Sentiment Driven Market Prediction: Evaluating News Influence on Stock Performance

Karthikeyan N[1], Jaswanth Kumar S[2], Dheepanraj S R[3], Mrs. Babisha A[4]
[1,2,3,4] *Panimalar Institute of Technology, Chennai, India*

Abstract - **In recent years, sentiment analysis has emerged as a powerful tool for understanding public perception and its impact on stock market behavior. This study advances traditional sentiment-based stock prediction by integrating cutting-edge AI techniques, feature engineering, and predictive modeling. We employ web scraping to collect financial news and apply transformer-based sentiment analysis models to extract nuanced sentiment insights. Aggregated sentiment scores, combined with technical stock indicators, form a comprehensive feature set to analyze the correlation between market sentiment and stock prices over time. For predictive modeling, we utilize hybrid approaches, including sentiment-augmented time-series models, deep learning architectures, and ensemble methods. A Long Short-Term Memory (LSTM) network is trained to predict stock price movements based on historical stock prices and sentiment scores. Trading signals are generated based on the predicted prices and sentiment trends, providing actionable buy, sell, or hold recommendations. The model's performance is evaluated using metrics such as Mean Absolute Error (MAE) and directional accuracy, ensuring robust and reliable predictions. The findings are integrated into an interactive dashboard, enabling traders and investors to make informed decisions based on sentiment-driven market signals. This data-driven investment tool enhances decision-making, offering actionable insights to traders and investors in volatile stock environments.**

Keywords: **Sentiment Analysis, Stock Market Prediction, Financial News Scraping, LSTM, Transformer Models, Trading Signals.**

I. INTRODUCTION

The stock market is a complex and dynamic financial system influenced by various factors, including economic indicators, geopolitical events, and investor sentiment. In recent years, sentiment analysis has gained significant traction in financial research as a means of evaluating how public perception, derived from news and social media, influences stock price movements. Traditional stock market prediction models have primarily relied on historical price patterns, fundamental analysis, and technical indicators. However, with the rapid dissemination of financial news and opinions across digital platforms, understanding market sentiment has become crucial for developing more accurate predictive models.

This research aims to bridge the gap between sentiment analysis and stock market forecasting by leveraging advanced AI techniques. We employ state-of-the-art sentiment analysis models to extract nuanced financial sentiment from news articles. These sentiment scores, combined with technical stock indicators, are used as key features to analyze their correlation with stock price trends over time.

The methodology of this study is designed to evaluate the influence of financial news sentiment on stock market movements. The workflow consists of multiple interconnected stages, including data collection, preprocessing, sentiment analysis, stock price modeling, and validation. Financial news data is collected through web scraping, while historical stock price data is retrieved using the yfinance API. The data undergoes preprocessing to ensure consistency and reliability, with text cleaning, tokenization, and lemmatization applied to news headlines, and normalization applied to stock prices.

A transformer-based model is employed for sentiment analysis, classifying news articles into positive, neutral, or negative categories. The sentiment scores are aggregated and linked to respective stock tickers and timestamps for further analysis. A Long Short-Term Memory (LSTM) network is trained to predict stock price movements based on historical stock prices and sentiment scores. Trading signals are generated based on the predicted prices and sentiment trends, providing actionable buy, sell, or hold recommendations.

The model's performance is evaluated using metrics such as Mean Absolute Error (MAE) and directional accuracy, ensuring robust and reliable predictions. The findings are integrated into an interactive dashboard, enabling traders and investors to make informed decisions based on sentiment-driven market signals. This research contributes to the field of financial AI, demonstrating how sentiment-aware predictive models can improve market analysis and decision-making in volatile stock environments.

II. LITERATURE REVIEW

The integration of sentiment analysis into stock market prediction has gained significant attention in recent years, as researchers explore the impact of public perception on financial markets. Traditional forecasting models have primarily relied on fundamental and technical analysis, but with the widespread availability of digital financial news and social media data, sentiment-driven approaches have emerged as a promising alternative. Sentiment analysis, a subfield of Natural Language Processing (NLP), classifies textual data into positive, negative, or neutral categories, allowing financial analysts to quantify market sentiment.

A. Sentiment Analysis in Stock Market Prediction

Recent studies have demonstrated the effectiveness of combining sentiment analysis with deep learning techniques for stock market prediction. For instance, Lee and Chen (2024) proposed a framework that integrates numerical/economic data with textual/sentiment data using deep learning models. Their study highlights how technical analysis can be enhanced by incorporating sentiment analysis, leading to more accurate predictions of stock prices and trends. Similarly, Kumar and Gupta (2022) explored the use of sentiment analysis and deep learning models, such as Long Short-Term Memory (LSTM), to predict stock movements in the Indian market. Their findings suggest that integrating historical prices with sentiment data significantly improves prediction accuracy.

B. Advanced NLP Techniques for Sentiment Analysis

The advent of advanced NLP techniques, such as transformer-based models, has revolutionized sentiment analysis in financial markets. Zhu and Yen (2024) employed BERTopic, an advanced NLP technique, to analyze the sentiment of topics derived from stock market comments. By integrating this analysis with deep learning models, they achieved enhanced stock price prediction performance. Another notable study by Gu et al. (2024) developed a FinBERT-LSTM model that combines the FinBERT NLP model with LSTM architecture. Their approach utilizes news categories related to the stock market and historical stock prices to predict stock prices, demonstrating the effectiveness of integrating domain-specific sentiment analysis with deep learning.

C. Sentiment-Augmented Machine Learning Models

Several studies have explored the integration of sentiment analysis with traditional machine learning models to optimize stock price forecasting. Talazadeh and Perakovic (2024) introduced a Sentiment-Augmented Random Forest (SARF) approach, which integrates sentiment analysis using the FinGPT generative AI model with the traditional Random Forest algorithm. Their results show that this hybrid approach significantly improves stock price forecast accuracy. Similarly, Mamillapalli et al. (2024) proposed GRUvader, an optimal gated recurrent unit network that incorporates lexicon-based sentiment analysis to identify sentiment features and their correlation with stock price movements. Their findings highlight the importance of sentiment-informed models in stock market prediction.

D. Investor Sentiment and Market Characteristics

The relationship between investor sentiment and stock market characteristics has also been a focus of recent research. Wang and Zhao (2024) explored the impact of individual investor sentiment on stock market characteristics, such as returns, volatility, and trading volumes, using the AAII sentiment survey index. Their study provides valuable insights into how investor sentiment influences market behavior and highlights the potential of sentiment analysis in understanding market dynamics.

E. Challenges and Future Directions

Despite the advancements in sentiment-driven stock market prediction, several challenges persist. One of the primary concerns is data noise, as financial news articles often contain biased opinions or speculative statements that may not accurately reflect market reality. Additionally, sentiment latency remains a challenge, as the stock market reacts at different speeds to different types of news, making it difficult to determine the exact timing of sentiment-driven price movements. Another limitation is the impact of algorithmic trading, where high-frequency trading algorithms can amplify sentiment-based signals, causing unpredictable fluctuations. Furthermore, sentiment models may struggle during market anomalies or black swan events, such as the COVID-19 pandemic, where investor behavior deviates significantly from historical sentiment patterns. To address these challenges, researchers are exploring real-time sentiment tracking using live data streams, multi-modal learning that incorporates textual, visual, and audio sentiment analysis, and explainable AI (XAI) techniques to enhance model interpretability. The integration of sentiment analysis with blockchain data for cryptocurrency market prediction is another emerging area of research, highlighting the expanding scope of sentiment-driven financial analysis.

Overall, sentiment analysis has proven to be a valuable tool in modern financial forecasting, offering a complementary perspective to traditional stock prediction methods. While existing studies have demonstrated promising results, improving sentiment-based models requires addressing challenges related to data quality, latency, and explainability. Future research should focus on developing adaptive trading strategies, incorporating diverse sentiment sources, and refining real-time sentiment tracking techniques to create more robust and actionable stock market prediction models. The continuous advancement of AI and NLP technologies is expected to further enhance the accuracy and reliability of sentiment-driven financial analysis, making it an indispensable component of market forecasting and investment decision-making.

III. EXISTING SYSTEMS

Traditional stock market prediction models primarily rely on historical stock price movements, technical indicators, and fundamental analysis. Many existing systems use machine learning models such as LSTM, SVM, and Random Forest to predict stock prices based on past trends. However, these models often lack sentiment-driven insights, making them less effective in capturing the impact of market news, investor sentiment, and external factors on stock price fluctuations. Some recent studies have attempted to incorporate sentiment analysis by analyzing news headlines or social media data, but they either use basic sentiment scoring techniques or lack a dedicated financial sentiment model like FinBERT. Moreover, many existing systems struggle with data validation and noise reduction, leading to unreliable predictions. While these approaches provide some level of prediction accuracy, they fail to fully integrate real-time sentiment analysis with machine learning models for better stock trend forecasting. Our proposed system bridges this gap by leveraging FinBERT for financial sentiment analysis and combining it with a robust predictive model, ensuring a more refined and data-driven stock market analysis.

IV. PROPOSED SYSTEM

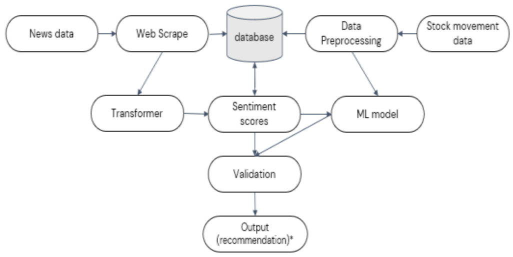

Fig. 1. Proposed system architecture

The proposed system integrates sentiment analysis and machine learning to enhance stock market prediction by leveraging financial news and stock movement data. The system begins by scraping financial news data from various sources and storing it in a database alongside historical stock movement data. The scraped news is processed using FinBERT, a sentiment analysis model, to generate sentiment scores that reflect market sentiment towards specific stocks. Simultaneously, the stock movement data undergoes preprocessing to extract relevant features for analysis. These sentiment scores and processed stock data are then fed into a machine learning model, which predicts future stock trends based on both market sentiment and historical price movements. A validation mechanism ensures that the predictions are reliable before generating an output recommendation, such as buy, sell, or hold signals. This system provides an enhanced decision-making framework by integrating sentiment-driven insights with predictive analytics for investors and traders.

V. METHODOLOGY

The methodology of this study is designed to evaluate the influence of financial news sentiment on stock market movements. The workflow consists of multiple interconnected stages, including data collection, preprocessing, sentiment analysis, stock price modeling, and validation. The following sections outline each step in detail, ensuring a structured and comprehensive approach to sentiment-driven stock market prediction.

A. Data Collection
The dataset comprises two primary sources:

Financial News Data:
News articles and headlines related to stocks are extracted through web scraping from financial sources such as the Economic Times. The scraping process utilizes the BeautifulSoup library to parse HTML content and extract relevant information, including headline text, timestamps, and article URLs. The scraped data is stored in a structured format for further analysis. This ensures a continuous flow of up-to-date financial news, which is critical for capturing real-time market sentiment.

Stock Market Data:
Historical stock price data is retrieved using the yfinance API, which provides access to real-time and historical market data. The dataset includes key variables such as opening price, closing price, high, low, trading volume, and market trends. The stock data is synchronized with the financial news data based on timestamps to analyze the correlation between news sentiment and stock price movements. This synchronization ensures that the sentiment extracted from a given news article corresponds to the stock market activity on the same or following trading day.

B. Data Preprocessing
Before analysis, the collected data undergoes preprocessing to ensure consistency and reliability:

1. Financial News Preprocessing:
Text Cleaning: Irrelevant characters, HTML tags, and special symbols are removed from the news headlines.

Tokenization: The text is split into individual words for further analysis.

Lemmatization: Words are reduced to their base forms to standardize the text and reduce dimensionality. This step ensures that variations of a word (e.g., "running" and "run") are treated consistently during analysis.

2. Stock Market Data Preprocessing:
Handling Missing Values: Missing data points are imputed using interpolation techniques such as forward fill and mean substitution.

Normalization: Stock prices are normalized using the MinMaxScaler to ensure consistent scaling across features. This prevents bias in machine learning models.

Technical Indicators: Additional features such as Simple Moving Averages (SMA) and daily returns are calculated to enhance the predictive power of the model. These indicators provide insights into market trends and volatility.

C. Sentiment Analysis Using Transformer-Based Models
To quantify market sentiment, a transformer-based model is employed to classify news articles into positive, neutral, or negative categories. The sentiment scores are calculated as follows:

$$\text{Sentiment Score} = P(\text{positive}) - P(\text{negative})$$

where P(positive) and P(negative) are the probability scores assigned by the model. The sentiment scores are then aggregated and linked to respective stock tickers and timestamps for further analysis. This numerical representation allows for direct incorporation of sentiment features into the predictive models.

The sentiment trends are visualized using time series plots to observe correlations between sentiment fluctuations and stock price changes. Additionally, feature importance analysis is conducted to determine the impact of sentiment on stock movements compared to traditional financial indicators.

D. Stock Price Prediction Using LSTM
A Long Short-Term Memory (LSTM) network is trained to predict stock price movements based on historical stock prices and sentiment scores. The LSTM model is structured as follows:

Input Layer: Accepts sequences of stock prices and sentiment scores.

Hidden Layers: Two stacked LSTM layers with dropout regularization to prevent overfitting.

Output Layer: A dense layer producing the predicted stock price.

The model is trained using the Mean Squared Error (MSE) loss function and the Adam optimizer. The training process involves splitting the data into training and testing sets, with the model being validated on unseen data to ensure generalizability. The LSTM model is particularly suited for stock market prediction due to its ability to capture long-term dependencies in sequential data.

E. Trading Signal Generation
Trading signals are generated based on the predicted stock prices and sentiment scores. The following rules are applied:

Buy Signal: Generated when the sentiment is positive and the predicted price shows an upward trend.

Sell Signal: Generated when the sentiment is negative and the predicted price shows a downward trend.

Hold Signal: Generated when the sentiment is neutral or the predicted price shows no significant trend.

Additional logic is incorporated to consider technical indicators such as moving averages for trend confirmation, resulting in stronger buy or sell signals. For example:

Strong Buy Signal: Generated when the sentiment is positive, the predicted price shows an upward trend, and the short-term moving average (SMA_50) is above the long-term moving average (SMA_200).

Strong Sell Signal: Generated when the sentiment is negative, the predicted price shows a downward trend, and the short-term moving average (SMA_50) is below the long-term moving average (SMA_200).

F. Model Validation and Performance Evaluation
To assess the model's accuracy, two validation methods are employed:

Backtesting: The trained model is tested on historical stock data to evaluate its ability to predict price movements.

Performance Metrics:
Mean Absolute Error (MAE): Measures the average absolute errors between predicted and actual values.

Root Mean Squared Error (RMSE): Penalizes large errors more than small errors.

Directional Accuracy: Measures the percentage of correctly predicted trends (upward or downward).

The results are visualized using time series plots to compare actual and predicted stock prices, providing insights into the model's performance. These metrics ensure a robust assessment of prediction reliability and help identify areas for improvement.

G. Output and Recommendations
The final output consists of:
Stock Trend Predictions: Predicted stock prices based on historical data and sentiment scores.

Sentiment Influence Analysis: Insights into how financial news sentiment affects stock prices.

Investment Recommendations: Actionable buy, sell, or hold signals based on the model's predictions.

The findings are integrated into an interactive dashboard, enabling traders and investors to make informed decisions based on sentiment-driven market signals. This data-driven investment tool enhances decision-making, offering actionable insights to traders and investors.

VI. RESULTS AND DISCUSSION

Our research introduces a transformative approach to stock market prediction by integrating sentiment analysis with deep learning models. Unlike conventional forecasting techniques that rely solely on historical price data, our model incorporates financial news sentiment to enhance predictive accuracy. The prediction of our model is included as follows, it shows the prediction for taken 'Paytm" stocks from actual to future process and sentiment analysis of the news results plot as follows.

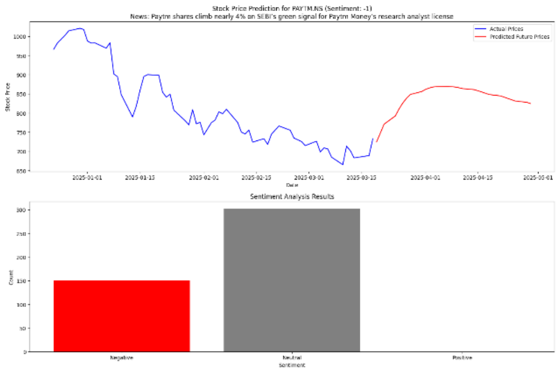

Fig. 2 Stock prediction and sentiment analysis plots

According to the results, the person can hold or sell their stocks in Paytm. The results validate the hypothesis that market sentiment plays a crucial role in stock price movements. By utilizing a sentiment-aware Long Short-Term Memory (LSTM) model, we observed that fluctuations in sentiment scores—particularly those extracted from financial news articles—had a direct correlation with stock price trends. This insight underscores the profound impact of investor psychology and market sentiment on trading behavior.

One of the key achievements of this research is the model's ability to capture directional trends with enhanced precision, outperforming traditional time-series models. The integration of sentiment analysis improved the model's understanding of market behavior, reducing the unpredictability often associated with stock price movements. Additionally, error metrics demonstrated the advantage of incorporating sentiment-driven features over standard numerical models.

By validating sentiment analysis as a critical component of stock price forecasting, this research bridges the gap between financial linguistics and quantitative market analysis. These findings reinforce the potential of sentiment-aware AI models to redefine investment strategies, risk assessment, and financial decision-making in a rapidly evolving economic landscape.

VII. CONCLUSION

This study establishes a new frontier in stock market prediction, proving that integrating sentiment analysis with deep learning enhances forecasting accuracy. Unlike traditional models that analyze only past stock prices, our approach reveals that news sentiment is a powerful predictor of future market movements.

Our findings have profound implications for financial institutions, traders, and AI-driven investment strategies. The ability to assess market sentiment in real-time can lead to better risk management, more informed trading decisions, and improved market predictions. This research highlights the necessity of moving beyond purely numerical models and adopting multi-dimensional AI approaches for financial forecasting. However, this study also opens doors for further exploration. Future advancements will focus on:

Real-time market intelligence – Enhancing AI models to process live financial news and social media data streams for instant sentiment evaluation.

Multi-factor analysis – Expanding predictive models to incorporate economic indicators, geopolitical events, and investor sentiment for more holistic forecasting.

Adaptive learning models – Developing AI-driven trading strategies that dynamically adjust based on evolving market sentiment.

By demonstrating the effectiveness of sentiment-driven stock prediction, this research lays the foundation for the next generation of AI-powered financial forecasting, offering a paradigm shift in the way markets are analyzed and predicted.

VIII. FUTURE ENHANCEMENT

Future enhancements for this research will focus on expanding the predictive capabilities of sentiment-aware stock forecasting models by integrating real-time financial news and social media sentiment analysis. The incorporation of multi-source data streams, including macroeconomic indicators, geopolitical events, and investor sentiment from diverse platforms, will further improve forecasting accuracy. Additionally, adaptive AI models that continuously learn and adjust based on evolving market trends will be developed to enhance long-term reliability. Advancements in explainable AI (XAI) will also be explored to provide greater transparency in decision-making, allowing investors and financial institutions to understand the underlying factors influencing predictions. Finally, the integration of reinforcement learning-based trading strategies will help create dynamic, self-optimizing models capable of making real-time investment decisions, thereby pushing the boundaries of AI-driven financial forecasting.

REFERENCES

[1] Lee, K. H., & Chen, Y. Z. (2024). Deep Learning for Stock Market Prediction Using Sentiment and Technical Analysis. *Springer Link*.

[2] Kumar, S., & Gupta, A. (2022). Stock Price Prediction Using Sentiment Analysis and Deep Learning for Indian Markets. *arXiv*.

[3] Zhu, E., & Yen, J. (2024). BERTopic-Driven Stock Market Predictions: Unraveling Sentiment Insights. *arXiv*.

[4] Gu, W., et al. (2024). Predicting Stock Prices with FinBERT-LSTM: Integrating News Sentiment Analysis. *arXiv*.

[5] Talazadeh, S., & Perakovic, D. (2024). SARF: Enhancing Stock Market Prediction with Sentiment-Augmented Random Forest. *arXiv*.

[6] Mamillapalli, A., et al. (2024). GRUvader: Sentiment-Informed Stock Market Prediction. *arXiv*.

[7] Wang, J., & Zhao, L. (2024). Sentiment and Stock Characteristics: Comprehensive Study of Individual Investor Sentiment. *Sciendo*.

[8] Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3), 1139-1168.

[9] Zhang, Y., & Skiena, S. (2010). Trading strategies to exploit blog and news sentiment. *Proceedings of ICWSM*.

[10] Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, 21(2), 129-151.

[11] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. *Proceedings of KDD*.

[12] Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. *Whitepaper.*