Sentiment Analysis of Social Media during Coronavirus and Its Correlation with Indian Stock Market Movements

Ahmad Amjad Mir

University of Wisconsin – Madison, USA

Corresponding Email: Ahmadamjadmir@gmail.com

Abstract:

The COVID-19 pandemic has deeply influenced our emotional state, leading to significant volatility in the Indian stock market throughout the year. Emotions have become a key factor in driving market movements, with the pandemic amplifying their effects. This research investigates the effectiveness of various machine learning algorithms in analyzing and predicting sentiment within the Indian stock market during the pandemic. It introduces a framework that assesses six different algorithms, identifying the ones with the highest accuracy. These findings could be instrumental in creating more robust prediction models moving forward.

Keywords: Prediction models, algorithm performance, market volatility, public sentiment, financial forecasting, emotional impact.

1. Introduction:

The COVID-19 pandemic has introduced unprecedented economic uncertainty, with global financial markets experiencing significant volatility. In India, the pandemic's impact on the stock market has been particularly notable, with dramatic fluctuations reflecting the broader uncertainties of the crisis[1]. As financial markets navigate the changing landscape, investors and analysts are increasingly seeking alternative data sources to better understand market trends. Social media platforms, in particular, have become key sources for capturing public sentiment and opinion[2].

Social media has become a crucial channel for disseminating information and expressing public sentiment during the pandemic. Social Media Platforms provide real-time updates and reflections on pandemic-related events, government policies, and economic conditions[3]. This constant stream of information provides a unique opportunity to analyze how public sentiment evolves in response to pandemic developments and how these sentiments may influence financial markets[4].

This research paper explores the relationship between social media sentiment during the COVID-19 pandemic and movements in the Indian stock market. By applying sentiment analysis techniques to social media data, this study investigates whether changes in public sentiment correlate with stock market fluctuations. Understanding this relationship can offer valuable insights into how sentiment drives market behavior, particularly during times of crisis. The results of this study offer valuable insights for investors, policymakers, and researchers looking to utilize social media data in financial forecasting and decision-making.

2. Literature Review:

The interplay between social media sentiment and financial markets has garnered considerable attention in recent research[5]. Numerous studies have established that social media platforms serve as valuable indicators of public sentiment, which can significantly impact market behavior. Studies have demonstrated that sentiment on Twitter may be used to forecast changes in the stock market, highlighting the importance of social media data in forecasting financial outcomes. The investigation indicated how public opinion fluctuations, as seen on social media, frequently precede changes in the market, providing early warning signs of market moves that may stimulate and motivate additional research [6].

Analysis of sentiment uses a variety of machine learning and natural language processing methods to examine data from social media platforms. Methods like subject modeling and sentiment classification are commonly employed to measure public sentiment and its influence on financial markets. For example, research has demonstrated that deep learning models for sentiment analysis can improve the precision of stock market forecasts by identifying subtle changes in sentiment from vast amounts of social media data.[6].

The COVID-19 pandemic has amplified the relevance of sentiment analysis in financial markets due to heightened public concern and increased social media activity. Social media sentiment became more volatile and closely linked to market movements[7]. This period of heightened uncertainty saw an increase in the volume and intensity of sentiment expressed online, reflecting the growing influence of social media on financial markets.

Additionally, studies on the Indian stock market during the pandemic have revealed that sentiment analysis can provide valuable insights into market dynamics[8]. A research explored the impact of COVID-19 related sentiment on Indian stock market indices, finding significant correlations

between sentiment trends and market behavior. Their work underscores the importance of better integrating sentiment analysis into financial models. This integration can provide policymakers with a more comprehensive understanding of market responses to pandemic-related developments[9].

This literature review provides a foundation for the current study by illustrating the established relationship between social media sentiment and financial markets, particularly during pandemic. It also highlights the methodological advancements in sentiment analysis and the specific relevance of these techniques to the Indian stock market during this crisis.

3. Methodology:

With the proliferation of data sources for collecting text messages, choosing the most relevant platforms for sentiment analysis is essential. For this study, four primary sources were identified: RSS feeds, which provide updates from various news sources; forum discussions, where individuals share their opinions and experiences; Twitter, a social media platform known for its real-time updates and public discussions; and news portals[10][10], which offer comprehensive coverage of current events.

The sentiment assessment has been done on the obtained text data using sophisticated natural language processing (NLP) methods. To identify whether the text was neutral, negative, or positive in sentiment, we used a number of machine learning techniques. These algorithms included support vector machines (SVM), a supervised learning model with associated learning algorithms that analyze data for classification and regression analysis; deep learning models like recurrent neural networks (RNNs) and transformers; and logistic regression, a statistical method for analyzing a dataset in which one or more independent variables determine an outcome. The selection of algorithms was made with the intention of capturing the subtleties of sentiment conveyed in various social media and news material categories [11].

The Bag of Words (BoW) method is fundamental to text analysis and natural language processing. It represents text data as a collection of words without considering their order or context. In this method, a text document is converted into a vector of word frequencies or binary indicators, where each word in the document's vocabulary is assigned a unique index. The resulting vector reflects the presence, absence, and sometimes frequency of each word in the document[12][12]. Despite its simplicity, BoW is effective for many text classification tasks, such as sentiment analysis, by providing a straightforward numerical representation of textual data.

This method is particularly useful in our research as it allows us to analyze large social media data efficiently and accurately.

The N-grams method is used in text analysis to extend the Bag of Words approach by considering sequences of words, or "n-grams," rather than individual words. An N-gram is a contiguous

sequence of 'n' words from a given text, where 'n' can be any positive integer. For example, in the phrase "financial markets are volatile," bigrams (2-grams) would be "financial markets," "markets are," and "are volatile." This method captures contextual relationships between words, providing a more nuanced text representation. By incorporating these word sequences, the N-grams method can improve the performance of text classification and sentiment analysis tasks by accounting for word order and co-occurrence patterns[13].

Formula:

$$N - grams = count\,(w2\,w1)\,/\,count\,(w2)$$

A commonly used approach in text mining and information retrieval for determining the relevance of words in a document in relation to a set of documents is the TF-IDF (Term Frequency-Inverse Document Frequency) model. The TF component calculates a word's frequency in a given document, indicating how important it is to that document. By taking into consideration the word's frequency in each document in the corpus, the IDF component modifies this metric and lessens the weight of popular terms that crop up frequently in a large number of documents. Combining these two metrics allows TF-IDF to be unique within the corpus and highlight significant terms inside a document, making it an effective tool for information retrieval and document classification tasks[14].

4. Suggested Model:

 Data is collected from four distinct sources: forum discussions, Twitter, news articles, and RSS feeds. This diverse text data undergoes a thorough pre-processing stage, which involves removing punctuation and stop words to clean the text. Additionally, tokenization and lemmatization are performed to standardize the text and reduce words to their base forms[15]. These steps ensure that the data is prepared for accurate analysis by reducing noise and variability[16].

Following pre-processing, the cleaned data is analyzed using three advanced natural language processing (NLP) techniques. A set is created for each of these techniques to capture the relevant textual information. After that, these feature sets are subjected to six machine learning algorithms to assess how well they classify sentiment into Positive, Negative, and Neutral categories. The overall framework for this sentiment analysis and prediction process is depicted in Fig.1:
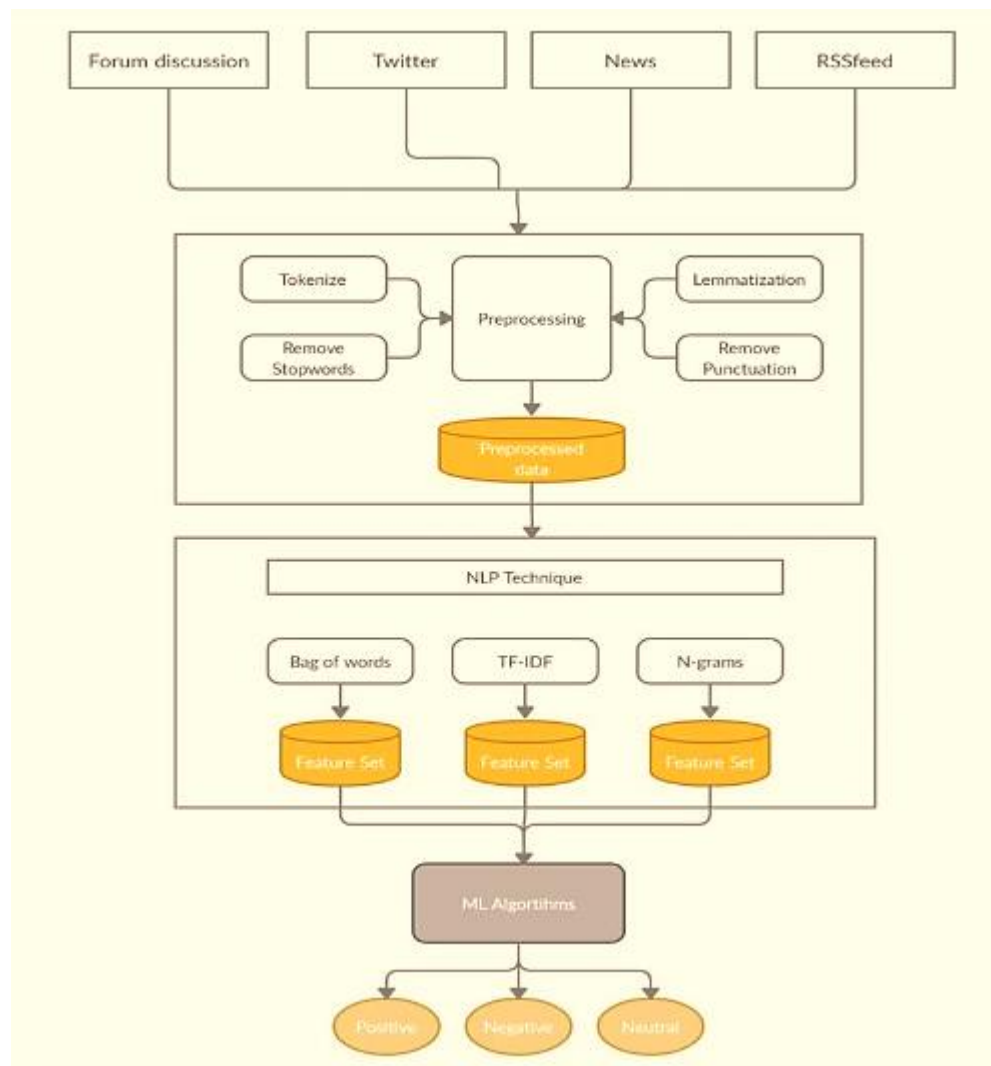
Fig.1: Suggested framework for sentiment analysis.

The suggested framework for sentiment analysis and prediction combines multiple NLP techniques and machine learning algorithms to classify sentiment from text data[17] accurately. It leverages Bag of Words, TF-IDF, and N-grams to generate feature sets, which are then analyzed using various algorithms to determine sentiment categories: Positive, Negative, or Neutral[18].

5. Results:

This research aims to determine the most effective approach for sentiment analysis and prediction concerning news about the Indian stock market, particularly during the COVID-19 pandemic[19]. The study compares various methods to evaluate their performance in classifying text into one of three sentiment categories: positive, negative, or neutral. By testing six algorithms with two different techniques, this research assesses their accuracy in categorizing sentiments.

Fig. 2 illustrates the data from January 1, 2020, to August 24, 2020. It shows the fluctuations in sentiment over this period, highlighting the variations and trends in sentiment movements.
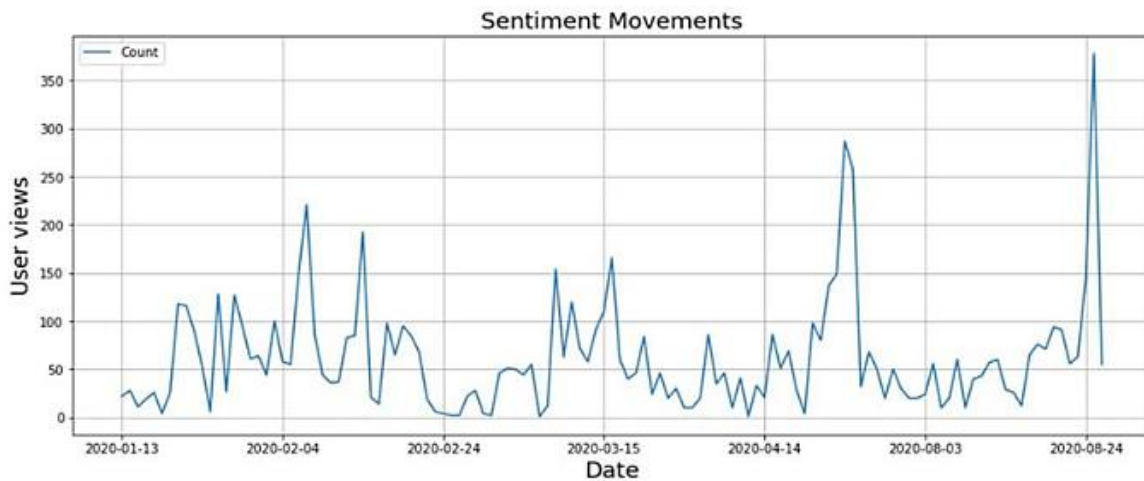
Fig.2: Sentiment Graph

This graph illustrates how sentiment fluctuated over the period, capturing shifts in public opinion and emotional responses to events as they unfolded. The visual representation highlights periods of heightened positivity and negativity, reflecting the broader impact of significant events and market dynamics during this time frame[20].

6. Future Directions:

Future studies in sentiment analysis for financial markets could investigate several promising avenues. One potential direction is the incorporation of advanced deep learning methods, such as transformers and attention mechanisms, to improve the precision and contextual understanding of sentiment predictions [21]. Broadening the dataset to encompass a variety of sources, including new social media platforms and multilingual content, could offer a more thorough perspective on market sentiment. Moreover, integrating real-time data and sentiment analysis into automated trading systems might enhance the responsiveness and adaptability of financial strategies. Examining the relationship between sentiment and other economic indicators, such as geopolitical events or macroeconomic factors, could also provide valuable insights into market dynamics. These advancements have the potential to refine predictive models and deepen our understanding of sentiment's impact on financial decision-making [22].

7. Conclusion:

This study highlights the crucial role of sentiment analysis in comprehending and forecasting movements in the Indian stock market, especially during the COVID-19 pandemic. By evaluating different NLP techniques and machine learning algorithms, the research has pinpointed effective

methods for improving sentiment classification accuracy. These findings emphasize the importance of integrating sentiment analysis with financial forecasting, revealing how public opinion and emotional responses can shape market behavior. As financial markets continue to develop, adopting advanced sentiment analysis methods and broadening data sources will enhance prediction precision and offer more valuable tools for investors and analysts.

References:

[1]   B. A. Abdelfattah, S. M. Darwish, and S. M. Elkaffas, "Enhancing the Prediction of Stock Market Movement Using Neutrosophic-Logic-Based Sentiment Analysis," Journal of Theoretical and Applied Electronic Commerce Research, vol. 19, no. 1, pp. 116-134, 2024.

[2]   P. Ray, B. Ganguli, and A. Chakrabarti, "Multivariate Bayesian Time-Series Model with Multi-temporal Convolution Network for Forecasting Stock Market During COVID-19 Pandemic," International Journal of Computational Intelligence Systems, vol. 17, no. 1, p. 170, 2024.

[3]   M. Bhardwaj, P. Mishra, S. Badhani, and S. K. Muttoo, "Sentiment analysis and topic modeling of COVID-19 tweets of India," International Journal of System Assurance Engineering and Management, vol. 15, no. 5, pp. 1756-1776, 2024.

[4]   D. Saravagi, S. Agrawal, and M. Saravagi, "Indian stock market analysis and prediction using LSTM model during COVID-19," International Journal of Engineering Systems Modelling and Simulation, vol. 12, no. 2-3, pp. 139-147, 2021.

[5]   N. B. Sing and R. G. Singh, "Investor attention and reaction in COVID-19 crisis: Sentiment analysis in the Indian stock market," Managerial Finance, vol. 49, no. 3, pp. 470-491, 2023.

[6]   P. M. Chandana, S. Mythri, S. F. Ahmad, and S. Siddiqui, "An Analytical Study of Sentimental Analysis During COVID-19 Pandemic," Advances in Computational Intelligence and Its Applications, p. 10, 2024.

[7]   R. Chandra and A. Krishna, "COVID-19 sentiment analysis via deep learning during the rise of novel cases," PloS one, vol. 16, no. 8, p. e0255615, 2021.

[8]   S. Singh Chauhan, K. Srinivasan, and T. Sharma, "A trans-national comparison of stock market movements and related social media chatter during the COVID-19 pandemic," Journal of Business Analytics, vol. 6, no. 3, pp. 203-216, 2023.

[9]   N. Chintalapudi, G. Battineni, and F. Amenta, "Sentimental analysis of COVID-19 tweets using deep learning models," Infectious disease reports, vol. 13, no. 2, pp. 329-339, 2021.

[10]   M. Costola, O. Hinz, M. Nofer, and L. Pelizzon, "Machine learning sentiment analysis, COVID-19 news and stock market reactions," Research in International Business and Finance, vol. 64, p. 101881, 2023.

[11]   N. Das, B. Sadhukhan, T. Chatterjee, and S. Chakrabarti, "Effect of public sentiment on stock market movement prediction during the COVID-19 outbreak," Social network analysis and mining, vol. 12, no. 1, p. 92, 2022.

[12]   P. Eachempati, P. R. Srivastava, and P. K. Panigrahi, "Sentiment analysis of COVID-19 pandemic on the stock market," American Business Review, vol. 24, no. 1, p. 8, 2021.

[13]   J. D. Fernandez and A. Kumar, "A machine learning perspective of the impact of COVID-19 on the Indian stock market," International Journal of Business Intelligence and Systems Engineering, vol. 2, no. 1, pp. 1-22, 2024.

[14]　I. Ghosh, E. Alfaro-Cortés, M. Gámez, and N. García-Rubio, "Role of proliferation COVID-19 media chatter in predicting Indian stock market: Integrated framework of nonlinear feature transformation and advanced AI," Expert Systems with Applications, vol. 219, p. 119695, 2023.

[15]　C. Gondaliya, A. Patel, and T. Shah, "Sentiment analysis and prediction of Indian stock market amid Covid-19 pandemic," in IOP conference series: materials science and engineering, 2021, vol. 1020, no. 1: IOP Publishing, p. 012023.

[16]　B. Singh, R. Dhall, S. Narang, and S. Rawat, "The outbreak of COVID-19 and stock market responses: An event study and panel data analysis for G-20 countries," Global Business Review, vol. 25, no. 3, pp. 606-631, 2024.

[17]　P. R. Jena and R. Majhi, "Are Twitter sentiments during COVID-19 pandemic a critical determinant to predict stock market movements? A machine learning approach," Scientific African, vol. 19, p. e01480, 2023.

[18]　M. VISHWAKARMA, V. SHUKLA, and P. KHOT, "BERT-Based Sentiment Analysis of Indian COVID-19 Tweets for Policy Making," 2024.

[19]　A. G. Katsafados, S. Nikoloutsopoulos, and G. N. Leledakis, "Twitter sentiment and stock market: a COVID-19 analysis," Journal of Economic Studies, vol. 50, no. 8, pp. 1866-1888, 2023.

[20]　H. S. Lee, "Exploring the initial impact of COVID-19 sentiment on US stock market using big data," Sustainability, vol. 12, no. 16, p. 6648, 2020.

[21]　P. Mehta, S. Pandya, and K. Kotecha, "Harvesting social media sentiment analysis to enhance stock market prediction using deep learning," PeerJ Computer Science, vol. 7, p. e476, 2021.

[22]　M. Radulescu, A. Rao, B. Doğan, and S. Abbas, "Co-movement between COVID-19, oil price and American stock market during 2020: Fresh investigation from partial and multiple wavelet methods," Resources Policy, vol. 95, p. 105194, 2024.