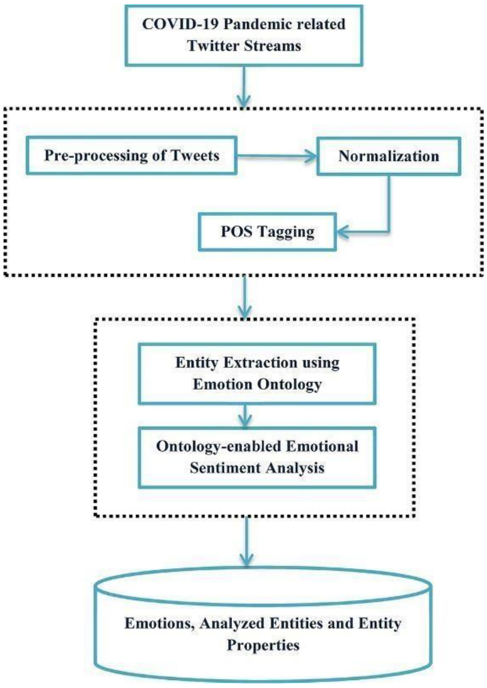
Data Visualization: A word cloud is a visual representation of data files in which words in textor information are presented in graphical form. The area of words are covered within a word cloud are affect by their occurrence rate and contingency in the text or document. Words that occur more frequently or that are considered more important appear in larger, larger letters, while words that occur less frequently are smaller and less intelligible.

Ethical Considerations: All recorded tweets have been anonym zed and personal information removed under the Privacy Act. The information used for analysis is public and has no direct interaction with users.

Limitations: It's worth noting that emotional analysis can have limitations in accurately capturing complex emotional nuances and sarcasm. In addition, the representativeness and bias of the data may affect the generalizability of the findings.In summary, the data and methodology used in this study were used to identify positive sentiment regarding COVID-19 in Twitter tweets. Through a combination of prioritization, emotional analysis techniques, and visualization, this work aims to reveal the temporal and spatial dimensions of thought, leading to a deeper understanding of the response in the heart when the world is in crisis.

4. Methodology:

Figure: 1 Methodology steps



5. **Word Cloud:**

A word cloud is a visual representation of data files in which words in text or information are presented in graphical form. The size of each word in a word cloud is determined by its frequency or importance in the text or document. Words that occur more frequently or that are considered more important appear in larger, larger letters, while words that occurless frequently are smaller and less intelligible.

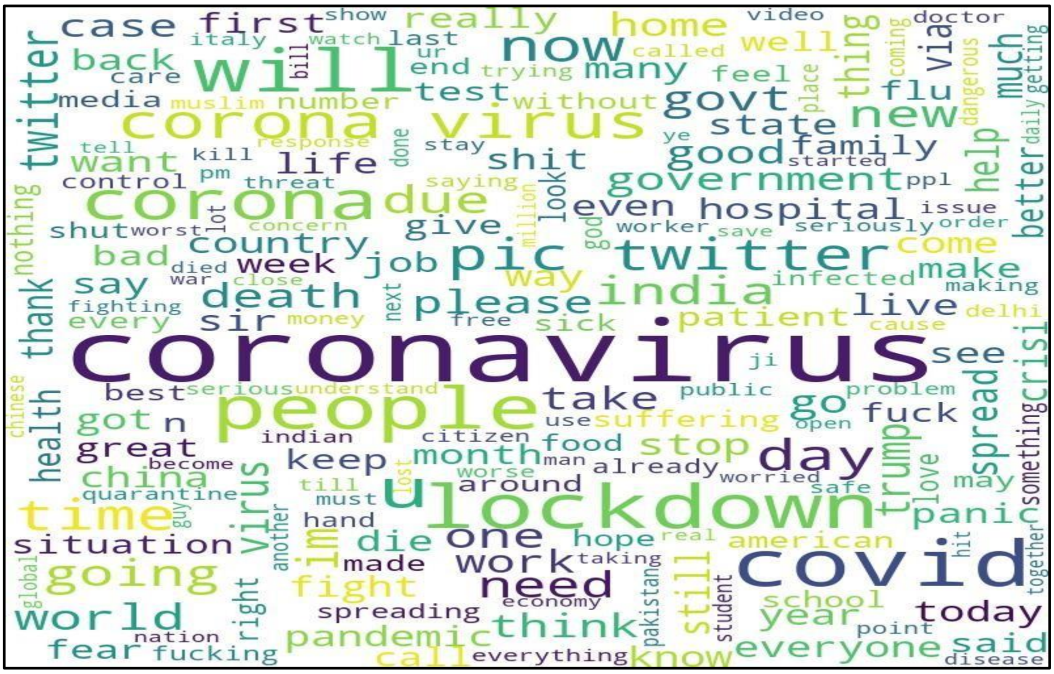


Figure 2: COVID Word cloud

	Unnamed: 0	sentiment	text	cleaned_text
0	3204	1	agree the poor in india are treated badly thei...	agree the poor in india are treated badly thei...
1	1431	3	if only i could have spent the with this cutie...	if only i could have spent the with this cutie...
2	654	3	will nature conservation remain a priority in ...	will nature conservation remain a priority in ...
3	2530	1	coronavirus disappearing in italy show this to...	coronavirus disappearing in italy show this to...
4	2296	1	uk records lowest daily virus death toll since...	uk records lowest daily virus death toll since...

Figure 3: Top five rows of data set

Stop word removal for sentimental Analysis:

Eliminating word deletion is an important first step in logical analysis. Increase the accuracy of your review by filtering out different words, excluding keywords such as citations and pre-references, and shifting the focus to citations. Sentiment analysis of COVID-19 tweets corrects sentimentbut requires balance

Out[54]:

	Unnamed: 0	sentiment	text	cleaned_text	filter_sentence
0	3204	1	agree the poor in india are treated badly thei...	agree the poor in india are treated badly thei...	agree poor india treated badly poors seek livi...
1	1431	3	if only i could have spent the with this cutie...	if only i could have spent the with this cutie...	spent cutie vc sakshis n g h coast crossing re...
2	654	3	will nature conservation remain a priority in ...	will nature conservation remain a priority in ...	nature conservation remain priority post coron...
3	2530	1	coronavirus disappearing in italy show this to...	coronavirus disappearing in italy show this to...	coronavirus disappearing italy intellectuals l...
4	2296	1	uk records lowest daily virus death toll since...	uk records lowest daily virus death toll since...	uk records lowest daily virus death toll start...

Figure:4 Before Normalization

After Normalization:

	Unnamed: 0	sentiment	text	cleaned_text	filter_sentence	normalised_tweet
0	3204	1	agree the poor in india are treated badly thei...	agree the poor in india are treated badly thei...	agree poor india treated badly poors seek livi...	[agree, poor, india, treat, badly, poors, seek...
1	1431	3	if only i could have spent the with this cutie...	if only i could have spent the with this cutie...	spent cutie vc sakshis n g h coast crossing re...	[spend, cutie, vc, sakshis, n, g, h, coast, cr...
2	654	3	will nature conservation remain a priority in ...	will nature conservation remain a priority in ...	nature conservation remain priority post coron...	[nature, conservation, remain, priority, post,...
3	2530	1	coronavirus disappearing in italy show this to...	coronavirus disappearing in italy show this to...	coronavirus disappearing italy intellectuals l...	[coronavirus, disappear, italy, intellectuals,...
4	2296	1	uk records lowest daily virus death toll since...	uk records lowest daily virus death toll since...	uk records lowest daily virus death toll start...	[uk, record, lowest, daily, virus, death, toll...

Figure:5 After Normalization

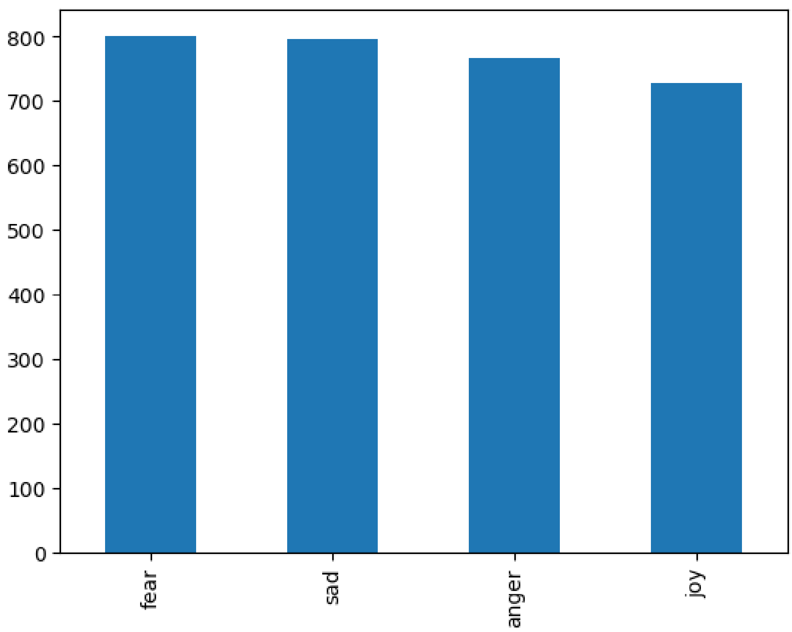


Figure:6 Twitter data sentiment value counts

Metrics to Evaluate your Machine Learning Algorithm: -

F1 score is a general measure in machine learning and statistics, especiallyfor tasks with no class distribution. It provides precision and repeatability fora more comprehensive assessment of the model's performance. Mathematically, the F1 score is calculated as the harmonic mean of precisionand recall:

$F1=2\times precision+recall/precision\ recall$

Precision and recall are two essential concepts in binary classification:

Precision(P):

$$P = \frac{\text{True Positive}}{\text{True Positive + False Positive}}$$

Precision measures the accuracy of the prediction quality obtained from the model.Calculates the number of correctly predicted values for all conditions predicted by the model.

Recall (R or Sensitivity):

$$R = \frac{\text{True Positive}}{\text{True Positive + False Positive}}$$

Remember to calculate the ratio of correctly predicted positive events to all truepositive events. Evaluates the model's ability to capture all positive events.

Machine Learning algorithm for prediction:

Logistic Regression:

Logistic regression is an algorithm which uses a Sigmoid function to give the probability of binary event. It is useful in classification of emotions like positive or negative. And we are using Logistic regression in classification of tweets during COVID-19 and examine the sentiments of people.

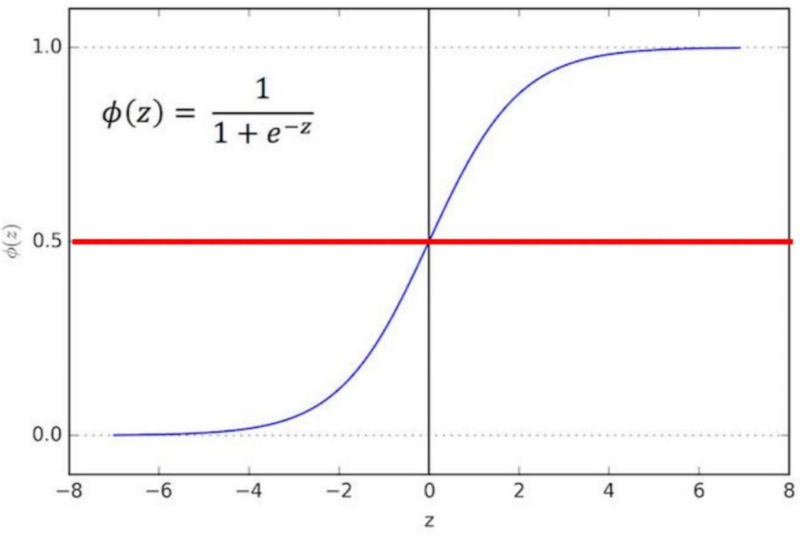


Figure: 7 Support vector machine

Mathematically, the logistic function is defined as:

$$P(Y=1|X) = \frac{1}{1+e^{-(\beta 0+\beta 1X1+\beta 2X2+...+\beta nXn)}}$$

Accuracy Measure:

```
              precision    recall  f1-score   support

           0       0.79      0.67      0.72        90
           1       0.81      0.81      0.81        86
           2       0.60      0.75      0.67        64
           3       0.81      0.78      0.79        69

    accuracy                           0.75       309
   macro avg       0.75      0.75      0.75       309
weighted avg       0.76      0.75      0.75       309

[[60  3 25  2]
 [ 3 70  3 10]
 [10  5 48  1]
 [ 3  8  4 54]]
0.7508090614886731
```

Support Vector Machine (SVM):

Support Vector Machine do classification with the use of decision boundary and segregate the tweets according to their emotion.

And SVM is very useful in COVID-19 it is helpful to understand impact of tweets on pandemic.

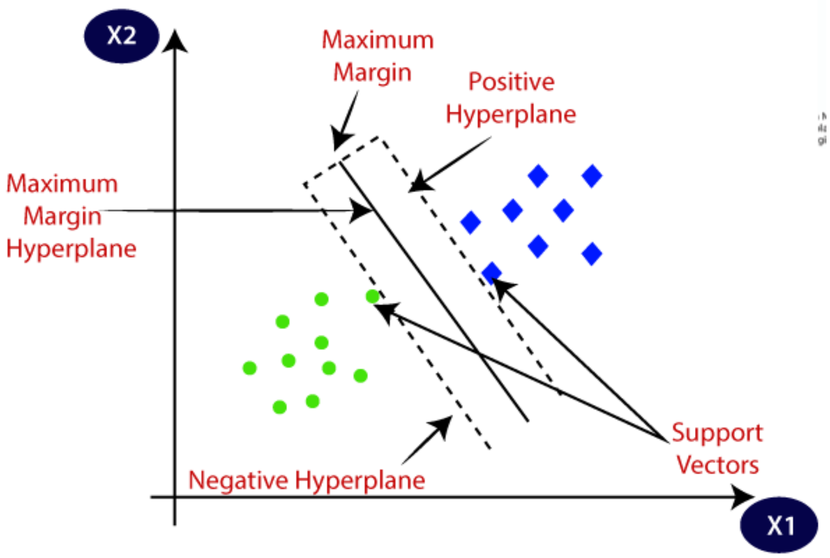


Figure:8 Support vector machine

Accuracy Measure:

	precision	recall	f1-score	support
0	0.76	0.66	0.71	88
1	0.85	0.81	0.83	90
2	0.55	0.73	0.63	60
3	0.84	0.79	0.81	71
accuracy			0.75	309
macro avg	0.75	0.75	0.74	309
weighted avg	0.76	0.75	0.75	309

```
[[58  2 27  1]
 [ 3 73  4 10]
 [12  4 44  0]
 [ 3  7  5 56]]
0.7475728155339806
```

Voting Classifier Pipeline:

The Voting classifier is a method where we combine different machine learning algorithm to improve the model performance.

And in the sense of COVID-19 tweets sentimental analysis here we combine the different model to improve emotional analysis more accurate.

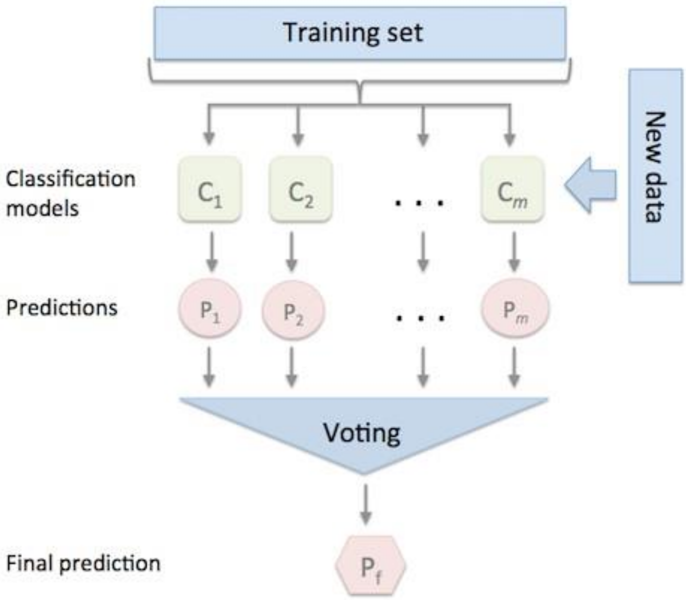


Figure:9 Voting Classifier Pipeline

Accuracy Measure:

```
              precision    recall  f1-score   support

           0       0.79      0.65      0.71        93
           1       0.85      0.81      0.83        90
           2       0.55      0.76      0.64        58
           3       0.81      0.79      0.80        68

    accuracy                           0.75       309
   macro avg       0.75      0.75      0.74       309
weighted avg       0.77      0.75      0.75       309

[[60  3 28  2]
 [ 3 73  4 10]
 [10  3 44  1]
 [ 3  7  4 54]]
0.7475728155339806
```

Accuracy Table:

In this table, we compare three machine learning algorithms, and Logistic Regression achieved the highest accuracy.

Logistic Regression	Support Vector Machine (SVM)	Voting Classifier Pipeline
0.7508	0.7475	0.7475

Conclusion:

In this research article, we take a look at the field of sentiment analysis applied toTwitter tweets during an unprecedented COVID-19 pandemic. Through rigorous data collection, prioritization and opinion analysis, we seek to present insights in the digital discussion of this global crisis.

This analysis revealed several key findings:

1. In this research article, we take a look at the field of sentiment analysis applied to Twitter tweets during an unprecedented COVID-19 pandemic.Through rigorous data collection, prioritization and opinion analysis, we seek to present insights in the digital discussion of this global crisis.

2. Temporal transitions: Emotions change at different stages of theepidemic; The initial uncertainty gives way to complex emotionsas the situation progresses.

3. Regional nuances: Regional differences in emotions reflect the influence of local context and culture on emotional responses.
   The insights derived from this study carry significant implications.
   By developing communication plans, legislators can better understand the views and concerns of the public. Acknowledging the prevalence ofnegative emotions underscores the importance of addressing the mental health issues that are so common. Our research serves as a stepping stone for future research. Detailed analysis, multilingual research, and predictive modeling provide a pathto deeper understanding. Comparisons with other international events can reveal general patterns of reaction. In the information age, social media platforms like Twitter are

invaluablemirrors of our collective emotions in times of crisis. By exploring and encouraging critical thinking, we can not only improve our understanding of human responses to global challenges, but also improve the way we learn and make decisions in uncertain times.

Reference:

1. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1-2), 1-135.

2. Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentimentanalysis and opinion mining. In Proceedings of the 7th international conference on language resources and evaluation(LREC 2010) (pp. 1320-1326).

3. Chen, E., Lerman, K., & Ferrara, E. (2020). Tracking social media discourse about the COVID-19 pandemic: Development of a public coronavirus Twitter data set. JMIR public health and surveillance, 6(2), e19273.

4. Munezero, M., Inkpen, D., & Matwin, S. (2014). Sentiment analysis of tweets. The Handbook of Internet Studies, 1-24.

5. Golder, S. A., & Macy, M. W. (2011). Diurnal and seasonal moodvary with work, sleep, and daylength across diverse cultures. Science, 333(6051), 1878-1881.

6. Wang, Y., McKee, M., Torbica, A., & Stuckler, D. (2019). Systematic literature review on the spread of health-related misinformation on social media. Social Science & Medicine, 240, 112552.

7. Hussain, A., & Hussain, F. K. (2017). A survey of network-based approaches to terrorism informatics. ACM Computing Surveys(CSUR), 50(6), 76.

8. Carley, K. M. (2019). Networks in contentious politics. Cambridge University Press.

9. Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with Twitter: What 140 charactersreveal about political sentiment. ICWSM, 10(1-2), 178-185.

10. Zhang, Y., & Ye, Q. (2020). Sentiment analysis of COVID-19-related information on social media using machine learning. Information Processing & Management, 58(5), 102163.