International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Sentiment Analysis In Financial Markets: Predicting Stock Trends Based On Social Media Sentiment

Gundra Kowshik

B. tech, GMR IT, Rajam 532127, India

ABSTRACT:

Sentiment analysis in financial markets involves using natural language processing (NLP) to gauge the emotional tone of social media posts and other online content related to companies and the economy. By identifying whether sentiments are positive, negative, or neutral, analysts can assess public opinion on specific stocks or market trends. The sentiment data is then used to predict potential movements in stock prices, helping investors make informed decisions based on the collective mood of the market. Present work looks at how emotions expressed on social media could be used to forecast changes in the stock market. Sites like Reddit and Twitter are full of thoughts that people post on businesses and the state of the economy in general. The sentiment, whether good, negative, or neutral, can be ascertained by attentively examining these posts using a technique known as (NLP). The connection between these attitudes and variations in stock prices is then investigated. The usefulness of sentiment analysis in tracking stock trends is evaluated by reviewing additional research. Problems like figuring out how accurate the data is and how quickly the markets react to new information are also taken into account. Furthermore, the possibility of enhancing forecasts by merging sentiment analysis and conventional financial metrics is explored.

Keywords: Social Media Sentiment, Stock Market Trends, Sentiment Analysis, Natural Language Processing, Financial Indicators, Market Prediction, Investment Strategies.

INTRODUCTION:

Predicting stock prices has always been a challenging task for financial experts due to the unpredictable nature of the market. In recent years, efforts have been made to enhance stock price predictions using machine learning and artificial intelligence techniques. The use of sentiment analysis has gained attention, where public opinions expressed on social media platforms, such as Twitter and Stock Twits, are analyzed. It is believed that these sentiments have a significant influence on stock price movements, as they often reflect the public's expectations of the market.

In several studies, social media data has been collected and processed using natural language processing (NLP) tools to extract sentiments from the text. The sentiments are then classified as positive, negative, or neutral, which are used to predict whether a stock price will rise or fall. Machine learning algorithms such as Support Vector Machine (SVM), Naive Bayes, Logistic Regression, and Random Forest have been employed to build models for these predictions.

One of the major findings from the research is that social media sentiment is positively correlated with stock price movements. Studies have shown that using sentiment data in combination with traditional stock market data can improve prediction accuracy. However, challenges such as data noise, fake reviews, and the lack of sentiment labels for all tweets have been highlighted.

Additionally, preprocessing of social media data, such as removing irrelevant words, normalizing the text, and converting sentiments into numerical scores, is a critical step in ensuring the accuracy of the models. By leveraging sentiment analysis, stock market predictions have been enhanced, proving that machine learning techniques and public sentiment can offer valuable insights into market trends. This approach provides a more comprehensive method for forecasting stock prices and is a growing field of research.

LITERATURE SURVEY:

1. This study develops a predictive method using Long Short-Term Memory (LSTM) networks for stock prices, integrating social media sentiment (from Twitter) with historical data. Sentiment analysis enhances prediction, yielding improved accuracy compared to traditional methods.
2. Focused on sentiment analysis from Twitter data, this work predicts stock prices for Dow Jones companies. By leveraging ensemble methods and LSTM models, it enables better stock market trend analysis, democratizing insights for non-experts.
3. A novel Transductive LSTM (TLSTM) model is proposed to integrate sentiment analysis with time-sensitive stock data. This approach improves prediction for underrepresented sentiments using the Off-policy Proximal Policy Optimization (PPO) algorithm, achieving higher accuracy for short-term forecasts.

4. Employing Random Forest classifiers, this paper analyzes the impact of combining social media sentiment and financial news with market data. It demonstrates the classifier's effectiveness through feature selection and reducing noise, leading to accurate predictions.
5. This work uses Naïve Bayes, SVM, and Logistic Regression for sentiment analysis on Stock Twits data, correlating daily sentiment polarity with stock price movements. The inclusion of historical data enhances prediction precision.
6. Integrating political events from Wikipedia with sentiment and historical stock data, the authors use ten machine learning algorithms, including Random Forest and Naïve Bayes, to predict trends. The study identifies external factors' impact on accuracy.
7. A Least Squares Support Vector Regression (LSSVR) model combines sentiment analysis from Twitter with stock data to predict vehicle sales, highlighting its adaptability for cross-domain market predictions.
8. Leveraging a hybrid genetic algorithm (HGA) with LSTM, this research focuses on stock trends in Taiwan, combining chip-based indicators and social network sentiment for improved predictions.
9. A review highlights practical applications of sentiment analysis in stock prediction, discussing the evolution of Transformer models like BERT for extracting sentiment features more effectively.
10. This paper employs Latent Dirichlet Allocation (LDA) for sentiment classification in English and Persian. Enhanced with part-of-speech tagging, the model outperforms traditional sentiment techniques, especially for low-resource languages.
11. Challenges in applying sentiment analysis to stock prediction are reviewed, emphasizing issues like data preprocessing and the need for scalable algorithms capable of capturing real-world sentiment trends.
12. Using deep learning models, this study analyzes sentiment on Stock Twits. The application of doc2vec and CNNs demonstrates superior performance over traditional models for stock movement forecasting.
13. Analyzing sentiment from Portuguese tweets during Brazil's elections, the authors use machine learning techniques to establish correlations between public mood and stock price fluctuations.
14. This comparative study examines the role of sentiment polarity during pandemics. Using VADER and lexicon-based methods, it identifies correlations between social sentiment shifts and financial market behaviors.
15. News sentiment analysis with VADER outperforms traditional models like SVM and Logistic Regression for intraday stock trends, showcasing the importance of fine-tuning lexicons for financial language.

METHODOLOGY:

The study aims to predict stock trends by leveraging social media sentiment analysis in financial markets. Sentiments expressed in posts from platforms like Twitter and Reddit are classified as positive, negative, or neutral using advanced models such as Transductive Long Short-Term Memory (TLSTM), Long Short-Term Memory (LSTM) with Ensemble Techniques, Random Forest, and Support Vector Machine (SVM). By integrating sentiment data with historical stock prices, trading volume, and technical indicators, the model enhances the accuracy of stock price predictions.

To achieve this, data is collected from social media platforms with timestamps and stock symbols, alongside historical stock market data. Sentiment data is pre-processed through noise removal, text tokenization, and sentiment scoring, while stock data is normalized and missing values are managed. Finally, sentiment and stock data are aligned using timestamps to ensure accurate predictions.

1. Transductive Long Short-Term Memory (TLSTM)

TLSTM is an advanced variant of the traditional LSTM that is designed to enhance prediction accuracy, particularly in time-sensitive scenarios such as stock market prediction. Unlike standard LSTMs, which treat all data equally over time, TLSTM focuses on recent data points, giving more weight to the most current information. This approach makes TLSTM well-suited for capturing short-term trends in stock prices, which can be highly influenced by social media sentiment during volatile periods.

- **Advantages:** TLSTM's ability to focus on recent trends makes it ideal for tracking stock market movements that are directly influenced by news and sentiment, such as political events or economic news. By dynamically adjusting its focus to the most recent data, TLSTM can better capture short-term shifts in the market.

- **Challenges:** TLSTM requires substantial computational power and large datasets to achieve reliable results. Additionally, by emphasizing recent data, TLSTM may underperform in scenarios where long-term trends are more significant. It may also overfit to short-term noise, especially in highly volatile market conditions.

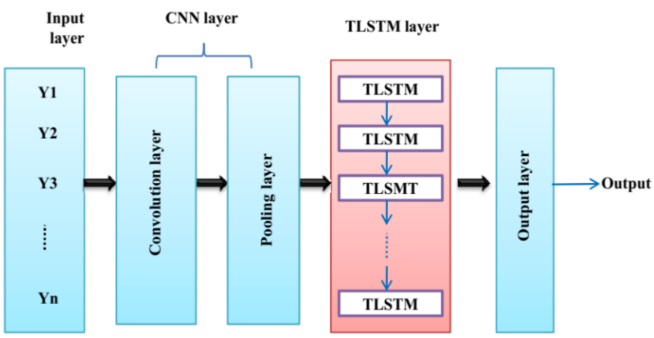

Figure 1. Architecture of TLSTM

2. Long Short-Term Memory (LSTM) with Ensemble Techniques:

LSTM models are widely used for sequence prediction tasks due to their ability to capture long-term dependencies in data. By combining multiple LSTM models using ensemble techniques, we can create a more robust predictive system that balances short-term and long-term trends.

- **Ensemble of Multiple LSTMs:** The multi-LSTM model uses several LSTM networks, each trained to capture different aspects of the time series data. These models are then combined, usually through methods like weighted averaging or stacking, to generate more accurate predictions.
- **Integration Techniques:** Methods such as Weighted Average and Differential Evolution are used to integrate the outputs from different LSTM models. These techniques help balance the strengths of individual models, improving overall prediction accuracy.
- **Advantages:** By combining multiple LSTM models, this ensemble approach increases the stability and accuracy of predictions, making it suitable for longer-term stock price forecasting.
- **Challenges:** Ensemble models can be computationally expensive and may require fine-tuning to avoid overfitting. Additionally, the complexity of these models increases the need for substantial data and computational resources.

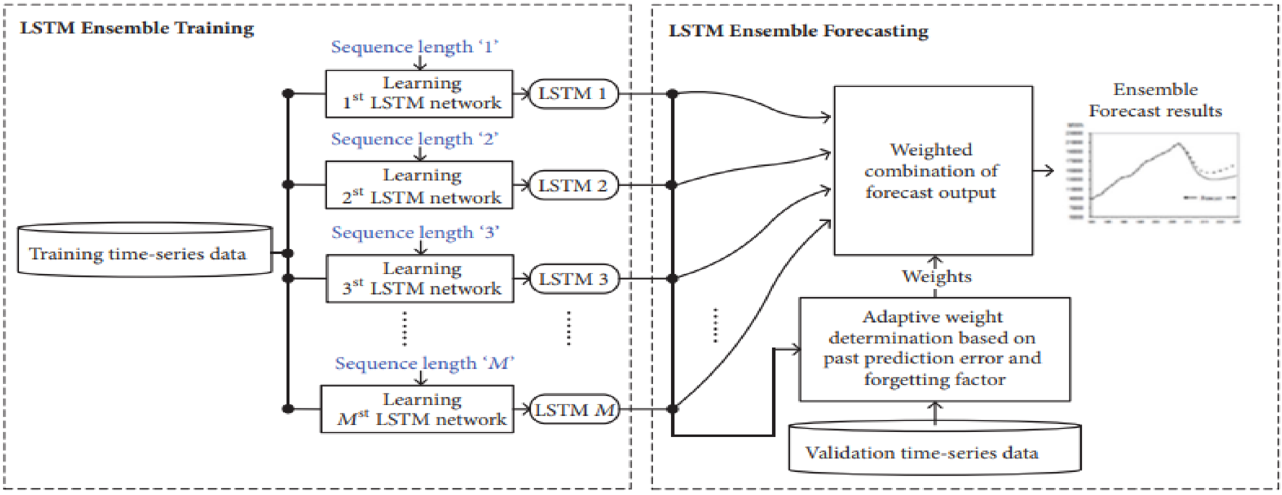

Figure 2. Architecture of LSTM with Ensemble Techniques

3. Random Forest Algorithm with Sentiment Analysis

Random Forest is a powerful machine learning algorithm that utilizes an ensemble of decision trees to improve prediction accuracy. By integrating sentiment analysis scores from social media posts, Random Forest can provide nuanced predictions that account for both sentiment and historical market data.

- **Multiple Decision Trees:** Random Forest constructs multiple decision trees, each of which makes a prediction based on a random subset of the features. The final prediction is based on the majority vote across all decision trees, which helps reduce overfitting and improves the generalization of the model.
- **Integration with Sentiment Analysis:** By incorporating sentiment scores from platforms like Twitter and Reddit, Random Forest can capture the mood of the market and use it as an additional feature for predicting stock trends.
- **Advantages:** Random Forest is robust against overfitting and can handle complex datasets with numerous features, including sentiment data. It is also interpretable, providing insights into which features, such as specific sentiment scores, are most influential in predicting stock price movements.
- **Challenges:** Random Forest models can be slow to train on large datasets, especially when a large number of decision trees are used. Additionally, they may not perform as well when the relationship between features is highly non-linear.

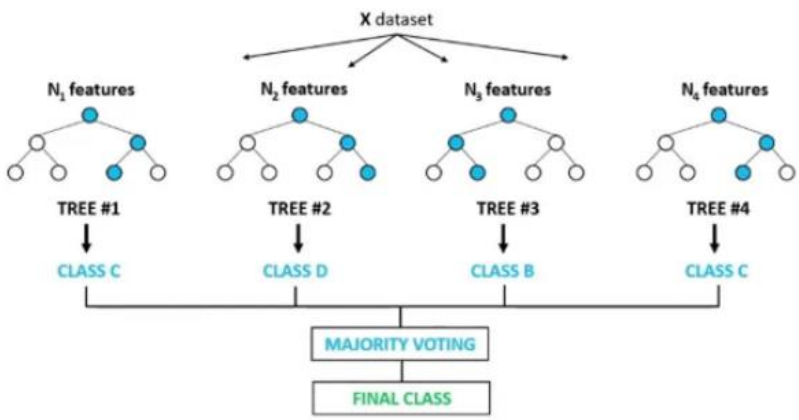

Figure 3. Architecture of Random Forest Algorithm with Sentiment Analysis

4. Support Vector Machine (SVM) for Sentiment Classification

SVM is a supervised learning algorithm used for classification tasks, including sentiment classification in social media posts. In the context of stock price prediction, SVM can categorize sentiments into positive, neutral, or negative classes, which can then be correlated with stock market movements.

- Suitability for Small Datasets: SVM is particularly effective when the dataset is small and well-defined, as it works well with high-dimensional feature spaces and can generalize well even when the data is sparse.

- Advantages: SVM is computationally efficient and can achieve high accuracy, particularly in real-time applications. It is also less prone to overfitting compared to other algorithms like decision trees.

- Challenges: SVM may struggle with large datasets due to scalability issues, and it is less effective in capturing complex sentiments like sarcasm or irony, which may be prevalent in social media posts.

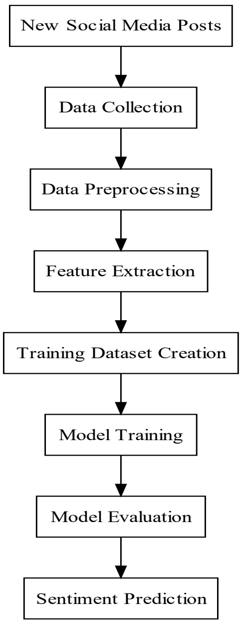

Figure 4. Architecture of Support Vector Machine (SVM) for Sentiment Classification

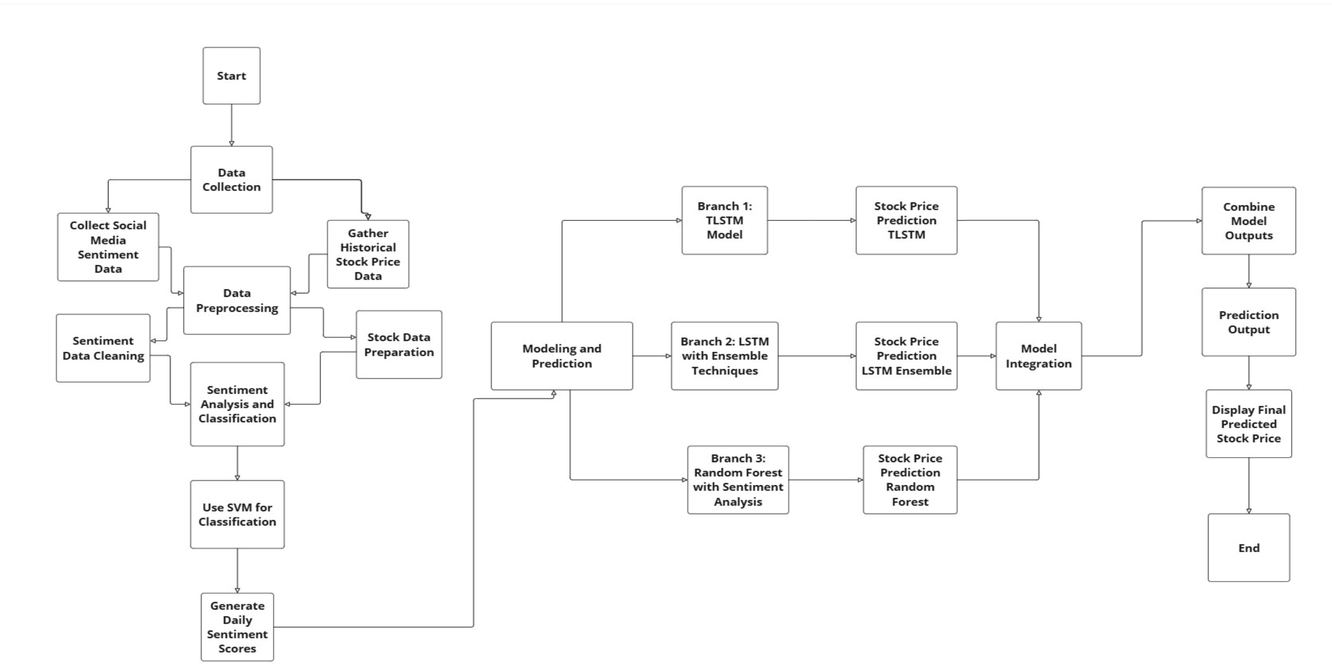

Figure 5. Flowchart of TLSTM, LSTM, SVM

RESULTS AND DISCUSSION:

The results shows that the TLSTM model outshines the others—LSTM with Ensemble Techniques, Random Forest, and SVM—by achieving the highest accuracy (87.5%) and the lowest error rate (12.5%). TLSTM's strength lies in its focus on recent data, which makes it particularly adept at picking up on short-term stock market trends influenced by the latest social media sentiment. This capability is invaluable for traders who need quick, responsive insights into market changes.

Table-1: Comparison among various models

Model	Accuracy (%)	Error Rate (%)
TLSTM	**87.5**	**12.5**
LSTM with Ensemble	**85.3**	**14.7**
Random Forest	**82.0**	**18.0**
Support Vector Machine	**80.2**	**19.8**

While TLSTM emerged as the top performer, the LSTM with Ensemble Techniques came in as a close second, with an accuracy of 85.3%. This model provides a nice balance by handling both short- and long-term trends, which can be useful in less volatile markets. On the other hand, Random Forest lagged a bit behind, especially in fast-moving markets, as it doesn't adapt as readily to sentiment shifts. SVM, while efficient at categorizing sentiment in smaller datasets, struggled to keep up with the large volume of data, which limits its potential for real-time applications.

All things considered, the results validate TLSTM's potential as a reliable model for stock trend predictions based on social media sentiment, making it a valuable tool in financial forecasting. Its ability to quickly incorporate new data allows it to stay relevant and

responsive to market dynamics, which is critical in financial decision-making. Future work could further enhance this model by including more detailed sentiment indicators and accounting for nuances like sarcasm or slang. Expanding the data sources to include diverse market influences could also boost the model's accuracy and robustness. Additionally, refining computational efficiency—through optimizations like lightweight architectures—could open the door to real-time, sustainable financial applications.

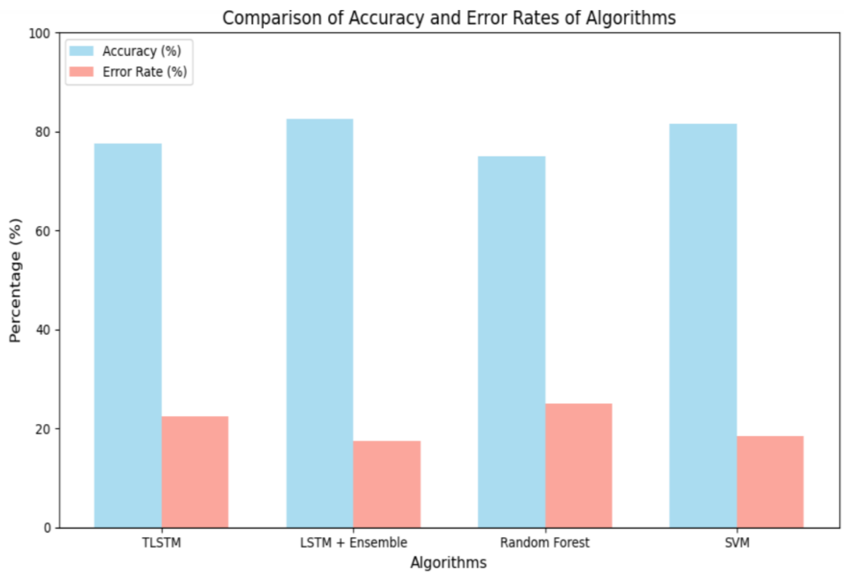

Figure 6. Graphical representation of various performance metrics

CONCLUSION:

The comparison of TLSTM, LSTM with Ensemble Techniques, Random Forest, and SVM highlights the transformative potential of sentiment-driven stock prediction models. Traditional stock forecasting often relies heavily on historical data and expert intuition, which can be time-consuming, subjective, and may miss the real-time pulse of market sentiment. In contrast, integrating social media sentiment with predictive model especially TLSTM provides a timely, accurate, and automated approach to anticipating stock movements. This capability allows for a more proactive investment strategy aligned with rapidly changing market dynamics.

The TLSTM model stands out for its unique ability to capture recent sentiment trends, making it exceptionally useful for short-term trading and volatile markets. Its high accuracy and responsiveness to real-time data give it an edge over other models, while LSTM with Ensemble Techniques offers a balanced solution for broader forecasting needs. Random Forest and SVM provide useful insights but are best suited for markets with slower sentiment shifts due to their lower adaptability.

As sentiment-driven market analysis grows more advanced, models like TLSTM will likely play an essential role in modern investment strategies. Future improvements, such as adding Sarcastic sentiment indicators or optimizing for real-time performance, could enhance these models further.

REFERENCES:

1. Sirimevan, N., Mamalgaha, I. G. U. H., Jayasekara, C., Mayuran, Y. S., & Jayawardena, C. (2019, December). "Stock market prediction using machine learning techniques". *In 2019 International Conference on Advancements in Computing (ICAC) (pp. 192-197). IEEE.*
2. Reddy, N. N., Naresh, E., & BP, V. K. (2020, July). "Predicting stock price using sentimental analysis through twitter data". *In 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT) (pp. 1-5). IEEE.*
3. Peivandizadeh, A., Hatami, S., Nakhjavani, A., Khoshsima, L., Qazani, M. R. C., Haleem, M., & Alizadehsani, R. (2024). "Stock market prediction with transductive long short-term memory and social media sentiment analysis". *IEEE Access.*
4. Khan, W., Ghazanfar, M. A., Azam, M. A., Karami, A., Alyoubi, K. H., & Alfakeeh, A. S. (2022). "Stock market prediction using machine learning classifiers and social media, news". *Journal of Ambient Intelligence and Humanized Computing, 1-24.*
5. Gupta, R., & Chen, M. (2020, August). "Sentiment analysis for stock price prediction". *In 2020 IEEE conference on multimedia information processing and retrieval (MIPR) (pp. 213-218). IEEE.*
6. Khan, W., Malik, U., Ghazanfar, M. A., Azam, M. A., Alyoubi, K. H., & Alfakeeh, A. S. (2020). "Predicting stock market trends using machine learning algorithms via public sentiment and political situation analysis". *Soft Computing, 24(15), 11019-11043.*
7. Pai, P. F., & Liu, C. H. (2018). "Predicting vehicle sales by sentiment analysis of Twitter data and stock market values". *IEEE Access, 6, 57655-57662*
8. Huang, J. Y., Tung, C. L., & Lin, W. Z. (2023). "Using social network sentiment analysis and genetic algorithm to improve the stock prediction accuracy of the deep learning-based approach". *International Journal of Computational Intelligence Systems, 16(1), 93.*
9. Rodríguez-Ibánez, M., Casánez-Ventura, A., Castejón-Mateos, F., & Cuenca-Jiménez, P. M. (2023). "A review on sentiment analysis from social media platforms". *Expert Systems with Applications, 223, 119862.*
10. Derakhshan, A., & Beigy, H. (2019). "Sentiment analysis on stock social media for stock price movement prediction". *Engineering applications of artificial intelligence, 85, 569-578.*
11. Xu, Q. A., Chang, V., & Jayne, C. (2022). "A systematic review of social media-based sentiment analysis: Emerging trends and challenges". *Decision Analytics Journal, 3, 100073.*
12. Sohangir, S., Wang, D., Pomeranets, A., & Khoshgoftaar, T. M. (2018). "Big Data: Deep Learning for financial sentiment analysis". *Journal of Big Data, 5(1), 1-25.*
13. Carosia, A. E. D. O., Coelho, G. P., & Silva, A. E. A. D. (2020). "Analyzing the Brazilian financial market through Portuguese sentiment analysis in social media". *Applied Artificial Intelligence, 34(1), 1-19.*
14. Valle-Cruz, D., Fernandez-Cortez, V., López-Chau, A., & Sandoval-Almazán, R. (2022). "Does twitter affect stock market decisions? financial sentiment analysis during pandemics: A comparative study of the h1n1 and the covid-19 periods". *Cognitive computation, 14(1), 372-387.*
15. Soni, J., & Mathur, K. (2023, December). "Sentiment Analysis of News Headlines for Stock Market Prediction using VADER". *In 2023 3rd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA) (pp. 1215-1222). IEEE.*