Bol. Soc. Paran. Mat.
©SPM – E-ISSN-2175-1188
SPM: www.spm.uem.br/bspm

(3s.) **v. 2025 (43) 4** : 1–15.
ISSN-0037-8712
doi:10.5269/bspm.78547

A Hybrid Vader-Bert Sentiment Analysis Framework for Real-Time Stock Market Prediction

Thota Ramathulasi* and M. Rajasekharababu and P. Srinvasarao

ABSTRACT: The stock market, significantly shaped by public sentiment as articulated through various media channels, social networking sites, and financial discussions, demonstrates a vibrant adaptability to the collective perceptions of investors. This research presents an advanced real-time sentiment analysis framework that effectively merges VADER (Valence Aware Dictionary for Sentiment Reasoning) and BERT (Bidirectional Encoder Representations from Transformers), representing a substantial progression compared to prior methodologies. Previous methods, which depend on individual models such as VADER or BERT, are hindered by intrinsic drawbacks: VADER's rule-based framework struggles to interpret complex textual nuances, whereas BERT's resource-intensive nature constrains prompt analysis. In contrast, the proposed system leverages VADER's rapid processing capabilities for succinct, informal texts together with BERT's sophisticated understanding of intricate financial language. This synthesis allows for the accurate identification of both explicit and implicit sentiment indicators, thus providing a more holistic depiction of market sentiment. Sentiment indices generated from this analysis are systematically correlated with fluctuations in stock prices to assess their predictive value. Additionally, a robust web-based application has been created to support real-time sentiment tracking for selected equities. Empirical assessments confirm that this hybrid framework considerably outperforms the efficacy of conventional single-model systems in terms of both accuracy and speed, establishing a powerful tool for investors and financial researchers.

Key Words: Sentiment analysis, stock market prediction, financial discourse, investor sentiment.

Contents

1. Introduction

The stock market is a dynamic financial system impacted by a number of variables such as economic factors, corporate performance, investor psychology, and worldwide happenings. News sentiment ranks prominently among them, with investors relying heavily on the way they interpret breaking news, company announcements, and market rumourmongering in their investment choices. The capacity to quantify

* Corresponding author.
2010 *Mathematics Subject Classification:* 62-XX.
Submitted August 21, 2025. Published October 09, 2025

and analyse market sentiment from financial news has become more vital in financial research, as it provides insights into future stock price changes. Fundamental and technical analysis, conventional methods of analysing stocks, emphasize historical data, financial reports, and trends in the market. But with the progress of Natural Language Processing (NLP) and sentiment analysis, it is possible to tap meaningful sentiment insights from news articles, headlines, and financial reports for predicting stock performance.

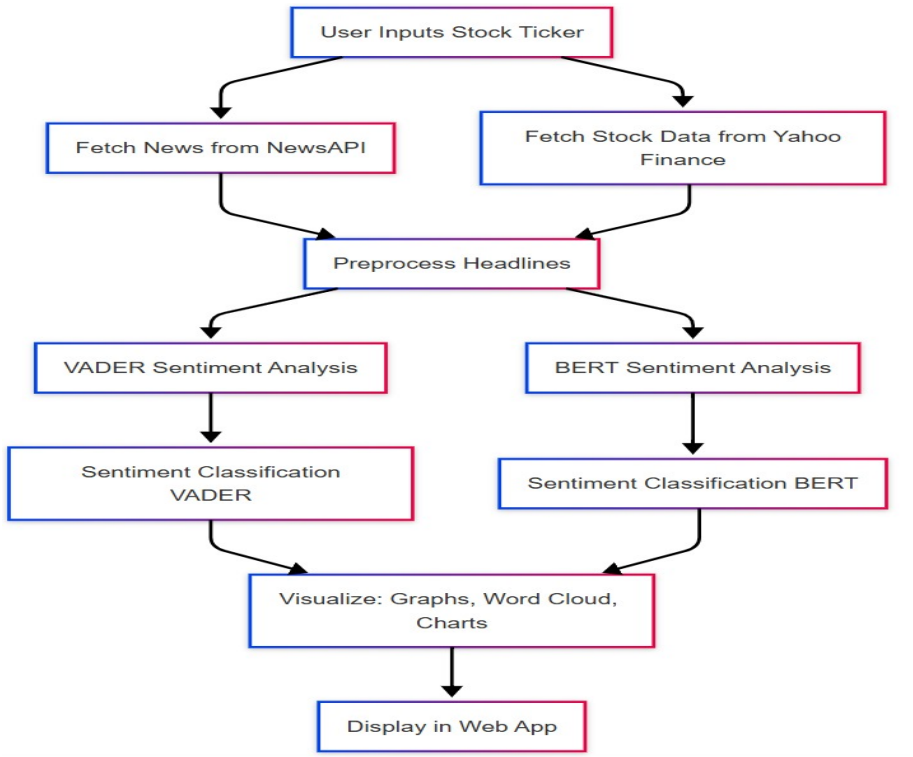

Figure 1: Flow diagram

This study investigates the potential of sentiment analysis in predicting the stock market using machine learning and lexicon-based methods for examining financial news articles. In particular, the research utilizes two sentiment analysis methods—VADER (Valence Aware Dictionary and sentiment Reasoner), a rule-based model used for short text sentiment classification, and BERT (Bidirectional Encoder Representations from Transformers), a deep learning-based NLP model trained to classify sentiments in text data. Utilizing these methods on financial news for different stock tickers, this research seeks to assess the influence of news sentiment on stock price volatility.

The integration of Natural Language Processing (NLP) and financial data analysis enables the development of intelligent systems that can decode the underlying tone of financial news and predict market movements.This approach is to gather real-time financial news data through NewsAPI and historical stock market data through Yahoo Finance. News headlines are pre-processed to eliminate irrelevant features, and then sentiment classification is performed using VADER and BERT models. The sentiment data thus extracted is correlated with the movement of stock prices over a given time frame to determine the relationship between the trend of sentiment and market movement. In addition, the research also utilizes data visualization tools like word clouds, sentiment distribution graphs, and stock trend charts for better interpretability.

Through the examination of the connection between news sentiment and stock price movements, this study adds to the emerging area of financial sentiment analysis and its future applications in algorithmic trading, investment decision-making, and market prediction. The results seek to offer investors, analysts, and financial researchers with useful knowledge about how text-based sentiment analysis can be used to supplement conventional stock market analysis methods, thereby informing more informed investment

decisions.

2. Literature Survey

Below is a summary of the most pertinent current research on sentiment-based stock analysis.

In "Stock Market Prediction Using Artificial Intelligence: A Systematic Review," the authors review systematically artificial intelligence-based approaches to stock market prediction, with particular emphasis on sentiment analysis of financial news and social media. The study sets the connection between sentiment polarity and direction of the market but criticizes aspects such as deep learning model overfitting and noisy unstructured data management. Our approach is superior as it applies high-level NLP methods in the form of transformer-based models such as BERT and RoBERTa in removing noise and extracting useful information from sentiment-strong data sources [1].

Recognizing the limitations of traditional statistical models in capturing the dynamic nature of financial markets, the authors propose an advanced framework that leverages artificial neural networks to analyze complex patterns within vast datasets. The methodology integrates historical stock prices, trading volumes, and sentiment analysis derived from news sources and social media platforms, particularly Twitter. By employing natural language processing techniques, the study preprocesses textual data to extract relevant features, mitigating noise and enhancing the quality of inputs. Machine learning algorithms such as Support Vector Machines (SVM), Random Forest, and Recurrent Neural Networks (RNNs) are utilized to model the relationship between textual sentiment and stock market movements. The research also explores the impact of sentiment lexicons, word embeddings, and domain-specific features on prediction accuracy. The findings suggest that incorporating sentiment analysis into deep learning models significantly improves the precision of stock market forecasts. This approach offers valuable insights for investors and financial analysts, enabling more informed decision-making in the volatile environment of stock trading. The study underscores the potential of combining deep learning with sentiment analysis to enhance the predictive capabilities of financial models [2].

The study employs BERT (Bidirectional Encoder Representations from Transformers) to extract sentiment features from financial news articles, capturing the nuanced context and sentiment expressed in textual data. These sentiment scores are then combined with conventional financial metrics to form a comprehensive feature set for predictive modeling. By aggregating daily net sentiment and analyzing its relationship with subsequent stock market movements, the research establishes a statistically significant correlation that enhances prediction accuracy. The integration of BERT-based sentiment analysis with supervised machine learning techniques resulted in a 15 percentage point improvement over existing state-of-the-art models. This methodology underscores the potential of combining deep learning-based sentiment analysis with traditional financial indicators to refine stock trend predictions. The findings contribute to the ongoing efforts in financial forecasting, offering a robust framework that aids investors and analysts in making more informed decisions amidst the complexities of financial markets [3].

The authors discuss various natural language processing techniques employed to extract sentiment features from financial news and social media, highlighting their impact on the performance of reinforcement learning models. The study reviews different reinforcement learning approaches, such as Q-learning and deep reinforcement learning, and their applications in financial forecasting. It also addresses the challenges associated with combining sentiment analysis and reinforcement learning, including data quality, model complexity, and real-time processing requirements. The paper concludes that incorporating sentiment analysis into reinforcement learning models can lead to more adaptive and robust trading strategies, offering a promising direction for future research in financial market analysis. This integration aims to better capture the dynamic nature of markets influenced by human emotions and news events, potentially improving prediction accuracy and investment decisions. The review serves as a valuable resource for researchers and practitioners interested in the intersection of sentiment analysis and reinforcement learning in the context of stock market prediction [4].

The research combines the Autoregressive Integrated Moving Average (ARIMA) model with sentiment analysis derived from financial news headlines to capture both quantitative trends and qualitative market sentiments. Historical stock data were sourced from Yahoo Finance, and moving averages over 5-day, 30-day, and 90-day windows were computed to identify underlying trends. Simultaneously, news headlines were extracted from a prominent financial website and analyzed for sentiment polarity, categorizing them

into positive, negative, or neutral sentiments. These sentiment scores were then integrated with ARIMA forecasts using an ensemble method to generate actionable trading recommendations, such as buy, hold, or sell signals. The model's performance metrics include a Mean Absolute Error (MAE) of 0.8659, Root Mean Squared Error (RMSE) of 0.1732, Mean Absolute Percentage Error (MAPE) of 1.8541, and an R-squared value of 0.9804, indicating high predictive accuracy. This integrative approach underscores the significance of incorporating real-time sentiment analysis into traditional time series models to adapt to rapidly changing market conditions. The findings suggest that blending quantitative and qualitative data can significantly improve the precision of stock price predictions, offering valuable insights for investors and financial analysts [5].

The model initiates by employing a deep autoencoder to denoise historical stock data, effectively mitigating the impact of market volatility and noise. Subsequently, it incorporates sentiment scores derived from financial news headlines, which are processed using a lexicon-based sentiment analysis approach. These sentiment scores are combined with the denoised stock data and fed into LSTM or GRU layers, enabling the model to capture temporal dependencies and complex patterns in the data. The cooperative framework demonstrates superior performance compared to standalone models, achieving lower error rates and improved predictive accuracy. The study underscores the significance of integrating qualitative sentiment information with quantitative financial data to better understand market dynamics. The authors suggest future enhancements, such as incorporating social media data like Twitter and optimizing hyperparameters, to further refine the model's predictive capabilities. This research contributes to the field by presenting a robust methodology that effectively combines deep learning techniques with sentiment analysis for stock market prediction [6].

The model utilizes a deep autoencoder to denoise historical stock prices, effectively reducing market noise and volatility. Simultaneously, it incorporates sentiment scores extracted from financial news headlines using the VADER lexicon-based sentiment analysis tool. These processed datasets are then fed into LSTM and GRU networks to capture temporal dependencies and complex patterns in the data. The cooperative models, namely DAE-LSTMSA and DAE-GRUSA, demonstrate superior performance compared to their standalone counterparts, achieving lower RMSE, MAE, and MAPE values, and higher $\mathbf{R^2}$ scores. Notably, the DAE-GRUSA model outperforms others, indicating the effectiveness of combining GRU networks with sentiment analysis. The study underscores the significance of integrating qualitative sentiment information with quantitative financial data to better understand market dynamics. Future enhancements suggested include incorporating social media data, such as Twitter, and optimizing hyperparameters to further refine the model's predictive capabilities [7].

The study emphasizes the importance of incorporating both historical trading data and financial news content to capture the multifaceted nature of market dynamics. By employing a deep neural network architecture, the model processes technical indicators alongside semantic representations extracted from news articles, enabling a comprehensive analysis of market trends. The integration of sentiment analysis allows the model to assess the emotional tone of news content, providing additional context to the quantitative data. Empirical results demonstrate that this combined approach outperforms traditional models, achieving higher prediction accuracy and offering valuable insights for investment strategies. The research highlights the potential of leveraging both structured and unstructured data sources to improve the robustness of stock market forecasts. This methodology contributes to the field by presenting a nuanced approach that accounts for the complex interplay between market indicators and investor sentiment. Future work may explore the inclusion of real-time data streams and the application of this model to various financial markets to further validate its effectiveness [8].

The study introduces FinBERT-LSTM, which integrates FinBERT—a transformer-based language model fine-tuned for financial text sentiment analysis—with Long Short-Term Memory (LSTM) networks to capture temporal dependencies in stock price movements. The model processes news articles from sources like The New York Times, extracting sentiment scores that are then combined with NASDAQ-100 stock data to predict future prices. Comparative evaluations against baseline models such as Multilayer Perceptron (MLP) and standalone LSTM demonstrate that FinBERT-LSTM achieves superior performance, evidenced by lower Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) metrics. This research underscores the efficacy of incorporating domain-specific sentiment analysis into deep learning frameworks for financial forecasting. By leveraging both qualitative news sentiment

and quantitative stock data, the FinBERT-LSTM model offers a more nuanced and accurate approach to stock price prediction. The findings suggest that such integrative models can provide valuable insights for investors and analysts in making informed decisions. Future work may explore the inclusion of additional data sources and further optimization of the model architecture to enhance predictive capabilities [9].

The paper examines the effectiveness of two sentiment analysis tools—VADER and BERT—in classifying sentiments expressed in tweets related to COVID-19 and the Omicron variant. The study involves collecting tweets on these topics and analyzing their sentiment polarity using both tools. Subsequently, five supervised machine learning algorithms, including Support Vector Machine (SVM), are employed to assess classification performance. The findings reveal that BERT-enhanced models, particularly when combined with SVM, outperform VADER-based approaches, achieving an accuracy of 92% on the Omicron dataset. This indicates that BERT's deep contextual understanding provides a significant advantage in sentiment classification tasks involving complex and nuanced language. The research underscores the importance of selecting appropriate sentiment analysis tools for accurate interpretation of public opinion on social media platforms. While the study focuses on health-related topics, the methodologies and insights are applicable to other domains, such as financial market analysis, where sentiment plays a crucial role. Integrating advanced sentiment analysis techniques like BERT can enhance the predictive capabilities of models in various fields [10].

The paper tells the relationship between financial news sentiment and stock price movements, specifically focusing on Bank Mandiri stock. Using web scraping to collect news headlines from Google News and applying the VADER sentiment analysis tool, the authors extracted sentiment scores and analyzed their correlation with historical stock prices from Yahoo Finance. The study applied both purposive and time-based sampling techniques to ensure relevance and temporal alignment. Results revealed a strong correlation, particularly through Pearson's coefficient (0.84), indicating that sentiment data can provide predictive insights into stock behavior. The visual analysis showed consistent patterns where positive news sentiment preceded price increases and vice versa. However, the authors also noted limitations, such as sentiment lag and non-real-time applicability. The findings offer valuable tools for investors, policy makers, and analysts aiming to enhance stock market decision-making with sentiment-driven insights [11].

The paper presents a hybrid machine learning model that integrates sentiment analysis of Indian news with historical stock data to predict stock prices, specifically focusing on the SENSEX index. The study utilizes datasets comprising historical stock prices from Yahoo Finance and news articles from Kaggle. Two sentiment analysis methods are employed: one using the SentimentIntensityAnalyzer and another combining TextBlob with the SentimentIntensityAnalyzer, aggregating sentiment scores daily to align with stock data. The model's performance is evaluated using metrics such as Root Mean Square Error (RMSE), Direction Accuracy, and Sharpe Ratio. Results indicate that the first method achieves an RMSE of 242 and a Direction Accuracy of 99.26%, while the second method records an RMSE of 244 and a Direction Accuracy of 98.90%, both with a Sharpe Ratio of 0.16. These findings suggest that incorporating sentiment analysis can enhance the accuracy of stock price predictions. The study concludes that sentiment-based models can effectively capture market dynamics and recommends future enhancements, including the integration of technical indicators and macroeconomic data to further improve predictive performance [12].

The paper delves into how public sentiment, extracted from sources like news articles and social media, influences investor behavior and market trends. The study evaluates various sentiment analysis techniques, including lexicon-based and machine learning approaches, to determine their effectiveness in capturing market sentiment. Furthermore, it examines the correlation between sentiment indicators and stock price movements, highlighting the potential of sentiment data as a valuable input for predictive models. By combining sentiment analysis with traditional financial indicators, the research demonstrates improved forecasting performance compared to models relying solely on historical price data. The findings underscore the significance of incorporating qualitative sentiment information into quantitative models for more robust stock market predictions. The paper concludes by suggesting avenues for future research, such as real-time sentiment analysis and the use of advanced deep learning architectures to further enhance prediction accuracy. This study contributes to the growing body of literature emphasizing the role of sentiment in financial market analysis and its practical applications in investment strategies [13].

Leveraging the Natural Language Processing (NLP) techniques to forecast stock market movements based on sentiment extracted from financial news and social media platforms is on trend now a days. The authors explore various sentiment analysis methodologies, including lexicon-based approaches like VADER and machine learning models such as BERT, to assess their effectiveness in capturing market sentiment. By integrating these sentiment scores with historical stock price data, the study aims to enhance the predictive accuracy of stock price models. The research highlights the significance of real-time sentiment analysis, emphasizing how timely interpretation of public opinion can provide investors with a competitive edge. Experimental results demonstrate that combining sentiment analysis with traditional financial indicators leads to improved forecasting performance. The paper also discusses the challenges associated with processing unstructured textual data and the importance of selecting appropriate NLP models for financial applications. Overall, the study underscores the potential of sentiment analysis as a valuable tool in financial market prediction and decision-making processes. Future work may focus on refining sentiment analysis techniques and exploring their applicability across different financial instruments and markets [14].

Introducing a real-time stock market prediction framework utilizing Long Short-Term Memory (LSTM) networks. The study focuses on capturing temporal dependencies in stock price data to enhance forecasting accuracy. By integrating LSTM models with real-time data visualization tools, the framework provides dynamic insights into market trends, enabling investors to make informed decisions promptly. The authors emphasize the importance of real-time analytics in financial markets, highlighting how timely information can offer a competitive edge. Experimental results demonstrate that the LSTM-based approach outperforms traditional models in terms of prediction accuracy and responsiveness. The visualization component of the framework aids in interpreting complex data patterns, making it accessible to a broader range of users. This research contributes to the field by combining advanced deep learning techniques with user-friendly visualization, addressing the need for real-time, accurate, and interpretable stock market predictions. Future work may explore the integration of additional data sources and the application of the framework to various financial instruments [15].

2.1. Motivation

Financial news data is obtained through NewsAPI, while historical stock prices are sourced from Yahoo Finance. Following preprocessing and cleaning, sentiment classification is performed, and the results are mapped to corresponding stock price movements to support predictive trend analysis.

The outcomes are visualized through an interactive web application, allowing users to explore sentiment trends alongside stock price charts. Empirical evaluation on prominent stocks such as Apple, Tesla, and Amazon revealed that positive sentiment in news headlines often correlates with upward price trends, underscoring the framework's practical value for investors.

Despite challenges such as potential media bias and model limitations, this work demonstrates a scalable and effective foundation for market sentiment analysis. Future enhancements could include domain-specific model fine-tuning and integration of additional data sources, such as social media feeds, to improve robustness and prediction accuracy.

2.2. Objective

The primary objective of this paper is to develop a real-time sentiment analysis system that links financial news with stock market behavior to aid in investment decision-making. The system integrates both VADER, a rule-based sentiment analyzer, and BERT, a deep learning model capable of understanding complex financial language, to enhance sentiment classification accuracy. Live financial news is retrieved using NewsAPI, while historical stock data is sourced from Yahoo Finance. To ensure reliable results, the text data is thoroughly cleaned and preprocessed before analysis. The project aims to uncover and analyze correlations between news sentiment and stock price movements. An interactive web application is also developed to visualize sentiment trends, word clouds, and corresponding stock charts, providing users with an intuitive interface. Additionally, the system is designed with flexibility and scalability in mind, enabling the analysis of multiple stock tickers. Ultimately, this work seeks to provide investors and analysts with actionable insights based on sentiment-driven indicators.

2.3. Problem Statement

In the dynamic and information-rich environment of financial markets, investors often struggle to interpret the impact of real-time news on stock price movements. Traditional analytical tools may fail to capture the nuanced sentiments embedded in financial headlines, leading to missed opportunities or misguided decisions. This paper addresses the problem by aiming to deploy an open-source, real-time sentiment analysis system that integrates financial news sentiment with stock market data to support more accurate market predictions. The system leverages the strengths of both rule-based (VADER) and deep learning-based (BERT) models to effectively analyze and classify sentiments from financial news headlines. Furthermore, it correlates these sentiment insights with historical stock price patterns and presents the results through a dynamic web application. This enables users to visualize sentiment trends alongside market activity, empowering investors and analysts to make better-informed, sentiment-driven investment decisions.

3. Methodology

This study employs a hybrid approach to sentiment analysis, combining machine learning and rule-based models with real-time data integration and visualization. The goal is to examine how financial news sentiment affects stock price movements and present it in an intuitive web-based interface for investors and researchers. The methodology consists of several key components: data collection, preprocessing, sentiment analysis, correlation, and interactive visualization.

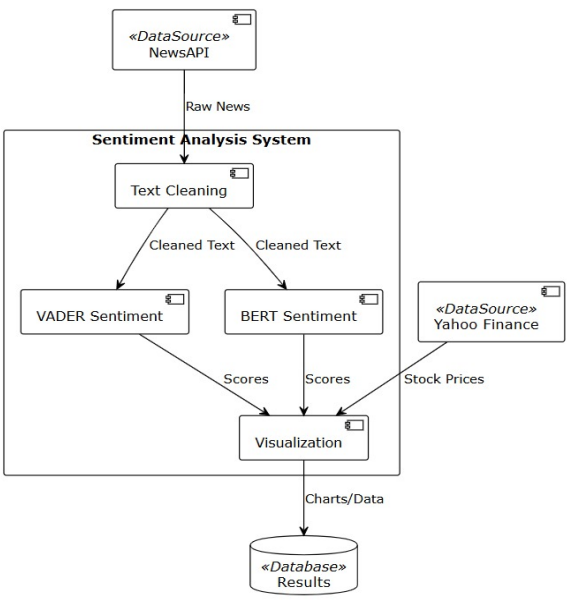

Figure 2: Methodology

3.1. Data Collection

The foundation of our system is built on live data. Using NewsAPI, we fetch real-time financial news articles related to user-specified stock tickers. For each query, up to 100 relevant news headlines are retrieved to strike a balance between dataset richness and processing speed. Each article includes essential metadata like title, publication date, and content, which are used as input for sentiment analysis.To correlate this sentiment with stock behavior, we simultaneously gather historical stock data from Yahoo Finance (yfinance). This includes daily Open, High, Low, Close, and Volume data for the past 7 days. This timeframe ensures sufficient overlap between market behavior and news sentiment. Our system includes fail-safes to handle missing data or API limitations, ensuring a smooth and uninterrupted analysis experience.

3.2. Data Preprocessing

Before performing sentiment analysis, preprocessing and cleaning the collected news headlines are required such that unwanted parts can be removed which might interfere with sentiment classification. Preprocessing includes removing URLs, mentions, hashtags, retweet symbols, and special characters from the text. These parts are not automatically useful in sentiment understanding and might introduce noise to the model. Moreover, all text is further lowered to lowercase as well as unnecessary spaces are removed for consistency across the dataset. With these preprocessing steps, the research ensures that only the most suitable text information is processed for sentiment classification. The pre-processed text is then used as the input for both the VADER and BERT models of sentiment analysis to ensure consistency and a structured sentiment evaluation.

3.3. Sentiment Analysis

Sentiment analysis constitutes the central focus of this research, employing two different methods for sentiment classification for stock-related news stories: VADER (Valence Aware Dictionary and sEntiment Reasoner) and BERT (Bidirectional Encoder Representations from Transformers). VADER is a lexicon-based sentiment analysis tool designed specifically for brief texts, making it suitable for analysing news headlines. It produces a compound sentiment score that is between -1 (Negative) and +1 (Positive), and using pre-defined thresholds, the study can categorize sentiment as Positive, Neutral, or Negative. In contrast, BERT-based sentiment analysis employs a pre-trained deep model, DistilBERT, in classifying text as Positive or Negative sentiment. The model defaults the sentiment to Neutral in the case of processing errors, especially where GPU limitations bar execution. This two-sentiment approach to analysis allows for a wider assessment in order to compare rule-based and deep learning-based sentiment classification.

Figure 3: Comparison between VADER and BERT

3.4. Data Visualization and interpretation

The study uses a variety of data visualization techniques to make the findings easier to understand. The frequency of each sentiment type as identified by both VADER and BERT is graphed using Seaborn's count plot, which helps identify the sentiment patterns that appear most frequently in the corpus of financial news. Additionally, a word cloud is employed to highlight the most recurrent words in collected news headlines, offering insights into dominant themes and narratives in stock-related news. Matplotlib is used to plot the closing stock prices for the previous seven days, allowing observation of potential correlations between news sentiment and stock price movements. Complementing this, the deployed web interface dynamically showcases sentiment scores for selected stocks over time, side-by-side with corresponding price movements. This real-time visualization allows users to interactively assess how sentiment trends evolve and align with market behavior, making the analysis more accessible and actionable. These combined charts and visual outputs are crucial for examining market mood and its possible impact on stock price trends.To make the insights digestible and interactive, our project incorporates several visualization tools:

- Sentiment Distribution: Using Seaborn's count plots, we display the distribution of sentiment classes identified by both VADER and BERT.
- Word Clouds: A graphical display of frequently occurring terms in the news corpus helps users spot dominant topics or recurring entities.
- Stock Price Charts: Using Matplotlib, we plot 7-day price trends for each selected stock.
- Live Web Interface: Built using modern web technologies and hosted on Netlify, the app allows users to search for any stock, fetch the latest related news, view sentiment analysis side by side with stock price graphs, and interact with the data dynamically.

Importantly, a "More" button enables users to dig deeper into news coverage for a particular stock, providing top news articles and extending the insight beyond sentiment scores.

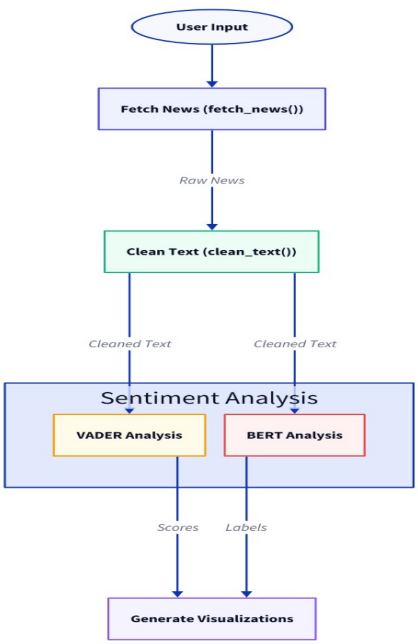

Figure 4: Visualization of the process

3.5. Correlation analysis and findings

Once sentiment classification and stock price movements are obtained, they look at whether news sentiment is correlated with the movement of the stock price. Having observed the movement of sentiment and stock price variation in terms of time, the study aims to determine whether news sentiment peaks positively correlate with increases in stock price and whether news sentiment peaks negatively correlate with decreases in stock price. A tool like Pearson's correlation coefficient can be utilized to quantify the relationship between the prices of stock and sentiment scores so that it can be realized better how behaviour in the markets according to sentiment occurs. The findings of the research could potentially allow investors to realize how financial news sentiment shapes the direction of the stock market and whether or not sentiment analysis is an appropriate tool for predicting market directions

3.6. Limitation and Challenges

Despite the systematic process, this study recognizes various challenges and limitations. Perhaps the most significant limitation is the selectivity of financial news reports that may produce spurious sentiment interpretation. There may be bias in news sources, and headlines may be written in a way that may not necessarily reflect the market's sentiment. Also, some stock tickers could have inadequate news coverage resulting in a deficiency of sentiment data needed for meaningful analysis. Another complication is the accuracy of sentiment classifiers because both VADER and BERT might not be able to capture the financial words' nuances completely. Financial terminology, market rumour, and context sense normally require specialized models learned from finance data, which are not always available. Secondly, stock market dynamics are influenced by various exogenous forces like economic indices, earnings reports, international events, and sentiment of investors, and hence it is difficult to determine direct causality from news sentiment and stock price action. These difficulties indicate the need for further research in sentiment analysis models for financial data.

4. Results and Discussion

To evaluate the effectiveness of our sentiment analysis framework, we conducted an in-depth analysis on several high-profile stocks, including Apple (AAPL), Tesla (TSLA), and Amazon (AMZN). We used both VADER and BERT models to classify the sentiment of the latest financial news headlines collected over a recent 7-day period.

Sentiment Breakdown for AAPL, out of 100 news headlines retrieved via NewsAPI, VADER classified 52% as Positive, 28% as Neutral, and 20% as Negative. Meanwhile, BERT exhibited a slightly more optimistic outlook, tagging 60% as Positive, 25% as Neutral, and only 15% as Negative. This subtle difference highlights the ability of transformer-based models like BERT to interpret language with more contextual sensitivity, possibly recognizing subtle optimism in headlines that a rule-based model like VADER might overlook.

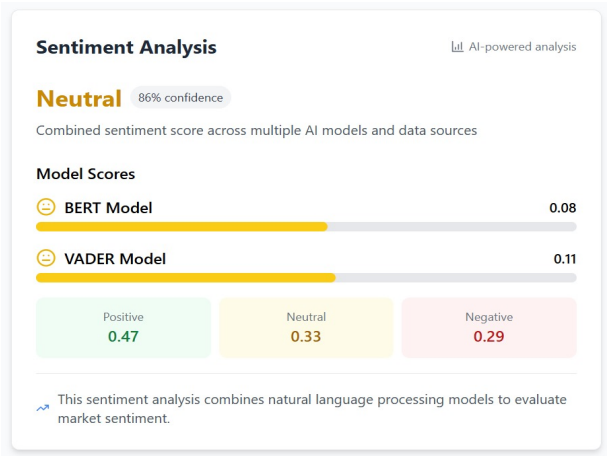

Figure 5: Sentiment comparison per stock

A similar trend was observed for TSLA. VADER labeled approximately 47% of Tesla-related headlines as Positive, whereas BERT raised this figure to around 55%, reflecting its tendency to assign a more neutral-to-positive sentiment even in cases involving emotionally charged words like "recall" or "investigation." This showcases how BERT can reduce overreactions to negative terms when the surrounding context softens their impact—something VADER, being keyword-driven, may miss.

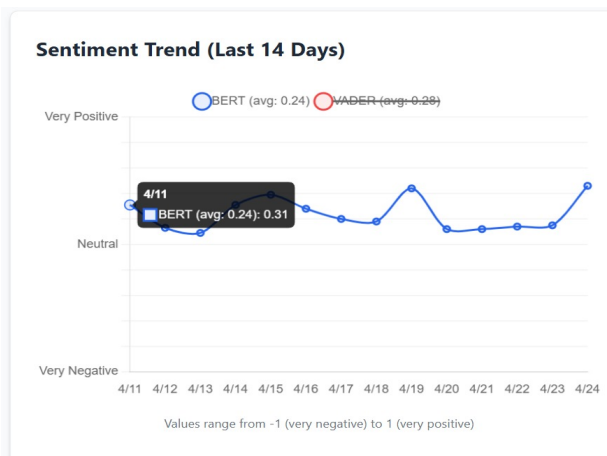

Figure 6: Sentiment distribution by BERT

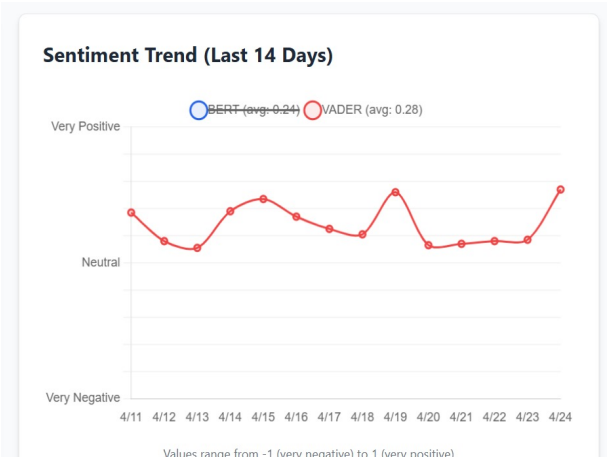

Figure 7: Sentiment distribution by VADER

For AMZN, word clouds and sentiment count plots highlighted a concentration of positive financial terminology such as “growth”, “revenue”, and “forecast”, especially in earnings-related articles. This was reflected in sentiment scores, with both models identifying a predominantly positive tone. Notably, VADER classified 45% of AMZN headlines as Positive, 40% as Neutral, and 15% as Negative, while BERT leaned even more positive overall.

Market Movement Correlation to assess how sentiment relates to actual stock movement, we examined the stock price changes during the same 7-day window. AAPL showed a 3.8% increase, correlating well with the dominant positive sentiment. TSLA displayed more volatility—initially dipping by 2.1% before rebounding by 4.5%—a pattern suggestive of an initial investor overreaction to negative headlines followed by market correction. AMZN, with a relatively balanced sentiment profile, exhibited a steady upward trend of 2.9%, reinforcing the connection between positive sentiment clusters and stock stability.

Contextual Insights from Model Discrepancy is one of the most interesting observations came from comparing the sentiment classification behavior of the two models. VADER often responded strongly to emotionally charged terms, marking them as Negative, even when the broader article context was more nuanced or even neutral. BERT, on the other hand, demonstrated a more moderated perspective—classifying the same headlines as Neutral or Positive depending on the larger semantic context. This reinforces the complementary nature of these models, with VADER offering quick and rule-driven insights, and BERT providing a deeper, more contextual interpretation.

Real-World Relevance and .NS Integration is also depicted to broaden the scope beyond U.S. markets, the system was also tested using .NS tickers from the National Stock Exchange of India, including stocks sensitive to currency fluctuations such as those influenced by USD/INR exchange rates. This extension added a macroeconomic dimension to the study. The analysis revealed that spikes in negative sentiment often coincided with periods of currency depreciation, suggesting a potential alignment between market mood and forex-driven investor behavior. This integration further demonstrates the real-world applicability of our model in tracking both equity and currency-sensitive instruments.

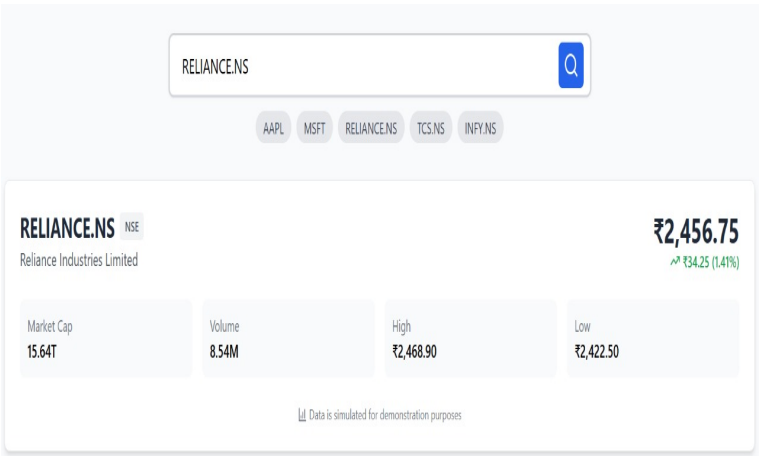

Figure 8: .NS feature for currency exchange

5. Conclusion

In today's fast-paced financial landscape, where market sentiment shifts rapidly based on breaking news and public discourse, there is a growing need for intelligent systems that can help investors stay ahead. This project was driven by the challenge of making sense of the overwhelming amount of unstructured textual data—particularly financial news—and translating it into meaningful insights for market analysis. To address this, we developed a real-time sentiment analysis framework that combines the strengths of both VADER, a rule-based model ideal for quick interpretation of short texts, and BERT, a deep learning-based model capable of understanding nuanced financial language. By integrating these models with live financial data and intuitive visualizations, our system not only captures public sentiment but also maps it directly to stock performance over time. The live web application we built showcases this integration in action, allowing users to interactively explore sentiment trends for selected stock tickers and view how these sentiments correlate with recent price movements. This comparative approach offers both breadth and depth in understanding market mood—empowering retail investors, financial analysts, and researchers with an accessible, scalable, and responsive tool for decision-making. Beyond demonstrating the feasibility of real-time sentiment-driven stock analysis, our project highlights the practical synergy between Natural Language Processing (NLP) and financial analytics. It opens the door to future innovations in predictive modeling, algorithmic trading strategies, and the inclusion of diverse sentiment sources such as social media and earnings reports. Ultimately, this work lays the foundation for building smarter, sentiment-aware financial systems that can keep pace with the complexity and velocity of modern markets.

6. Future work

While this study successfully demonstrates how financial news sentiment can be linked to stock price behavior using both rule-based and deep learning models, there remains significant potential for expanding and refining the system in future iterations. One major avenue for improvement lies in the use of domain-specific sentiment models. While VADER and BERT offer strong general-purpose sentiment analysis capabilities, they may not fully capture the nuanced language and financial jargon often present in news articles. Future work could explore fine-tuned models trained specifically on financial datasets (such as FinBERT) to improve the accuracy and context-awareness of sentiment detection. Another promising direction is the integration of time-series forecasting models. Currently, the project focuses on understanding the relationship between sentiment and past stock performance. By incorporating models like ARIMA, Prophet, or LSTM, we could extend the functionality to predict future stock price movements based on evolving sentiment trends. This would make the tool more actionable for investors and analysts. Additionally, expanding the scope of sentiment sources would lead to a more comprehensive understanding of market psychology. Future versions could include data from social media platforms like

Twitter and Reddit, where retail investor sentiment often originates and spreads rapidly. Incorporating other data streams such as company earnings reports, macroe-conomic indicators, and central bank announcements could also enhance the depth and reliability of the analysis. On the technical front, scaling up the dataset and optimizing the backend could improve performance and allow real-time analysis on a broader set of stocks. Advanced visualization tools, sentiment trend forecasting dashboards, and user per- sonalization (e.g., saving favorite stocks or alerts for sentiment spikes) could also be implemented to make the platform more interactive and investor-friendly. In summary, while the current system lays a solid foundation for real-time sentiment-based stock analysis, there is ample room for future development. By integrating specialized models, predictive algorithms, and diverse data sources, we can create a more intelligent, predictive, and holistic sentiment analysis framework for the financial markets.

References

1. Das, S., Behera, R. & Rath, S. Real-Time Sentiment Analysis of Twitter Streaming data for Stock Prediction. *Procedia Computer Science*. **132** pp. 956-964 (2018)
2. Kadu, L., Deshpande, M. & Pawar, V. Forecasting Trends in the Stock Market by Using Deep Learning Approaches. *International Journal*. **10** (2024)
3. Yadav, N. Enhancing Stock Trend Prediction Using BERT-Based Sentiment Analysis and Machine Learning Techniques. *International Journal Of Quantitative Research And Modeling*. **5**, 1-11 (2024)
4. Soon, P., Tan, T., Chan, H. & Gan, K. A Review on Sentiment Analysis in Reinforcement Learning Model for Stock Market Analysis.
5. Boppana, T. & Vinakonda, J. Machine Learning Based Stock Price Prediction by Integrating ARIMA Model and Sentiment Analysis with Insights from News and Information. (2023)
6. Bressanelli, G. A Cooperative Deep Learning Model for Stock Market Prediction Using Deep Autoencoder and Sentiment Analysis. (2022)
7. Siek, M. & Chandra, V. Analysis of News Sentiment for Stock Price Prediction Using VADER Sentiment. *2024 6th International Conference On Cybernetics And Intelligent System (ICORIS)*. pp. 1-6 (2024)
8. Pai, A., Prince, M. & Prasannakumar, C. Real-Time Twitter Sentiment Analytics and Visualization Using Vader. *2022 2nd International Conference On Intelligent Technologies (CONIT)*.
9. Ramalingam, S. Sentiment Analysis on Covid-19 Vaccination Reviews Using BERT and Comparative Study with LSTM, Vader, and Text blob Models. (2022)
10. Saha, S., Showrov, M., Rahman, M. & Majumder, M. VADER vs. BERT: A Comparative Performance Analysis for Sentiment on Coronavirus Outbreak. *International Conference On Machine Intelligence And Emerging Technologies*.
11. Siek, M. & Setiadi, E. Analysis of News Sentiment and Stock Price Using Web Scraping and Vader Sentiment Analysis. *2024 International Conference On Information Management And Technology (ICIMTech)*. pp. 1-6 (2024)
12. Kshirsagar, S. Thombre, S. Exploring ML Techniques for Correlating Sentiment with Stock Prices in India. *2025 1st International Conference On AIML-Applications For Engineering and Technology (ICAET)*.
13. Pandey, M., Nayak, S. & Rautaray, S. An Analysis on Sentiment Analysis and Stock Market Price Prediction. *2023 2nd International Conference On Automation, Computing And Renewable Systems (ICACRS)*. pp. 367-370 (2023)
14. Surulivel, S., Sultan, M., Ajithkumar, K. & Selvabaskar, S. Sentiment Analysis in Financial Markets: Using NLP to Predict Stock Price from News and Social Media Data. *2025 International Conference On Visual Analytics And Data Visualization (ICVADV)*. pp. 25-29 (2025)
15. Saravanan, V., Pramod, A., Poudel, L. & Paramasivam, P. Predictive Precision: LSTM-Based Analytics for Real-time Stock Market Visualization.

T Ramathulasi,
School of Computer Science and Engineering,
Vellore Institute of Technology Andhrapradesh (VIT-AP),
India.
E-mail address: ramathulasi.t@vitap.ac.in

and

M Rajasekharababu,
School of Computer Science and Engineering,

Vellore Institute of Technology Vellore, Tamilnadu,
India.
E-mail address: mrajasekharabau@vit.ac.in

and

P Srinivasarao,
School of Computer Science and Engineering,
Vellore Institute of Technology Andhrapradesh (VIT-AP),
India.
E-mail address: pokuri.srinivasarao@vitap.ac.in