

Temporal Evolution of Public Health Sentiment: A Longitudinal Analysis

Samaneh MADANIAN^{a,1} and Vahid BAKHTIARI^b, Vincent Feng^a

^a*Department of Data Science and Artificial Intelligence, AUT, Auckland, New Zealand*

^b*School of Computing and Engineering, University of West London, London, UK*

Abstract. This study advances our understanding of public health crisis communication by conducting a longitudinal analysis. As COVID-19 has been the largest public health crisis to date, we performed sentiment analysis on it. While previous research focused on discrete time periods, our study examines the arc of pandemic-related discourse from 2020 to 2022, revealing long-term patterns in public sentiment evolution. Using advanced natural language processing techniques and temporal pattern analysis, we identify key transition points in public health discourse and sentiment, offering insights for future crisis communication strategies.

Keywords. Public Health, Text Mining, Digital Health, NLP, Disaster Response

1. Introduction and Background

A public health crisis, such as the COVID-19 pandemic, can impact billions of lives. In such circumstances, Social Media (SM) can become a vital channel for the public to share their opinions, concerns, and feelings regarding [1]. SM is indispensable in disseminating health-related information, with a significant proportion of people using these platforms to voice their healthcare concerns during the pandemic [2; 3]. For example, during the COVID-19 pandemic, as the number of confirmed cases rose and quarantine rules tightened, people became increasingly active in expressing their sentiments through social media. This makes SM a valuable dataset for public health crisis communication, which has undergone significant evolution throughout the COVID-19 pandemic. This necessitates a deeper understanding of how public sentiment and discourse patterns change over extended periods, which could help public health authorities make more informed decisions regarding their strategies to contain public health crises.

One key method for understanding public emotions is sentiment analysis. Sentiment analysis, also known as emotion AI or opinion mining, is a technique that uses text analysis, natural language processing (NLP), biometrics, and computational linguistics to systematically extract, identify, and quantify subjective information and emotional states. The metadata produced from social media can be processed and analysed to extract useful information, which can help decision-makers make more effective decisions for future crises and improve their responses to them.

Among various social media platforms, X (known as Twitter) provided an Application Programming Interface (API) that allowed researchers to analyse, learn

¹ Corresponding Author: Samaneh Madanian, sam.madanian@aut.ac.nz.

from, and engage in conversations programmatically. Several studies demonstrated that crisis-related tweets could offer valuable insights into specific situations [4; 5]. During COVID-19, [6] highlighted a relationship between COVID-19-related tweets and geographic locations. Public fears expressed on social media platforms were analysed by [7] during the early stages of the pandemic, while [8] examined COVID-19-related tweets to track the public's response to the pandemic. [15] aimed to gain insights into the pandemic situational information by exploring trends on social media. [9] developed a model to analyse shifting trends and sentiments, identifying false and misleading information about COVID-19 on Twitter. Their findings contributed to the early detection and elimination of fake news.

Consequently, it can be concluded that these strategies provided governments with insights into public reactions and helped them address the crisis. However, despite the existence of several studies on the sentiment analysis of COVID-19-related tweets, there has been limited research explicitly focused on long-term public health concerns. Therefore, this research aims to identify public health concerns expressed in COVID-19-related tweets and compare the differences between the first and second waves of the COVID-19 pandemic. The first wave occurred from mid-March to the end of June 2020, while the second wave occurred from mid-September to the end of December 2020. During these two waves, public health concerns likely shifted as people became more knowledgeable about the virus and how to protect themselves. Thus, it was both meaningful and necessary to explore the differences in public sentiment between the first and second waves of the pandemic. Such insights could help authorities and organizations optimise their strategies and responses to the global crisis. As a result, the primary objective of this research is to extract public sentiment and concerns from the tweet dataset during the COVID-19 outbreak using data analysis techniques, including machine learning (ML), natural language processing (NLP), and cloud computing services such as Amazon AWS. The study first identified the feasibility of extracting public sentiment and concerns from the tweet dataset, then analysed the sentiment differences between the two waves of the pandemic.

2. Method

We used IEEE Coronavirus (COVID-19) Tweets Dataset [10]. We used only two-period datasets: March 20, 2020, to June 30, 2020, and September 15, 2020, to December 21, 2020, which encompass the two outbreak waves of the COVID-19 pandemic. Once the tweet data were hydrated, several pre-processing steps were carried out to clean the dataset for analysis. These steps included removing non-English tweets, duplicate tweets, and links, as well as removing mentions, hashtags, and other irrelevant characters (e.g., punctuation) to make the data more suitable for analysis. Following pre-processing, the data were analysed using NLP techniques and machine learning tools (ML). We specifically focused on analyzing the tweet volume between two waves, their sentiments, and the developing trend of public sentiment.

3. Results and Discussion

This research aimed to extract public sentiment and healthcare concerns related to COVID-19 by analysing large-scale Twitter data during two significant pandemic waves.

The study leveraged advanced techniques in NLP and ML. We used Amazon Web Services (AWS) cloud computing to process and analyse tweets effectively. The application of AWS cloud services helped address the limitations of traditional ML and NLP methods. AWS provided efficient handling of unstructured data and scalable solutions for large-scale data analysis, with flexible and cost-effective resources based on actual usage. Through this analysis, several significant findings emerged, shedding light on public reactions during the pandemic and offering evidence to guide future public health crisis management strategies. Notably, public healthcare concerns were effectively identified and examined using Twitter data.

From mid-March to June 2020, the first wave of the outbreak experienced a significant rise in cases and deaths. The daily confirmed cases increased from 158,000 to 1.3 million, and deaths rose from 7,000 to 32,000 [2]. In the second wave, spanning from mid-September to December 2020, daily confirmed cases escalated from 2.1 million to 4.5 million, with deaths rising from 36,000 to 80,000. This global context served as a backdrop for understanding public sentiment throughout the pandemic. This global trend provided context for understanding public sentiment during the pandemic. In the first wave, there was a clear increase in tweet volume, starting with 50,000 daily tweets, rising to 150,000 mid-waves, and reaching 250,000 by the end. This trend indicates a growing engagement among the public as the pandemic continued. In contrast, the second wave exhibited a more stable tweet volume, fluctuating between 120,000 and 350,000 tweets daily. This steadiness suggests that the public's reaction became more consistent during the second wave, potentially reflecting a greater level of adjustment or habituation to the ongoing situation. Notably, COVID-19-related tweets surged significantly during the first wave and remained consistently high throughout the second wave, highlighting sustained public engagement.

Figure 1 illustrates the sentiment polarity of tweets during the two waves. Comparing the two waves, the percentage of people who held positive sentiments in the second wave is higher, while neutral and negative sentiments in the second wave are slightly lower than in the first wave. This shift suggests that, as people became more familiar with the pandemic and its measures, their overall outlook improved, as fewer expressed negative sentiments during the second wave. Also, the public became more positive as they adapted to the crisis and government measures.

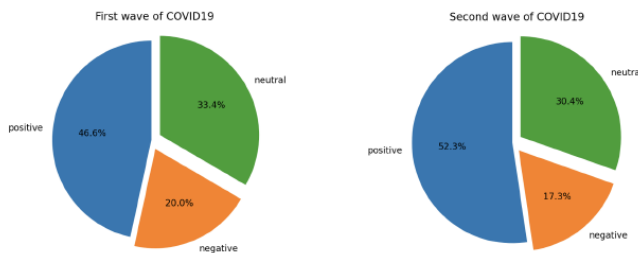


Figure 1. Sentiment polarity of COVID-19-related tweets during the first and second waves.

Figure 2 depicts public sentiment by geographic region, revealing a predominantly more positive reaction in developed countries that have better access to information and healthcare resources. The analysis of frequently mentioned words within the tweets highlighted an evolution in public concerns. During the first wave, the primary focus centered on government responses and testing. In contrast, the second wave saw discussions shift toward vaccines, quarantine, and personal well-being. Additionally, the

study identified regional differences in sentiment, with urban areas generally expressing more positive sentiments compared to rural areas. This finding emphasizes the necessity for region-specific public health responses.



Figure 2. Public sentiment by geographic locations.

Regarding key government interventions, sentiment toward measures such as lockdowns, mask mandates, and social distancing was largely positive, as illustrated in Figure 3. The analysis can reveal public sentiment toward government policies, such as Lockdown rules, wearing face masks, and maintaining social distance in public areas. The result could insist the government review and improve the policies and strategies.



Figure 3. Sentiment polarity of COVID-19-related tweets in health concerns.

The implications of this study are significant for policymakers and public health authorities. Real-time monitoring of public sentiment during a crisis enables adjustments to communication strategies and interventions that better align with public concerns, which align with literature findings in [11]. Furthermore, the positive shift in sentiment observed during the second wave suggests that effective communication and transparent government actions cultivate public trust over time. This observation is consistent with the findings of [12], which suggest that government transparency during health crises has a significant impact on public trust and adherence to health measures. Our findings extend this work by demonstrating how sentiment evolves across multiple pandemic waves, providing a longitudinal perspective that was lacking in previous studies. The regional differences in sentiment identified in our analysis align with the findings of [13],

who observed variations in public responses to COVID-19 across different geopolitical zones. However, our study goes further by correlating these differences with access to healthcare resources and information. The research also highlights the importance of utilizing scalable and flexible technology, such as cloud services, to efficiently process and analyze large volumes of data, thereby providing a practical framework for future research and crisis management that relies on real-time data analysis.

However, the study has faced some limitations. It exclusively analyzed English-language tweets, which restricts the representation of global public sentiment. Future research could also consider employing more advanced sentiment analysis models to enhance the accuracy and depth of sentiment detection and incorporate multilingual data to capture a broader array of public concerns. Additionally, while this study focused solely on text-based sentiment analysis, future investigations could explore visual and multimedia content, which are increasingly prevalent on social media platforms like Twitter. Further research should also consider employing more advanced sentiment analysis models, such as deep learning techniques, to enhance the accuracy and depth of sentiment detection. Expanding the analysis to include real-time data collection during forthcoming health crises would enable governments to respond more effectively to emerging public concerns.

Reference

- [1]. Y. Wu and D.W. Feng, Beyond reporting: The communicative functions of social media news during the COVID-19 pandemic, in: *Discourses, Modes, Media and Meaning in an Era of Pandemic*, Routledge, 2022, pp. 178-200.
- [2]. World Health Organization, Who coronavirus disease (COVID-19) Dashboard: World Health organization, 2020, in, World Health Organization, 2020.
- [3]. S. Schumacher and N. Kent, 8 charts on internet use around the world as countries grapple with COVID-19, (2020).
- [4]. S. Madanian, H. Rasoulipannah, and J. Yu, Stress Detection on Social Network: Public Mental Health Surveillance: Public mental health surveillance, in: *Proceedings of the 2023 Australasian Computer Science Week*, Association for Computing Machinery, Melbourne, VIC, Australia, 2023, pp. 170–175.
- [5]. K. Zahra, M. Imran, and F.O. Ostermann, Automatic identification of eyewitness messages on twitter during disasters, *Information Processing & Management* **57** (2020), 102107.
- [6]. J. Samuel, G.G.M.N. Ali, M.M. Rahman, E. Esawi, and Y. Samuel, COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification, *Information* **11** (2020), 314.
- [7]. S. Boberg, T. Quandt, T. Schatto-Eckrodt, and L. Frischlich, Pandemic populism: Facebook pages of alternative news media and the corona crisis--A computational content analysis, *arXiv preprint arXiv:2004.02566* (2020).
- [8]. K. Jahanbin and V. Rahmanian, Using twitter and web news mining to predict COVID-19 outbreak, *Asian Pacific journal of tropical medicine* **13** (2020), 378-380.
- [9]. K. Sharma, S. Seo, C. Meng, S. Rambhatla, and Y. Liu, Covid-19 on social media: Analyzing misinformation in twitter conversations, *arXiv preprint arXiv:2003.12309* (2020).
- [10]. R. Lamsal, Design and analysis of a large-scale COVID-19 tweets dataset, *Applied Intelligence* **51** (2021), 2790-2804.
- [11]. H. Rasouli Panah, S. Madanian, and J. Yu, Disaster Health Care and Resiliency: A Systematic Review of the Application of Social Network Data Analytics, *Disaster Medicine and Public Health Preparedness* **18** (2024), e334.
- [12]. J. Liu, Y. Shahab, and H. Hoque, Government Response Measures and Public Trust during the COVID-19 Pandemic: Evidence from Around the World, *British Journal of Management* **33** (2022), 571-602.
- [13]. S.R. Rufai and C. Bunce, World leaders' usage of Twitter in response to the COVID-19 pandemic: a content analysis, *Journal of Public Health* **42** (2020), 510-516.