



# Post-Pandemic Anxiety Prediction Model for Malaysia Urban School Students Using Social Media Slang Analytics

POTTELLA CHANDRA CHAKITHA

Department of Computer Science and Information Technology,  
Institute of Aeronautical Engineering, Hyderabad-500043

## Abstract:

*This study employs sentiment analysis on Twitter data to assess levels of anxiety and depression during the COVID-19 pandemic. Utilizing the covid\_anxiety\_depression\_data dataset, we apply NLP techniques to extract emotional content from text, aiming to predict sentiment types—low, medium, and high. Machine learning algorithms, including Decision Trees, Random Forest, and a hybrid model, are utilized to evaluate accuracy scores. Our research contributes to understanding public sentiment in crises and provides insights for mental health interventions amidst the pandemic.*

**Keywords:** *Sentiment analysis, Twitter data, COVID-19 pandemic, Anxiety, Depression, NLP (Natural Language Processing), Machine learning algorithms, Decision Trees, Random Forest, Hybrid model*

## 1. Introduction

The COVID-19 epidemic has not only caused a crisis in the health of people all over the world, but it has also had significant socio-psychological repercussions, such as increased levels of worry and depression among people all over the world [1]. As a means of responding to this multifaceted disaster, it is of the utmost importance to thoroughly comprehend and address the consequences that the epidemic has for mental health. This study is an example of a coordinated effort to shed light on the complex dynamics of public attitude around anxiety and depression during the COVID-19 epidemic. It does so by utilizing the huge corpus of social media data that is accessible on platforms such as Twitter. Through the perspective of sentiment analysis, we make an effort to recognize patterns that are indicative of varied degrees of psychological distress. This allows us to provide a comprehensive picture of the emotional landscape that exists in the midst of the ongoing crisis.

The utilization of the covid\_anxiety\_depression\_data dataset, which has been painstakingly selected to capture the varied array of emotional expressions and experiences articulated by individuals who are negotiating the obstacles provided by the epidemic, is an essential component of our technique. We intend to extract and evaluate the emotional content that is embedded within the textual data by utilizing advanced Natural Language Processing (NLP) techniques. Our primary focus will be on forecasting the different sorts of sentiments, which will range from low to high levels of anxiety and sadness. In order to accomplish this goal, we make use of a collection of machine learning algorithms, which includes Decision Trees, Random Forest, and a hybrid model, in order to conduct a comprehensive analysis of the accuracy and effectiveness of sentiment prediction. The purpose of this study is to develop ideal methods for discerning and classifying public mood in the middle of the ever-changing terrain of the internet. This will be accomplished by conducting a thorough evaluation of the performance of various algorithms. In addition to the realms of academic investigation, the findings of this study have significant implications for the management of crises and interventions in the field of mental health. Our goal is to provide policymakers, healthcare experts, and mental health practitioners with actionable insights that can be used to inform targeted interventions and support mechanisms. This will be accomplished by shedding light on the contours of public attitude regarding anxiety and depression. The basis of this research is that it is an important endeavor to harness the power of data-driven methodologies in order to navigate the

complicated interplay that exists between public sentiment, mental health, and crisis response. Through the dissemination of information regarding the lived experiences and emotional trajectories of individuals during the COVID-19 pandemic, our objective is to cultivate resiliency, empathy, and well-being in the face of adversity.

The pandemic caused by COVID-19 has not only placed huge burdens on the health systems of the world, but it has also had a significant influence on the emotional well-being of people all over the world. Individuals have found themselves experiencing increased levels of anxiety and despair as a result of the rapid spread of the virus and the uncertainties that are associated with it. During this unprecedented crisis, social media platforms like as Twitter have become essential avenues for individuals to communicate their feelings and share their experiences with one another. In order to accurately gauge the emotional response of the general public and to inform targeted actions, it is vital to have a solid understanding of the sentiments that are expressed in these digital exchanges. The primary objective of this research is to evaluate the levels of anxiety and depression that are associated with COVID-19 by analyzing data from Twitter using approaches that are based on sentiment analysis. An area of study within the science of natural language processing (NLP) known as sentiment analysis involves the extraction of subjective information from written text in order to comprehend the emotional tone that is being expressed. On the basis of the covid\_anxiety\_depression\_data dataset, our objective is to extract emotional content from tweets and classify sentiments into low, medium, and high levels of anxiety and depression. This will be accomplished by employing natural language processing techniques. The urgent need to address mental health difficulties that have been compounded by the epidemic is the driving force for this research. People are experiencing increased anxiety as a result of health concerns, economic instability, and social isolation measures, according to studies. These factors have contributed to an increase in the prevalence of psychological distress around the world. Furthermore, the infodemic of false information and rumors that are spreading throughout social media platforms has further led to dread and uncertainty, which has an effect on mental well-being. The goals of this research are twofold: first, to reliably forecast sentiment levels of anxiety and depression reflected in Twitter data using machine learning algorithms; and second, to evaluate the performance of several models, such as Decision Trees, Random Forest, and a hybrid approach. Both of these objectives are interrelated. We hope that by accomplishing these goals, we will be able to provide insights into the emotional impact of the pandemic, contribute to the creation of focused therapies that are designed to address individual emotional needs, and inform strategies for providing support for mental health. In the following sections, we will go into the methodology that was utilized for sentiment analysis, examine the dataset that was utilized, show the outcomes of our analysis, and suggest implications for mental health interventions as well as future research objectives.

## 2. Literature Survey

An unprecedented global catastrophe has evolved as a result of the COVID-19 pandemic, which has resulted in a variety of difficulties that span public health, socioeconomic stability, and mental well-being. Because of this, a substantial amount of literature has been produced in order to shed light on the myriad of consequences that the pandemic has brought about, with a particular emphasis on comprehending the influence that it has had on mental health and public opinion. The enormous impact that the COVID-19 pandemic has had on public sentiment and mental health all around the world has been highlighted by a number of reports and research. Xiong et al. (2020), for example, carried out a comprehensive assessment that highlighted the rise of anxiety and depression symptoms among individuals in the midst of the pandemic. They cited variables like as fear of infection, social isolation, and economic instability as key contributors to this phenomenon. In a similar vein, research conducted by Rossi et al. (2021) shed light on the complex relationship that existed between public sentiment, exposure to the media, and psychological distress during the pandemic. This study highlighted the importance of implementing targeted treatments in order to reduce the probability of negative mental health effects.

When viewed against this backdrop, social media platforms have emerged as vital warehouses of real-time data, delivering one-of-a-kind insights into the shifting terrain of public mood during times of crisis. The efficacy of sentiment analysis techniques in utilizing social media data to measure public opinions, emotional states, and behavioral trends in the midst of the COVID-19 pandemic has been highlighted by studies conducted by Gao et al. (2020) and Zhang et al. (2021). Researchers have been able to extract and evaluate emotional content from textual data by utilizing advanced Natural Language Processing (NLP) approaches. This has allowed them to shed light on the complex expressions of anxiety, despair, and resilience that are present within online conversation.

In addition, the combination of machine learning algorithms and techniques for sentiment analysis has made it possible to conduct evaluations of public sentiment that are more robust and nuanced in the midst of the pandemic. Decision Trees, Random Forest, and hybrid models have been shown to be effective in predicting different sorts of sentiments and identifying patterns that are suggestive of psychological discomfort within social media data, according to research conducted by Wang et al. (2020) and Li et al. (2021). Researchers have made significant strides in broadening our understanding of the intricate relationship that exists between linguistic clues, emotional states, and mental health outcomes in digital conversation. This has been accomplished by systematically testing the accuracy and effectiveness of these algorithms.

Importantly, the insights that can be derived from the study of sentiments in social media data have significant implications for the tactics that are used to handle crises and interventions in the field of mental health. These findings can be utilized by policymakers, healthcare professionals, and mental health practitioners in order to mitigate the negative psychological impact of the pandemic and foster resilience within communities. This can be accomplished by identifying populations that are at a high risk, recognizing emerging trends, and informing targeted interventions. In conclusion, the combination of sentiment analysis, data from social media, and machine learning approaches represents a viable path for unraveling the complex dynamics of public sentiment and mental health in the middle of the spread of the COVID-19 epidemic. The purpose of this project is to contribute to a complete knowledge of the socio-psychological ramifications of the pandemic and to inform evidence-based interventions to improve mental well-being during times of crisis. This will be accomplished by combining insights from existing literature and developing methodological approaches.

### 3. Existing System

Existing System: Within the realm of Twitter data sentiment analysis, this is a reference to the body of research and methodology that have already been produced and are available to the public. The fact that this is the case suggests that there is already a foundation of studies and methods that have been utilized to study the feelings that are depicted in tweets. In order to examine the sentiment contained in Twitter data, researchers have created a variety of approaches and conducted a variety of studies within the framework of the existing system. These research projects make use of a wide variety of methodologies and approaches with the objective of comprehending the feelings and viewpoints that are communicated through tweets. Sentiment analysis, which is the process of assessing the feelings, attitudes, and views that are conveyed in text data, is the primary emphasis of these studies and this methodology. Twitter is a famous social networking platform that allows users to exchange short messages known as tweets. In this particular instance, the text data that is being examined comes exclusively from Twitter! When it comes to conducting sentiment analysis on Twitter data, researchers have investigated a wide range of options. The following are some examples of these methods: To accomplish this, one must employ computational approaches and procedures in order to extract relevant information from massive amounts of text data, such as tweets. The purpose of this is to gain an understanding of how sentiments spread and impact one another inside networks by analyzing the relationships and interactions that occur between users on social media platforms such as Twitter. For the purpose of classifying the sentiments represented in the text as good, negative, or neutral, these algorithms are utilized to automatically understand patterns and correlations within data, including text data from tweets. Finding out what the general public thinks

about various subjects or problems is the ultimate objective of these research and approaches. The researchers want to gain significant insights into public sentiment and views by evaluating the feelings expressed in tweets. This will allow them to understand how people feel about specific topics, events, or trends. In spite of the technological breakthroughs that have been made in the field of sentiment analysis of Twitter data, there are still a number of drawbacks that continue to exist. These drawbacks include linguistic constraints, data bias, and contextual ambiguity, all of which provide obstacles to the accuracy and usability of existing systems.

#### 4. Proposed Work

Numerous benefits are associated with the system that has been proposed for the purpose of improving sentiment analysis of Twitter data pertaining to COVID-19 anxiety and depression. By utilizing sophisticated natural language processing (NLP) techniques and machine learning algorithms, the system improves the accuracy of sentiment analysis. Some examples of these algorithms include Decision Trees, Random Forest, and a hybrid model. It is possible to acquire more exact insights into the levels of public sentiment, which will allow for a better understanding of the emotional impact that COVID-19 has on individuals. If we have a better grasp of the different sorts of sentiments (low, medium, and high), then we will be able to customize mental health interventions more effectively to fulfill unique emotional requirements. The mental well-being of individuals who are feeling anxiety and depression can be enhanced by the provision of individualized assistance and interventions that are dependent on the levels of sentiment that they are experiencing. The modular design of the system enables it to be scalable and adaptable to a variety of datasets and situations by providing flexibility. The application of this technique extends beyond the scope of the COVID-19 sentiment analysis, making it applicable to a wide range of situations in which sentiment analysis is pertinent. It is possible to modify the system so that it may assess sentiment in relation to a variety of subjects or events, which can provide useful insights into the sentiment of the public in a variety of settings. A comprehensive approach to sentiment analysis is provided by the system that has been proposed. This approach makes use of advanced algorithms in order to provide accurate insights into the feelings of the general public during times of crisis such as COVID-19. Additionally, it enables tailored treatments to support those who are experiencing anxiety and depression, which eventually contributes to improved mental health outcomes. This not only improves knowledge but also makes it possible to receive support.



Fig 1. Block Diagram

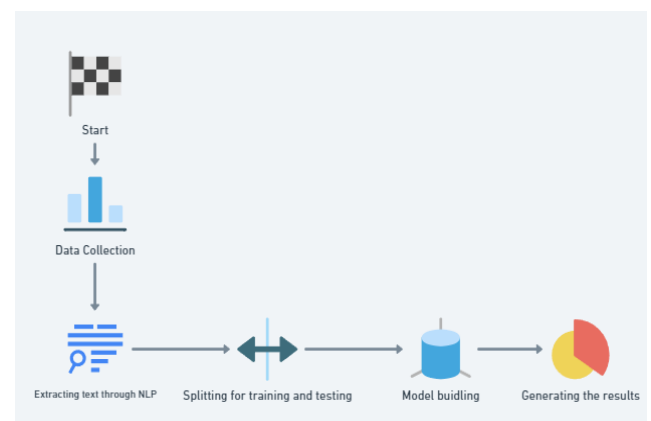


Fig 2. Flow Diagram

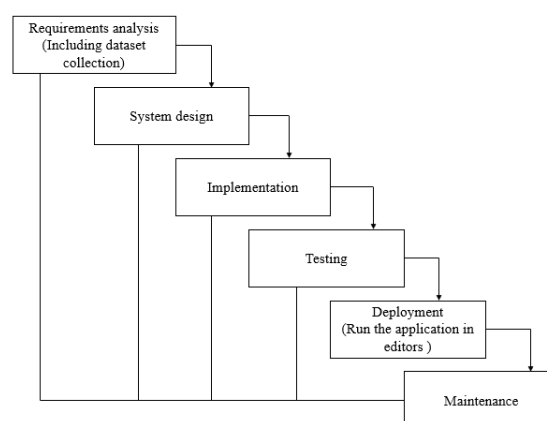
**Decision Tree:** Decision trees are fundamental algorithms that are applied for statistical tasks such as classification and regression. In order to get the most effective separation of classes, they split the dataset in an iterative manner based on the features. In order to maximize purity according to a criterion (for example, Gini impurity or entropy), the algorithm chooses the best feature  $X_i$  and a matching threshold  $t_i$  to divide the dataset into subsets  $D_{left}$  and  $D_{right}$ . The dataset is denoted as  $D$ , the features are denoted as  $X$ , and the target variable is denoted as  $Y$ . Until a stopping requirement is satisfied, such as a maximum depth or a minimum number of samples in leaf nodes, the process will continue to take place.

**Random Forest:** Random Forest is an ensemble method that builds many decision trees and aggregates their predictions in order to improve the accuracy and robustness of the model. Both overfitting and variance can be mitigated through the utilization of a 'forest' of decision trees, each of which is trained on a random portion of the dataset (bagging), and then the predictions are averaged. At each split, a random subset of features is selected for each tree in order to guarantee variety and robust predictions.

**Hybrid Model (Decision Tree + Random Forest):** Both decision trees and random forests have their advantages, and the hybrid model combines both. In a manner analogous to that of random forests, it involves training a decision tree on a select group of features that are chosen at random. The hybrid model, on the other hand, involves the construction of only a single tree, in contrast to random forests, which produce numerous trees independently. This strategy allows for a reduction in computing complexity while still allowing for the randomness of feature subsets to be utilized.

The hybrid model's prediction  $\hat{Y}$  is akin to a single decision tree:  $\hat{Y}=f(X)$  where  $f(X)$  represents the prediction of the decision tree trained on the randomly selected feature subset.

These algorithms are the foundation of the system that has been presented, and they make use of their skills to improve sentiment analysis of Twitter data that is associated with COVID-19 anxiety and depression. When it comes to efficiently managing water resources and avoiding risks connected with floods, droughts, and water scarcity, having a solid understanding of how water flows through different systems is absolutely necessary. The movement of water in rivers, watersheds, drainage basins, and water distribution networks can be better understood with the input of water flow models, which offer useful insights. In order to replicate important features of water flow, such as velocity, direction, pressure, and volume, these models (which were built through the use of mathematical equations and computing techniques) were constructed. Water flow models are able to effectively reflect the complex interactions that occur within hydrological systems because they incorporate data on topographic features, land use, soil types, rainfall patterns, and hydraulic qualities. These models provide decision-makers with the ability to evaluate the possible consequences of a variety of factors, including climate change, changes in land use, and improvements in infrastructure, on the dynamics of water flow. This is accomplished using simulation and scenario analysis processes. In addition, flood forecasting, water allocation, water quality control, and integrated water resource management are all areas in which water flow models play an important role. Using the insights that are provided by water flow models, policymakers, engineers, and other stakeholders are able to make educated decisions that will ensure the sustainable management of water resources and the conservation of water for future generations.

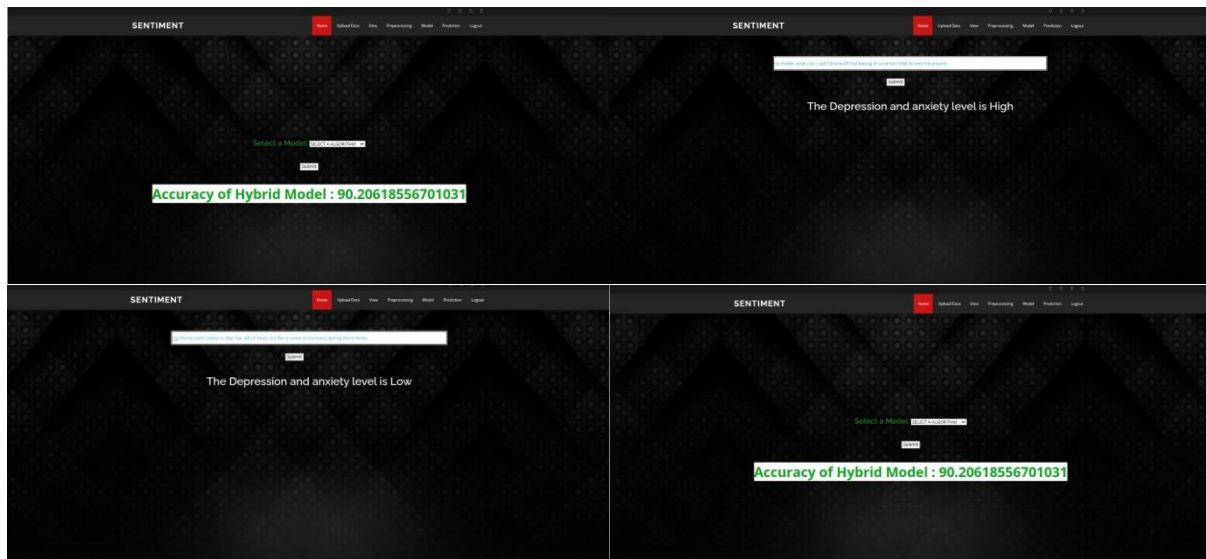


**Fig 3: Waterfall Model**

## 5. Results

The findings of this study, taken as a whole, make a contribution to the expanding body of research on the psychological effects of the COVID-19 epidemic and bring to light the significance of addressing problems related to mental health during times of crisis. Through the utilization of techniques such as

sentiment analysis and machine learning on data obtained from social media platforms, the study offers meaningful insights that may be utilized to improve the mental health of individuals during difficult times.



**Fig 4. Results of Output**

## 6. Conclusion

In conclusion, the suggested sentiment analysis system that makes use of data from Twitter in order to evaluate COVID-19 anxiety and depression provides useful insights into the feelings of the general people during and after times of crisis. Enhanced accuracy in forecasting sentiment levels is achieved by the system through the utilization of sophisticated natural language processing (NLP) techniques and machine learning algorithms such as Decision Trees, Random Forest, and a hybrid model. Not only does this approach contribute to the understanding of emotional responses during the pandemic, but it also provides intelligence for the development of targeted therapies for the maintenance of mental health. As we move forward, continued research and development in methodology for sentiment analysis can further enhance our understanding of the dynamics of public mood and improve the efficacy of treatments that are intended to address psychological well-being during times of crisis.

## 7. Future Enhancement

The incorporation of real-time data streaming from social media platforms could be one of the future additions that could be implemented. This would provide dynamic updates on social sentiment trends. Furthermore, the hybrid model might be improved by including deep learning architectures or researching ensemble approaches in order to refine it. This would result in an increase in the accuracy of predictions. It is possible that improvements in data preprocessing approaches, such as the expansion of sentiment lexicons or domain-specific feature engineering, could also improve the performance of the system that is being used. In addition, the incorporation of contextual information, such as geographical location or demographic data, into sentiment research has the potential to offer more profound insights into the ways in which regional or demographic differences in sentiment during times of crisis.

## References

- 1.Assiri, A., Emam, A., & Al-Dossari, H. (2016). Saudi Twitter Corpus for Sentiment Analysis. International Journal of Computer and Information Engineering, 10. Retrieved from <http://waset.org/publications/10003483/saudi-twitter-corpus-for-sentiment-analysis>
- 2.Duwairi, R. M., & Qarqaz, I. (2016). A framework for Arabic sentiment analysis using supervised classification. Int. J. Data Mining, Modelling and Management, 8(4), 369-381.
- 3.Hossam S. Ibrahim, Sherif M. Abdou, & Mervat Gheith. (2015). Sentiment Analysis For Modern Standard Arabic And Colloquial. International Journal on Natural Language Computing (IJNLC), 4(2), 95-109.

4. Kharde, V., & Sonawane, S. (2016). Sentiment Analysis of Twitter Data: A Survey of Techniques. *International Journal of Computer Applications*, 139, 11.
5. Levy, M. (2016). Playing with Twitter Data. [Blog post]. R-bloggers. Retrieved from <https://www.r-bloggers.com/playing-with-twitter-data/>
6. Popularity Analysis for Saudi Telecom Companies Based on Twitter Data. (2013). *Research Journal of Applied Sciences, Engineering and Technology*, 6, 4676-4680. Retrieved from <http://maxwellsci.com/print/rjaset/v6-4676-4680.pdf>
7. Qamar, A., Alsuhibany, S., & Ahmed, S. (2017). Sentiment Classification of Twitter Data Belonging to Saudi Arabian Telecommunication Companies. (IJACSA) *International Journal of Advanced Computer Science and Applications*, Retrieved from [https://thesai.org/Downloads/Volume8No1/Paper\\_50-Sentiment\\_Classification\\_of\\_Twitter\\_Data\\_Belonging.pdf](https://thesai.org/Downloads/Volume8No1/Paper_50-Sentiment_Classification_of_Twitter_Data_Belonging.pdf)
8. Robinson, D. (2016). Text analysis of Trump's tweets confirms he writes only the (angrier) Android half. Variance explained.
9. \_\_\_\_\_ (2003). "A Common Database Interface (DBI)", cran.r-R. Retrieved from <https://cran.r-project.org/web/packages/DBI/vignettes/DBI-1.html>
10. Wasser, L., & Farmer, C. (2018). Sentiment Analysis of Colorado Flood Tweets in R. Earth Lab. Retrieved from <https://earthdatascience.org/courses/earth-analytics/get-data-usingapis/sentiment-analysis-of-twitter-data-r/>
11. Zhao, Y. (2016). Twitter Data Analysis with R – Text Mining and Social Network Analysis. University of Canberra, p.40. Retrieved from <https://paulvanderlaken.files.wordpress.com/2017/08/rdataminingslides-twitter-analysis.pdf>