INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 3, March 2026 | ISSN: 2456-4184 | IJNRD.ORG

STOCK PREDICITION BASED ON SENTIMETAL ANALYSIS

Kuldeep Patil[*1], Tejas Jadhav[*2], Shreyash Jadhav [*3], Umesh Mantale [*4]

[*1,2,3]*Student, Department of Computer Engineering Terna Engineering College, Nerul, Navi Mumbai, Maharashtra, India*

[*4]*Lecturer, Department of Computer Engineering, Terna Engineering College, Nerul, Navi Mumbai, Maharashtra, India*

Abstract: This survey examines and consolidates recent research on stock market prediction methods that utilize Sentiment Analysis (SA). Traditional approaches, based on technical and fundamental analysis, often fail to capture the impact of investor sentiment, which is derived from unstructured data sources such as financial news and social media. To provide a clearer understanding of the field, this study categorizes existing work based on data sources, sentiment analysis techniques—including lexicon-based methods, deep learning models, and Large Language Models—and predictive algorithms such as LSTM, RNN, and SVM. The overall findings indicate that incorporating sentiment-based features significantly enhances prediction accuracy and helps in modelling the complex and non-linear behaviour of financial markets. The survey also identifies key challenges, including data noise, market non-stationarity, and limited model interpretability, and suggests potential directions for future research to overcome these issues.

Index Terms - Natural Language Processing (NLP), Investor Sentiment, Financial News, Social Media Data, Deep Learning, Forecasting.

I. INTRODUCTION

Predicting stock prices has been a long-standing challenge explored by researchers in both finance and computer science, traditionally relying on technical indicators and fundamental analysis. In recent years, the expansion of digital platforms has brought attention to the importance of investor sentiment and collective behaviour, which are often embedded in large volumes of unstructured textual data. Sentiment Analysis (SA), supported by advancements in Natural Language Processing (NLP), has emerged as an effective method for quantifying these subjective influences and enhancing forecasting models.

In the modern financial landscape, investors actively track real-time stock movements across major exchanges such as the Indian markets (NSE/BSE) and the US markets (NYSE/NASDAQ). Although traditional metrics like price patterns and trading volume remain relevant, they fail to fully represent the psychological and emotional drivers behind market fluctuations. Public sentiment derived from financial news, social media platforms, and online discussions plays a critical role in influencing stock behaviour.

Sentiment Analysis allows automated systems to process vast amounts of textual data and extract meaningful sentiment indicators that may precede or align with market trends. By combining these sentiment signals with real-time stock data from both Indian and international markets, more robust and accurate prediction models can be developed. This study integrates data from financial news APIs and social media sources to create a sentiment-aware stock prediction system, offering practical insights for investors, analysts, and researchers.

II. LITERATURE REVIEW

Recent advancements in stock market prediction highlight the increasing use of artificial intelligence, machine learning, and sentiment analysis to improve forecasting performance. Research by Bollen et al. and Mittal demonstrates that sentiment extracted from social media platforms, particularly Twitter, can serve as an effective indicator of market trends. These studies emphasize that emotional and behavioural factors significantly influence stock price movements, encouraging the integration of sentiment-based features with traditional financial analysis.

Natural Language Processing techniques have played a central role in extracting sentiment from textual data. Early contributions in opinion mining introduced structured approaches for analysing user opinions, while subsequent developments provided practical tools and frameworks for implementing NLP systems. Further progress in neural network-based language models and word embeddings has improved the ability to understand semantic relationships, leading to more accurate sentiment classification.

Machine learning models have also contributed significantly to stock prediction by enabling advanced pattern recognition and time-series analysis. Foundational theories in statistical learning support the development of classification algorithms, while deep

learning architectures such as Long Short-Term Memory (LSTM) networks are widely used for handling sequential financial data. Modern frameworks facilitate efficient implementation of these models, improving scalability and predictive performance.

Additionally, reinforcement learning techniques have introduced adaptive strategies for dynamic decision-making in trading environments. Despite these advancements, most existing studies focus on isolated components such as sentiment extraction or predictive modelling. There remains a gap in developing an integrated system that combines sentiment analysis, stock prediction, and intuitive visualization within a single platform. This limitation motivates the development of a unified and comprehensive framework.

III. PROPOSED SYSTEM

The proposed system is an intelligent stock prediction platform that leverages sentiment analysis to enhance market forecasting. It is designed to analyse both textual and numerical data to generate accurate stock movement predictions while minimizing manual effort. The system integrates data acquisition, sentiment processing, predictive modelling, and visualization into a unified framework.

The architecture consists of several interconnected components, including a user interface, backend processing unit, sentiment analysis module, prediction engine, and database. A web-based interface developed using Flask allows users to interact with the system, monitor stock trends, and view prediction results in real time.

The system collects financial data from sources such as Yahoo Finance, including historical stock prices and relevant market indicators. In parallel, textual data is gathered from financial news platforms and social media sources using APIs. This textual information is analysed to extract public sentiment related to specific stocks or market conditions.

A sentiment processing module evaluates the collected text using Natural Language Processing techniques and assigns sentiment scores representing positive, negative, or neutral opinions. These scores are combined with stock market features such as price movements and trading volume.

The prediction module utilizes machine learning algorithms to analyse the integrated dataset and generate actionable outputs such as Buy, Sell, or Hold signals. The system also includes visualization features to display sentiment trends, stock price movements, and prediction outcomes. All processed data and user interactions are stored in a database for efficient retrieval and analysis.
The overall system architecture of the proposed model is illustrated in Fig. 1.

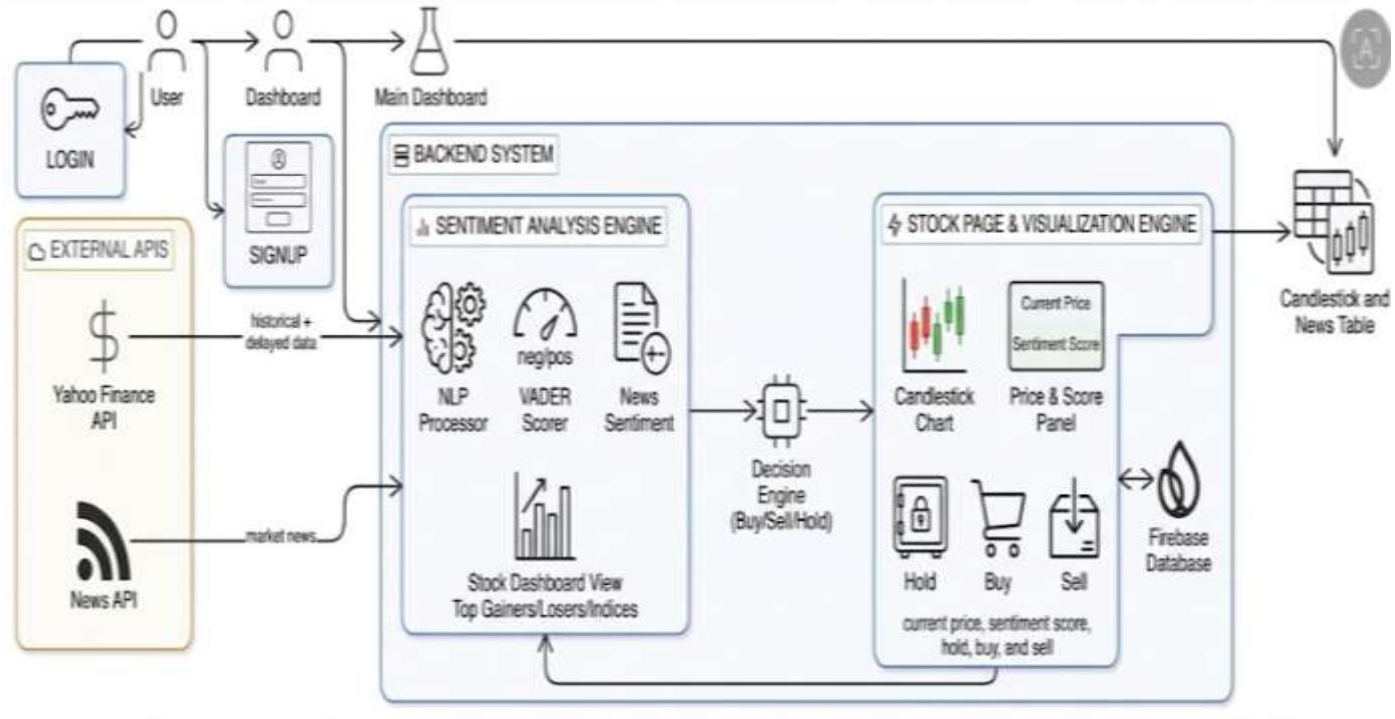

Fig. 1. System Architecture of Sentiment-Based Stock Prediction System

IV. METHODOLOGY

The system follows a systematic approach for predicting stock market behavior by combining sentiment analysis with financial data analytics. The methodology consists of several stages, including data collection, preprocessing, sentiment extraction, feature integration, model development, and result visualization.

4.1 Data Collection

The system gathers two types of data: textual data and numerical stock data. Financial news articles and social media content are collected using APIs to capture market sentiment. At the same time, stock price data, including historical and recent values, is obtained from platforms such as Yahoo Finance.

4.2 Data Preprocessing

The collected textual data undergoes cleaning to remove noise such as special characters, stop words, and irrelevant content. Similarly, stock data is checked for missing or inconsistent values. The processed data is then structured into a format suitable for analysis.

4.3 Sentiment Analysis

The cleaned textual data is analysed using sentiment analysis techniques such as VADER or machine learning-based models. Each text entry is assigned a sentiment score indicating its polarity. These scores are aggregated over time to compute an overall sentiment indicator for each stock.

4.4 Feature Engineering and Integration

The sentiment scores are combined with stock-related features such as opening price, closing price, highest and lowest values, trading volume, and technical indicators like moving averages. The integration is performed based on timestamps to ensure alignment between sentiment and market data.

4.5 Predictive Modelling

Machine learning algorithms, including Logistic Regression, Random Forest, and Support Vector Machine (SVM), are applied to the integrated dataset. These models learn patterns between sentiment trends and stock price movements to predict future market behaviour. The output is classified into Buy, Sell, or Hold recommendations.

4.6 Model Evaluation

The performance of the models is assessed using evaluation metrics such as accuracy, precision, recall, and F1-score. The dataset is divided into training and testing subsets to validate model reliability and effectiveness.

4.7 Visualization and User Interface

The results are presented using visualization tools to display stock price trends, sentiment variations, and prediction outcomes. A web application built with Flask provides users with an interactive interface to explore data and predictions in real time.

4.8 Data Storage and Management

All relevant data, including stock information, sentiment scores, and prediction results, is stored in a database system. This ensures efficient data management and supports future analysis.

This methodology creates a continuous workflow that integrates sentiment analysis with stock prediction, enabling more accurate and data-driven investment decisions. The overall workflow of the system is shown in Fig. 2.

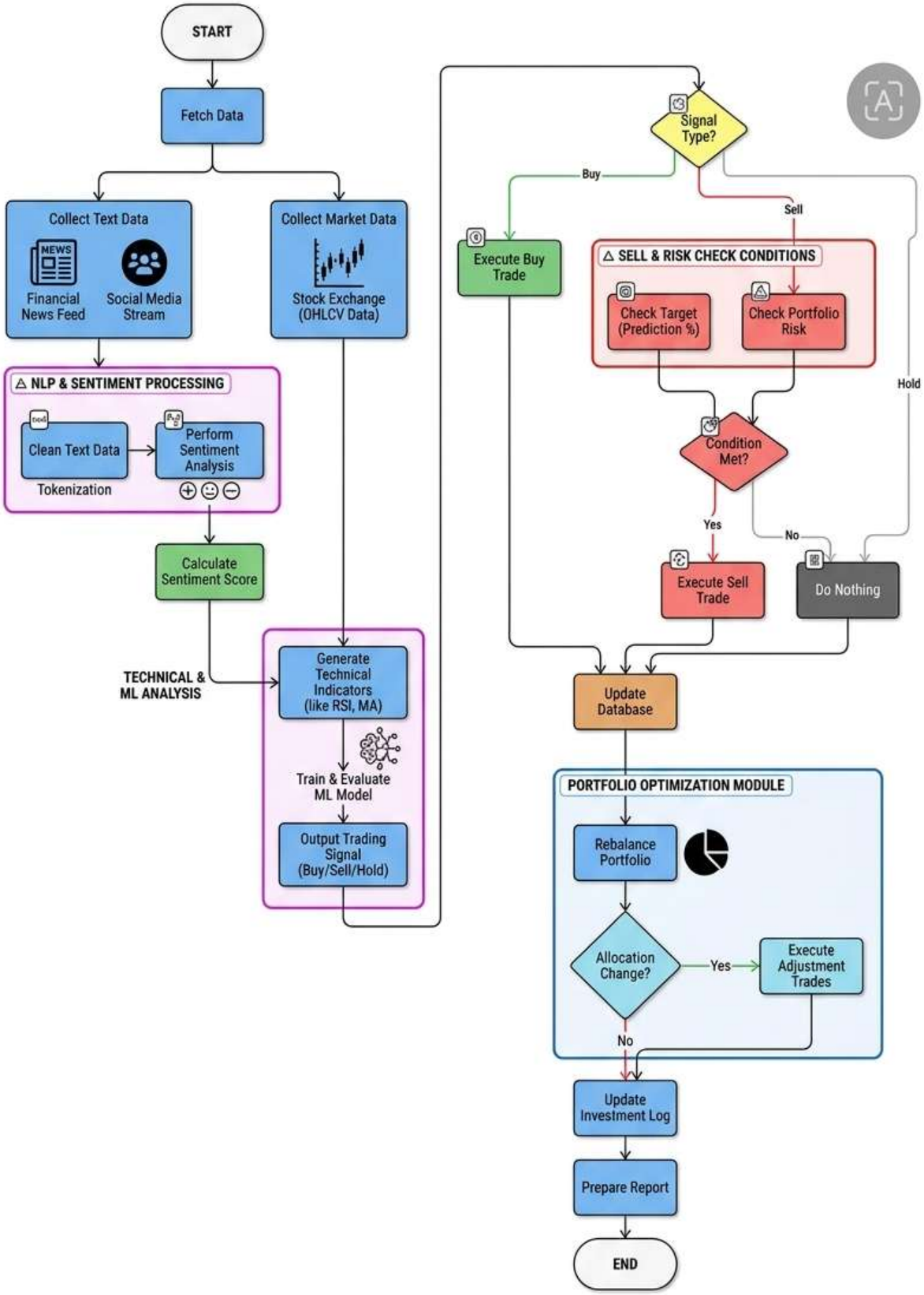

Fig. 2. Flowchart Of Stock Predication Based On Sentimental Analysis

V. ALGORITHMS

The proposed system applies a combination of sentiment analysis and machine learning algorithms to predict stock market behaviour. These algorithms collectively process textual sentiment data and numerical stock features to generate reliable Buy, Sell, or Hold predictions.

5.1 Sentiment Analysis Algorithm

This algorithm extracts sentiment information from financial news and social media data to quantify public opinion.
Algorithm Steps:

1. Collect textual data related to selected stocks
2. Clean the text by removing stop words, symbols, and irrelevant content
3. Tokenize and normalize the text
4. Apply a sentiment analysis model (e.g., VADER or ML-based classifier)
5. Assign sentiment scores (positive, negative, neutral)
6. Compute a combined sentiment score (compound value)
7. Aggregate scores over a fixed time period to obtain overall sentiment

Decision Criteria:

- Positive sentiment → Indicates potential upward movement (Buy)
- Negative sentiment → Indicates potential downward movement (Sell)
- Neutral sentiment → No strong signal (Hold)

5.2 Feature Integration Algorithm

This algorithm combines sentiment scores with stock market data to form a unified dataset.
Algorithm Steps:

1. Input sentiment scores and stock price data
2. Extract stock features such as Open, Close, High, Low, and Volume
3. Align sentiment data with stock data based on timestamps
4. Normalize and scale features if required
5. Create a structured dataset for model training

5.3 Machine Learning Prediction Algorithm

This algorithm uses supervised learning techniques to predict stock movements.
Algorithm Steps:

1. Input integrated dataset
2. Split data into training and testing sets
3. Train models such as Logistic Regression, Random Forest, and SVM
4. Learn relationships between sentiment and stock features
5. Generate predictions on test data

Decision Criteria:

- Predicted upward trend → Buy
- Predicted downward trend → Sell
- No clear trend → Hold

5.4 Random Forest Algorithm

Random Forest is used as a primary model due to its ability to capture complex relationships.
Algorithm Steps:

1. Create multiple decision trees using random subsets of data
2. Train each tree independently
3. Combine predictions using majority voting
4. Output final classification

Decision Criteria:

- Majority votes Buy → Buy
- Majority votes Sell → Sell
- Mixed predictions → Hold

5.5 Model Evaluation Algorithm

This algorithm measures the effectiveness of prediction models.
Algorithm Steps:

1. Compare predicted values with actual outcomes
2. Compute evaluation metrics such as Accuracy, Precision, Recall, and F1-score
3. Analyse performance across different models

4. Select the best-performing model

5.6 Visualization Algorithm

This module presents results in an understandable graphical format.
Algorithm Steps:

1. Input prediction results and stock data
2. Generate charts for stock prices and sentiment trends
3. Display Buy/Sell/Hold signals visually
4. Update graphs dynamically based on user input

5.7 Decision-Making Algorithm (Integrated Logic)

This algorithm combines outputs from sentiment and machine learning models.
Algorithm Steps:

1. Collect outputs from sentiment analysis and prediction model
2. Assign weights or priority to signals
3. Combine signals to produce a final decision

Decision Criteria:
• Strong positive sentiment + upward prediction → Buy
• Strong negative sentiment + downward prediction → Sell
• Conflicting or weak signals → Hold

5.8 Data Storage and Update Algorithm

This algorithm manages system data efficiently.
Algorithm Steps:

1. Store stock data, sentiment scores, and predictions in database
2. Update records periodically with new data
3. Retrieve data when requested by user interface
4. Ensure consistency and accuracy of stored data

VI. RESULTS AND DISCUSSION

The proposed sentiment-based stock prediction system was evaluated using data from multiple stocks across different time periods. The system effectively combined historical stock data with sentiment information extracted from financial news and social media sources. The sentiment analysis module successfully identified positive, negative, and neutral trends, which were then integrated with numerical market features for prediction.

Machine learning models such as Logistic Regression, Support Vector Machine, and Random Forest were tested to analyse prediction performance. Among these, the Random Forest model demonstrated better accuracy due to its capability to capture complex and non-linear relationships between sentiment and stock data. The integration of sentiment features significantly improved prediction outcomes compared to models relying solely on historical price data.

The system generated clear Buy, Sell, and Hold signals based on combined sentiment and market indicators. Visualization tools helped in understanding stock trends and sentiment fluctuations over time, making the system more interpretable and user-friendly. Overall, the results indicate that incorporating sentiment analysis enhances prediction reliability and provides meaningful insights into market behaviour.

VII. CONCLUSION AND FUTURE SCOPE

A. Conclusion

The proposed system presents an intelligent approach to stock market prediction by integrating sentiment analysis with machine learning techniques. Unlike traditional methods that rely only on numerical data, this system incorporates public opinion derived from textual sources, enabling a more comprehensive understanding of market dynamics. The use of NLP-based sentiment extraction combined with predictive models improves the accuracy of forecasting and supports better decision-making. Additionally, the inclusion of visualization and a user-friendly interface makes the system practical for both researchers and investors.

B. Future Scope

The system can be further enhanced by incorporating advanced deep learning models such as Long Short-Term Memory (LSTM) and Transformer-based architectures for improved prediction accuracy. Integration of real-time streaming data from multiple sources can make the system more responsive to sudden market changes. Expanding the system to include additional data types, such as economic indicators and global financial news, can further improve performance. Future improvements may also include mobile application support, personalized investment recommendations, and advanced risk management strategies to make the platform more comprehensive and scalable.

REFERENCES

[1] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.
[2] A. Mittal and A. Goel, "Stock prediction using Twitter sentiment analysis," *International Journal of Computer Applications*, vol. 11, no. 1, pp. 1–5, 2012.
[3] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. ACM SIGKDD*, 2004, pp. 168–177.
[4] Y. Goldberg, "A primer on neural network models for natural language processing," *Journal of Artificial Intelligence Research*, vol. 57, pp. 345–420, 2016.
[5] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.
[6] S. Hochreiter and J. Schmid Huber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[7] F. Chollet, *Deep Learning with Python*. Manning Publications, 2017.
[8] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer, 1995.
[9] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, 2018.
[10] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*. O'Reilly Media, 2009.