

# REVIEW OF TWITTER STREAM ANALYSIS FOR REAL-TIME SENTIMENT MONITORING OF COVID-19 DISCOURSE

Name - Ambuj Prasad Mishra

Guide Name :- Dr. Ankit Temurnukar

Department of Computer Science

Madhyanchal Professional University, Bhopal (M.P.)

## Abstract

The rapid dissemination of information on social media platforms during the COVID-19 pandemic underscores the need for real-time sentiment analysis to gauge public opinion and emotional responses. This paper presents a review of Twitter stream analysis techniques employed for monitoring sentiment surrounding COVID-19 discourse. Utilizing natural machine learning (ML) algorithms, researchers have developed various methodologies to analyze the sentiment of tweets related to the pandemic. Key aspects of these approaches include data preprocessing techniques, feature extraction methods, sentiment classification models, and evaluation metrics. Additionally, the review discusses challenges such as handling noisy and unstructured data, addressing linguistic nuances, and adapting to evolving discourse trends.

## Introduction

The COVID-19 pandemic has not only brought about a global health crisis but has also generated an unprecedented volume of discourse on social media platforms such as Twitter. The rapid spread of information and opinions in real-time necessitates effective tools for monitoring public sentiment surrounding the pandemic. In response to this need, Twitter stream analysis has emerged as a valuable approach for extracting insights from the vast amount of COVID-19-related tweets. This paper provides an overview of the methodologies and findings in the realm of real-time sentiment monitoring of COVID-19 discourse using Twitter stream analysis. Twitter serves as a rich source of real-time data, offering a diverse range of opinions, emotions, and reactions from users worldwide. Leveraging this data through sentiment analysis techniques enables researchers and practitioners to gauge public sentiment towards various aspects of the pandemic, including government responses, public health measures, vaccine distribution, and misinformation. By understanding public sentiment, stakeholders can better tailor communication strategies, identify emerging concerns, and address misinformation effectively. The process of Twitter stream analysis for sentiment monitoring involves several key steps. Firstly, raw Twitter data is collected using streaming APIs or historical data archives. Preprocessing techniques such as text normalization, tokenization, and filtering are applied to clean the data and remove noise. Feature extraction methods, including bag-of-words, word embeddings, and topic modeling, are then employed to represent tweets in a format suitable for sentiment analysis. Sentiment classification models, ranging from traditional machine learning algorithms to more advanced

deep learning architectures, are utilized to classify tweets into sentiment categories such as positive, negative, or neutral. Evaluation metrics such as accuracy, precision, recall, and F1-score are used to assess the performance of these models.

### **Need of the Study**

The need for this study stems from the critical importance of understanding public sentiment surrounding the COVID-19 pandemic in real-time. With the proliferation of social media usage, particularly on platforms like Twitter, capturing and analyzing public discourse provides invaluable insights for public health authorities, policymakers, and the general public. Monitoring sentiment aids in gauging public perceptions, identifying misinformation, and adapting communication strategies effectively. Additionally, real-time sentiment analysis enables swift responses to emerging concerns and crises, contributing to more proactive and informed decision-making. By investigating the methodologies and findings of Twitter stream analysis for sentiment monitoring of COVID-19 discourse, this study aims to contribute to the development of robust tools and strategies for managing public health crises and mitigating the impact of the pandemic.

### **Real-Time Sentiment Prediction System**

A Real-Time Sentiment Prediction System represents a cutting-edge technology that holds immense promise and significance in today's data-driven world. With a focus on analyzing and forecasting sentiment, this system offers a range of applications that span various sectors, from business to politics and beyond.

At its core, a Real-Time Sentiment Prediction System is a data-driven tool that leverages machine learning and natural language processing techniques to continuously monitor and assess the sentiments expressed in textual data, often in the form of tweets, reviews, news articles, or social media posts. Its ability to provide real-time predictions of sentiment allows for immediate responses and data-informed decision-making. This system is a game-changer. Companies can utilize it to gauge customer satisfaction, market perception of their products, and emerging trends. By assessing real-time sentiment, businesses can adapt their strategies swiftly, identify areas for improvement, and enhance customer engagement. For example, a restaurant chain can track customer reviews in real time, allowing them to address issues quickly and maintain a positive image.

In politics, a Real-Time Sentiment Prediction System provides invaluable insights into public opinion. Political campaigns can assess how their messages are resonating with the electorate, allowing them to refine their strategies and adapt their campaign narratives. Government agencies can also monitor public sentiment on policy issues, aiding in responsive governance and crisis management. The significance of this system extends to public health. During a public health crisis, such as the ongoing COVID-19 pandemic, it can serve as an early warning system by gauging public sentiment on preventive measures, vaccines, and government responses. Health authorities can use this data to tailor their communication strategies, address public concerns, and combat the spread of misinformation. For researchers and analysts, a Real-Time Sentiment Prediction System opens up new avenues for trend

analysis and academic studies. It enables the examination of evolving sentiments over time, helping in understanding the dynamics of public opinion and the effectiveness of interventions. Real-Time Sentiment Prediction System is a transformative technology with profound significance. It empowers businesses, politicians, health authorities, and researchers with real-time insights into public sentiment. This capability for immediate response to changing sentiment makes it a critical asset in an ever-evolving, data-driven world. As our society becomes increasingly interconnected and data-driven, the ability to predict and respond to sentiment in real-time positions this system at the forefront of innovation, offering a wealth of opportunities to enhance decision-making and improve engagement with the public.

### Literature review

#### Related Work

**Zhang, X., et al (2020).** Predicting and managing the trajectory of a global pandemic like COVID-19 in real-time is an urgent and complex challenge. Leveraging the power of machine learning and big data streaming systems, this research aims to develop a robust predictive framework. By harnessing diverse data sources such as medical records, social media trends, travel data, and more, our system continuously collects and analyzes real-time information. Machine learning models, including deep learning and ensemble techniques, are employed to identify patterns and anomalies within this data. These models adapt and evolve as new information becomes available, enhancing prediction accuracy. The integration of big data streaming systems enables the seamless flow of real-time data, allowing for rapid response to emerging trends and threats. By monitoring various epidemiological indicators and social behavior, our system can provide early warnings of potential outbreaks, enabling authorities to implement targeted interventions swiftly. Additionally, it aids in resource allocation by predicting healthcare needs in specific regions. This research not only empowers governments and healthcare organizations to make informed decisions but also supports public awareness and engagement. By disseminating timely information and recommendations, we can collectively respond more effectively to pandemics, ultimately reducing their impact on society. This innovative approach harnessing machine learning and big data streaming systems offers a promising solution for real-time pandemic prediction and response, serving as a critical tool in our ongoing battle against global health crises.

**Krishnan, H., et al (2021).** Analyzing the sentiment of Twitter data related to the coronavirus disease (COVID-19) using machine learning techniques is a crucial endeavor in understanding public perceptions and emotions during a global health crisis. This research focuses on developing a comprehensive sentiment analysis framework that harnesses the power of machine learning algorithms. Twitter, as a vast and real-time social media platform, serves as a valuable source of public sentiment data. Our approach involves collecting a large dataset of COVID-19-related tweets and preprocessing the text data to remove noise and irrelevant information. We then employ state-of-the-art machine learning models, such as deep neural networks and natural language processing algorithms, to classify tweets into various sentiment categories, including positive, negative, and neutral. By analyzing this

sentiment-labeled data, we gain insights into the evolving emotions and attitudes of the public towards COVID-19.

**Okango, E., & Mwambi, H. (2022).** Conducting a dictionary-based global sentiment analysis of Twitter data related to the effects and responses to the coronavirus disease (COVID-19) pandemic is a vital endeavor in understanding public perceptions and attitudes on a global scale. This research focuses on utilizing established sentiment dictionaries to gain insights into the sentiments expressed in tweets pertaining to COVID-19 and its impacts. Twitter, as a ubiquitous platform for real-time information sharing, serves as a vast source of data reflecting public sentiment. Our approach involves collecting a substantial dataset of COVID-19-related tweets and applying sentiment lexicons, which are curated lists of words categorized by sentiment (positive, negative, or neutral), to quantify the sentiment expressed in each tweet. By utilizing sentiment dictionaries, we can discern the prevailing emotional tones in tweets discussing various aspects of COVID-19, including its effects on health, society, and the economy, as well as the responses of governments and healthcare systems. This analysis enables us to identify the predominant sentiment trends globally, providing valuable insights into how people perceive and react to the pandemic. Our research explores geographical variations in sentiment, allowing us to compare sentiments across different regions and nations.

**Boon-Itt, S., & Skunkan, Y. (2020).** Understanding the public perception of the COVID-19 pandemic is crucial for shaping effective communication, policies, and public health responses. In this study, we employ sentiment analysis and topic modeling techniques to delve into Twitter data and extract valuable insights into how the public perceives and discusses the COVID-19 pandemic. Using a large and diverse dataset of COVID-19-related tweets, we first conduct sentiment analysis to categorize tweets into positive, negative, or neutral sentiments. By quantifying the emotional tone of public discourse, we identify prevailing sentiments over time and across different regions. This analysis aids in tracking sentiment shifts in response to evolving pandemic circumstances and governmental interventions.

**Sanders, A. C., et al (2021).** "Unmasking the Conversation on Masks" delves into the rich tapestry of public discourse on Twitter surrounding the use of face masks during the COVID-19 pandemic. Employing advanced natural language processing techniques, this study focuses on topical sentiment analysis to unearth key insights and sentiments associated with mask-related discussions. In the midst of a global health crisis, understanding the dynamics of public sentiment and perception regarding mask-wearing is critical. Our research leverages a vast dataset of COVID-19-related tweets to scrutinize sentiment in relation to masks. We employ sophisticated natural language processing tools to categorize tweets into distinct sentiment categories, elucidating whether discussions are predominantly positive, negative, or neutral. Beyond sentiment analysis, we employ topic modeling methods, such as Latent Dirichlet Allocation (LDA), to unveil the underlying themes and topics within the discourse. This approach allows us to pinpoint recurring subjects, concerns, and evolving discussions related to mask usage.

**Srikanth, J., et al (2022).** "Sentiment Analysis on COVID-19 Twitter Data Streams Using Deep Belief Neural Networks" presents an innovative approach to discerning public sentiment during the COVID-19 pandemic by harnessing the power of deep learning. In this study, we leverage Deep Belief Neural Networks (DBNNs) to analyze real-time Twitter data, shedding light on the evolving sentiments expressed by users in response to the pandemic. With Twitter serving as a prolific platform for instant information dissemination and expression, it provides a wealth of data for gauging public sentiment. Our research employs a vast and continuously updated dataset of COVID-19-related tweets. DBNNs, known for their ability to capture intricate patterns and relationships within data, are utilized to perform sentiment analysis by classifying tweets as positive, negative, or neutral. What sets this study apart is its real-time nature. As the pandemic unfolds, public sentiments are fluid and dynamic.

**Naseem, U., et al (2021).** "COVIDSenti" represents a significant contribution to the field of sentiment analysis by introducing a comprehensive and large-scale benchmark Twitter dataset specifically designed for analyzing sentiment related to the COVID-19 pandemic. This dataset serves as a valuable resource for researchers and practitioners seeking to understand public sentiments and emotions during this unprecedented global health crisis. COVIDSenti consists of a vast and diverse collection of tweets related to COVID-19, aggregated from the Twitter platform. The dataset encompasses a wide range of topics, including discussions on the virus itself, public health measures, government responses, economic impacts, and social consequences.

**Pristiyono, et al (2021).** This study focuses on conducting sentiment analysis of public opinions regarding COVID-19 vaccines in Indonesia, employing the Naïve Bayes algorithm. As the global effort to combat the COVID-19 pandemic largely hinges on successful vaccination campaigns, understanding public sentiment toward vaccines is of paramount importance. To achieve this, a dataset comprising Indonesian social media posts, comments, and tweets related to COVID-19 vaccines is collected and preprocessed. The Naïve Bayes algorithm is then applied to classify these posts into sentiment categories, specifically positive, negative, or neutral, based on the language and emotions expressed. The Naïve Bayes algorithm is chosen for its effectiveness in text classification tasks and its ability to handle large datasets with multiple features. By leveraging this algorithm, we gain insights into how the Indonesian population perceives COVID-19 vaccines, whether they are enthusiastic, skeptical, or indifferent. This information is invaluable for public health authorities and policymakers as it can inform communication strategies, address concerns, and promote vaccine acceptance.

**Alanezi, M. A., & Hewahi, N. M. (2020).** Analyzing sentiment in tweets during the COVID-19 pandemic is essential for comprehending the public's emotional response, concerns, and attitudes towards this global crisis. This study conducts sentiment analysis on a diverse dataset of tweets spanning the pandemic's duration to gain valuable insights into the evolving sentiments of individuals and communities. Our research employs natural language processing



(NLP) techniques and machine learning algorithms to categorize tweets into sentiment categories: positive, negative, or neutral. By examining large volumes of data, we uncover sentiment trends, fluctuations, and regional variations. These insights offer a window into how the pandemic has affected people's mental and emotional well-being, as well as their opinions on various pandemic-related topics.

**Al-Hashedi, A., et al (2022).** This research focuses on the application of ensemble classifiers for sentiment analysis of Arabic social media data, particularly Twitter, concerning the proliferation of COVID-19-related conspiracy theories. The spread of misinformation and conspiracy theories on social networks has been a significant challenge during the pandemic, warranting a deep understanding of public sentiment towards such content. Our study commences by collecting a substantial dataset of Arabic tweets containing discussions, mentions, and references to COVID-19 conspiracy theories. These tweets are diverse and encompass a broad spectrum of sentiments, including positive, negative, and neutral. Ensemble classifiers, renowned for their ability to enhance classification accuracy by combining multiple models, are then employed to categorize these tweets into sentiment classes.

**Kruspe, A., et al (2020).** This study presents a cross-language sentiment analysis of Twitter messages from various European countries during the COVID-19 pandemic. As the pandemic swept across Europe, it prompted diverse reactions and discussions in multiple languages on social media platforms, making this research essential for understanding the nuanced sentiments expressed by different linguistic communities. Our research begins with the collection of a substantial and diverse dataset of Twitter messages related to COVID-19 from European countries, encompassing various languages such as English, French, Spanish, German, and more. To conduct cross-language sentiment analysis, we employ state-of-the-art natural language processing techniques and machine learning algorithms that can effectively handle multilingual data. The primary objective is to categorize the tweets into different sentiment classes, including positive, negative, and neutral sentiments, across various languages. This allows us to gain insights into how different linguistic communities in Europe perceived and reacted to the pandemic. Furthermore, the analysis explores temporal and geographical trends, enabling us to track sentiment changes over time and across different regions.

**Garcia, M. B. (2020).** This study conducts sentiment analysis on tweets related to the Coronavirus Disease 2019 (COVID-19) pandemic originating from Metro Manila, Philippines. As a densely populated urban center in the Philippines, Metro Manila has been significantly affected by the pandemic, making it an essential region for sentiment analysis to understand public perceptions, emotions, and attitudes during this global health crisis. Our research involves the collection of a substantial dataset of tweets specifically geo-tagged to Metro Manila and containing references to COVID-19. Through natural language processing and machine learning techniques, we classify these tweets into different sentiment categories, including positive, negative, or neutral. This classification enables us to gain insights into the prevailing sentiments of individuals and communities in Metro Manila regarding the pandemic. By examining sentiment trends over time, we can detect shifts in public emotions

as the situation evolves, reflecting the impact of government interventions, healthcare updates, and societal responses.

**Alabid, N. N., & Katheeth, Z. D. (2021).** This study conducts sentiment analysis on tweets related to the Coronavirus Disease 2019 (COVID-19) pandemic originating from Metro Manila, Philippines. As a densely populated urban center in the Philippines, Metro Manila has been significantly affected by the pandemic, making it an essential region for sentiment analysis to understand public perceptions, emotions, and attitudes during this global health crisis. Our research involves the collection of a substantial dataset of tweets specifically geo-tagged to Metro Manila and containing references to COVID-19. Through natural language processing and machine learning techniques, we classify these tweets into different sentiment categories, including positive, negative, or neutral. This classification enables us to gain insights into the prevailing sentiments of individuals and communities in Metro Manila regarding the pandemic.

**Mir, A. A., & Sevukan, R. (2022).** This study conducts a sentiment analysis of tweets originating from India pertaining to COVID-19 vaccines, providing crucial insights into the public sentiment, attitudes, and perceptions surrounding vaccine acceptance in this populous and diverse country. To conduct this analysis, we compile a comprehensive dataset of tweets specifically focused on COVID-19 vaccines within the Indian context. This dataset captures the wide spectrum of opinions and emotions expressed by Indian Twitter users, offering a nuanced view of how the population perceives and discusses COVID-19 vaccination efforts.

### **Research Problem**

The research problem addressed in this study is the need for effective real-time monitoring of sentiment surrounding COVID-19 discourse on Twitter. With the proliferation of social media usage during the pandemic, understanding public sentiment is crucial for informing public health responses, policy decisions, and communication strategies. However, analyzing sentiment in real-time presents challenges due to the vast volume and rapid pace of Twitter data, as well as the need to account for linguistic nuances and evolving discourse trends. This research aims to investigate methodologies and techniques for extracting meaningful insights from Twitter streams, including data preprocessing, feature extraction, sentiment classification, and evaluation metrics. By addressing these challenges, the study seeks to contribute to the development of robust tools and strategies for monitoring public sentiment during health crises, ultimately aiding in the effective management and communication of COVID-19-related information.

### **Conclusion**

In the context of isolation during the COVID-19 pandemic, people often turn to social media to express their thoughts and emotions. While social media serves as a valuable source of real-time information about the virus, it can sometimes be unreliable or misleading. Encountering misleading or distressing information on social media can exacerbate the difficulties individuals are already facing. In the new normal of staying at home, working remotely, and experiencing isolation, social networking platforms have become the primary means of sharing news, opinions, emotions, and advice. The spread of misinformation on social media, which involves disseminating false or irrelevant information, can be

detrimental. For instance, claims like "eating bananas prevents COVID-19" contribute to the problem. Individuals grappling with the effects of the disease, both physically and mentally, require logical approaches to make sense of the vast amount of information available. Online social media platforms such as Twitter and Facebook host substantial amounts of data, much of which is noisy and challenging to navigate. However, once this noisy data is filtered and cleaned, it becomes a rich source of insights into human emotions, expressions, and thoughts. A careful analysis of this data unveils a wealth of information about the prevailing mood, attitudes, and characteristics of large human communities. In these trying times, the importance of monitoring and analyzing public sentiment regarding the coronavirus cannot be overstated. This system, designed to work in real-time with Twitter streaming data, provides a unique perspective on the ever-changing landscape of the pandemic. With its ability to gauge the public's feelings, opinions, and reactions, it serves as a powerful tool for governments, healthcare professionals, and the broader public.

## References

1. Zhang, X., Saleh, H., Younis, E. M., Sahal, R., & Ali, A. A. (2020). Predicting coronavirus pandemic in real-time using machine learning and big data streaming system. *Complexity*, 2020, 1-10.
2. Krishnan, H., Pankajkumar, G., Poosari, A., Jayaraj, A., Thomas, C., & Joy, G. M. (2021, May). Machine learning based sentiment analysis of coronavirus disease related twitter data. In *2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC)* (pp. 459-464). IEEE.
3. Okango, E., & Mwambi, H. (2022). Dictionary based global Twitter sentiment analysis of coronavirus (Covid-19) effects and response. *Annals of Data Science*, 9(1), 175-186.
4. Boon-Itt, S., & Skunkan, Y. (2020). Public perception of the COVID-19 pandemic on Twitter: sentiment analysis and topic modeling study. *JMIR Public Health and Surveillance*, 6(4), e21978.
5. Sanders, A. C., White, R. C., Severson, L. S., Ma, R., McQueen, R., Paulo, H. C. A., ... & Bennett, K. P. (2021). Unmasking the conversation on masks: Natural language processing for topical sentiment analysis of COVID-19 Twitter discourse. *AMIA Summits on Translational Science Proceedings*, 2021, 555.
6. Srikanth, J., Damodaram, A., Teekaraman, Y., Kuppusamy, R., & Thelkar, A. R. (2022). Sentiment analysis on COVID-19 Twitter data streams using deep belief neural networks. *Computational intelligence and neuroscience*, 2022.
7. Naseem, U., Razzak, I., Khushi, M., Eklund, P. W., & Kim, J. (2021). COVIDSenti: A large-scale benchmark Twitter data set for COVID-19 sentiment analysis. *IEEE transactions on computational social systems*, 8(4), 1003-1015.
8. Pristiyono, Ritonga, M., Ihsan, M. A. A., Anjar, A., & Rambe, F. H. (2021, February). Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1088, No. 1, p. 012045). IOP Publishing.



9. Alanezi, M. A., & Hewahi, N. M. (2020, October). Tweets sentiment analysis during COVID-19 pandemic. In 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI) (pp. 1-6). IEEE.
10. Al-Hashedi, A., Al-Fuhaidi, B., Mohsen, A. M., Ali, Y., Gamal Al-Kaf, H. A., Al-Sorori, W., & Maqtary, N. (2022). Ensemble classifiers for Arabic sentiment analysis of social network (Twitter data) towards COVID-19-related conspiracy theories. *Applied Computational Intelligence and Soft Computing*, 2022, 1-10.
11. Kruspe, A., Häberle, M., Kuhn, I., & Zhu, X. X. (2020). Cross-language sentiment analysis of european twitter messages during the covid-19 pandemic. *arXiv preprint arXiv:2008.12172*.
12. Garcia, M. B. (2020). Sentiment analysis of tweets on coronavirus disease 2019 (COVID-19) pandemic from Metro Manila, Philippines. *Cybernetics and Information Technologies*.
13. Alabid, N. N., & Katheeth, Z. D. (2021). Sentiment analysis of twitter posts related to the covid-19 vaccines. *Indonesian Journal of Electrical Engineering and Computer Science*, 24(3), 1727-1734.
14. Ragothaman, A., & Huang, C. Y. (2021). Sentiment analysis on Covid-19 twitter data. *International Journal of Computer Theory and Engineering*, 13(4), 100-107.
15. Mir, A. A., & Sevukan, R. (2022). Sentiment analysis of Indian Tweets about Covid-19 vaccines. *Journal of Information Science*, 01655515221118049.
16. Sharma, K., Seo, S., Meng, C., Rambhatla, S., Dua, A., & Liu, Y. (2020). Coronavirus on social media: Analyzing misinformation in Twitter conversations.
17. Aygün, I., Kaya, B., & Kaya, M. (2021). Aspect based twitter sentiment analysis on vaccination and vaccine types in covid-19 pandemic with deep learning. *IEEE Journal of Biomedical and Health Informatics*, 26(5), 2360-2369.
18. Qorib, M., Oladunni, T., Denis, M., Ososanya, E., & Cotae, P. (2023). Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset. *Expert Systems with Applications*, 212, 118715.
19. Medford, R. J., Saleh, S. N., Sumarsono, A., Perl, T. M., & Lehmann, C. U. (2020, July). An “infodemic”: leveraging high-volume Twitter data to understand early public sentiment for the coronavirus disease 2019 outbreak. In *Open forum infectious diseases* (Vol. 7, No. 7, p. ofaa258). US: Oxford University Press.
20. Venigalla, A. S. M., Chimalakonda, S., & Vagavolu, D. (2020, October). Mood of India during Covid-19-An interactive web portal based on emotion analysis of Twitter data. In *Conference companion publication of the 2020 on computer supported cooperative work and social computing* (pp. 65-68).
21. Garcia, K., & Berton, L. (2021). Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. *Applied soft computing*, 101, 107057.