© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1719159

Stock Market Prediction with Sentiment Analysis

JAYSHREE PANSARE[1], NARINDER SINGH[2], VINAY KOTWAL[3], KALHAN KOUL[4], ADITYA GUNDETI[5]
[1, 2, 3, 4, 5]*Department of Computer Engineering, MES Wadia College of Engineering, Pune, India*

Abstract- ***Stock market prediction is a complex and challenging task due to the volatile and non-linear nature of financial markets. Traditional prediction models mainly rely on historical price data, ignoring the influence of public sentiment and financial news. This research proposes a hybrid model that combines machine learning techniques with sentiment analysis to improve stock price prediction accuracy. Historical stock data and financial news headlines are collected and preprocessed. Sentiment scores are extracted using Natural Language Processing (NLP) techniques and combined with stock market indicators. A Long Short-Term Memory (LSTM) model is implemented for prediction. Experimental results show that integrating sentiment analysis with historical stock data improves prediction accuracy compared to models that rely solely on price data. The proposed system demonstrates the importance of textual information in financial forecasting.***

Keywords: Stock Market Prediction, Sentiment Analysis, LSTM, Machine Learning, NLP, Time Series Forecasting

I. INTRODUCTION

Stock markets play a crucial role in the global economy by facilitating capital formation and wealth creation. Investors aim to predict stock price movements to maximize profits and minimize financial risks. However, stock prices are highly volatile and influenced by multiple factors such as company performance, economic indicators, political events, global crises, and public sentiment.

Traditional stock prediction methods mainly rely on historical price trends and technical indicators using statistical models such as Autoregressive Integrated Moving Average (ARIMA) and linear regression. Although these models are effective for analyzing time-series data, they fail to capture non-linear patterns and external influences that significantly impact stock prices. Moreover, these approaches do not consider qualitative factors like news reports, investor opinions, and social media discussions, which often drive short-term market fluctuations.

With the rapid growth of social media platforms, financial blogs, and online news portals, a large amount of unstructured textual data is generated every day. This textual information reflects public opinion and investor sentiment, which can strongly influence market behavior. Sentiment analysis, a subfield of Natural Language Processing (NLP), enables the extraction of emotional tone (positive, negative, or neutral) from textual data. By analyzing sentiment from financial news and social media posts, it becomes possible to estimate market mood and incorporate it into predictive models.

Recent advancements in Machine Learning (ML) and Deep Learning (DL) techniques have significantly improved stock price forecasting. In particular, Long Short-Term Memory (LSTM) networks, a type of Recurrent Neural Network (RNN), are highly effective for time-series prediction because they can learn long-term dependencies and handle sequential data efficiently. LSTM models overcome the vanishing gradient problem and are well-suited for capturing complex, non-linear patterns in financial datasets.

This paper proposes a hybrid prediction system that integrates historical stock price data with sentiment analysis derived from financial news and social media sources. The sentiment scores are combined with numerical market indicators to create an enhanced feature set. An LSTMbased deep learning model is then applied to predict future stock price movements. The main objective of this research is to improve prediction accuracy by incorporating both quantitative and qualitative factors into a unified framework.

The proposed system aims to:

- Analyze historical stock price data.

- Extract sentiment scores from textual data using NLP techniques.
- Combine numerical and sentiment features.
- Train and evaluate an LSTM-based prediction model.

By integrating sentiment analysis with deep learning techniques, this research contributes toward building a more robust and intelligent stock market prediction system.

II. REVIEW OF LITERATURE RESEARCH

Recent research has focused on integrating sentiment analysis with machine learning techniques to enhance stock market prediction accuracy.

Nagaraj et al. [1] utilized historical stock price data along with sentiment extracted from news articles and social media platforms as input. Their methodology involved applying machine learning models to a hybrid feature set combining numerical and textual sentiment data. The system produced stock trend predictions and achieved up to 97% accuracy, particularly during volatile market conditions.

Agrawal and Mukherjee [2] considered financial news, Reddit discussions, and historical stock data as input features. They employed a BERT-based transformer model to extract contextual sentiment representations from textual data, which were then integrated with historical market data for prediction. Their deep learning-based approach generated stock movement predictions and achieved 98.92% accuracy, outperforming traditional time-series models that rely solely on historical prices.

Sridhara et al. [3] incorporated investor sentiment data, lexical financial text, and historical stock prices as input to their system. Their methodology combined Latent Dirichlet Allocation (LDA) for topic modeling with an LSTM neural network for time-series prediction. Additionally, the Sparrow Search Algorithm (SSA) was used to optimize LSTM parameters. The output of their model was an improved stock price prediction system with enhanced accuracy due to optimized feature selection and parameter tuning.

Zhu and Yen [4] utilized financial textual documents and market-related news as input data. They implemented a BERTopic framework that integrates topic modeling with BERT embeddings to capture fine-grained sentiment features. These features were fed into a deep learning prediction model. The system produced improved stock forecasting results, demonstrating that contextual topicbased sentiment representation enhances prediction performance.

Annalakshmi et al. [5] used financial news articles and historical stock data as inputs. Their methodology applied FinBERT, a financial-domain-specific language model, to extract sentiment scores from news content. These sentiment features were combined with historical numerical data and processed using an LSTM network for forecasting. The hybrid deep learning model produced stock price predictions with approximately 95% accuracy.

Hotasi and Satya [6] considered news sentiment data and Indonesian stock market historical data as inputs. They applied the Naïve Bayes algorithm for sentiment classification and used the K-Nearest Neighbor (KNN) algorithm for stock trend prediction. While the system achieved 90% training accuracy, the validation accuracy dropped to 43.5%, indicating overfitting and limited generalization capability.

Mujhid et al. [7] used Twitter data along with historical market data as inputs. Sentiment scores were extracted using the VADER sentiment analysis tool. They compared LSTM and BiLSTM models for prediction. The output analysis showed that LSTM performed slightly better when sentiment features were included, whereas BiLSTM showed better performance when sentiment input was excluded.

Xu and Keseli [8] incorporated tweet sentiment data, technical indicators, and historical stock prices as input features. Their methodology involved an attention-based LSTM model designed to capture important temporal dependencies and relevant sentiment signals. The output demonstrated improved stock prediction accuracy, particularly when tweets posted between market close and the next market opening were used, indicating higher predictive influence during that time window.

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1719159

Overall, these studies highlight that hybrid models integrating sentiment analysis with deep learning architectures such as LSTM, attention mechanisms, and transformer-based models significantly enhance stock market prediction performance compared to traditional statistical approaches. However, challenges such as overfitting, real-time implementation, and cross-market generalization remain areas for further research.

III. PROPOSED SYSTEM ARCHITECTURE

The proposed system integrates historical stock market data with sentiment analysis of financial news to improve stock price prediction accuracy. The architecture consists of multiple sequential modules, including data collection, preprocessing, feature engineering, sentiment analysis, model training, evaluation, and final prediction. The workflow of the proposed system is illustrated in Fig. 1.

Fig. 1. Workflow of Proposed Stock Market Prediction System

A. Data Collection

The first stage involves collecting two types of data:

1. Stock Market Data

Historical stock data is collected using the Yahoo Finance API. The dataset includes:

- Open price
- High price
- Low price
- Close price
- Trading
- volume

2. News Dataset

Financial news headlines related to the selected company (Apple Inc. in this case) are collected. These textual data sources help capture public opinion and investor sentiment.

B. Data Preprocessing

Raw data often contains missing values, noise, and inconsistencies. Therefore, preprocessing is performed as follows:

- Handling missing values
- Cleaning irrelevant or duplicate entries
- Normalizing numerical features using MinMaxScaler
- Cleaning text data by removing URLs, special characters, and converting text to lowercase

This step ensures that both numerical and textual data are structured and suitable for further processing.

C. Feature Engineering

Feature engineering enhances the predictive capability of the model by creating additional meaningful attributes from raw data. The following features are generated:

- Moving Average (MA_10, MA_50)
- Daily Returns
- Trading Volume analysis
- Technical indicators
- Sentiment score

These engineered features provide deeper insights into market trends and price momentum.

D. Sentiment Analysis

Sentiment analysis is performed on financial news headlines using the TextBlob library. The process includes:

- Text preprocessing
- Polarity score extraction (range: -1 to +1)
- Classification into Positive, Negative, or Neutral sentiment
- Calculation of daily average sentiment score

The daily sentiment score is then merged with stock market data to form a combined dataset.

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1719159

E. Combined Dataset Formation

After preprocessing and feature extraction, numerical stock indicators and sentiment scores are merged into a unified dataset. This combined dataset serves as input to the prediction model.

The feature vector includes:

- Close price
- Moving averages
- Returns
- Volume
- Sentiment score

F. Model Training (LSTM)

A Long Short-Term Memory (LSTM) neural network is used for stock price prediction. LSTM is chosen because it effectively captures time-series dependencies and long-term patterns in sequential data.

The dataset is divided into:

- 80% training data
- 20% testing data

The model architecture includes:

- Multiple LSTM layers
- Dropout layers for regularization
- Dense layers for final prediction

The model is trained using the Adam optimizer with Mean Squared Error (MSE) as the loss function.

G. Model Evaluation

Model performance is evaluated using the following metrics:

- Root Mean Square Error (RMSE)
- Mean Absolute Error (MAE)
- R^2 Score
- Mean Absolute Percentage Error (MAPE)
- Accuracy (derived from MAPE) Graphical analysis includes:
- Actual vs Predicted Price comparison
- Error distribution histogram
- Scatter plot of actual vs predicted values
- Sentiment composition pie chart

H. Final Prediction Output

The final output of the system provides:

- Predicted stock prices
- Trend analysis
- Model performance metrics
- Visualization dashboards

This architecture demonstrates that combining sentiment analysis with technical indicators significantly improves stock price forecasting performance.

IV. TOOLS AND TECHNOLOGIES USED

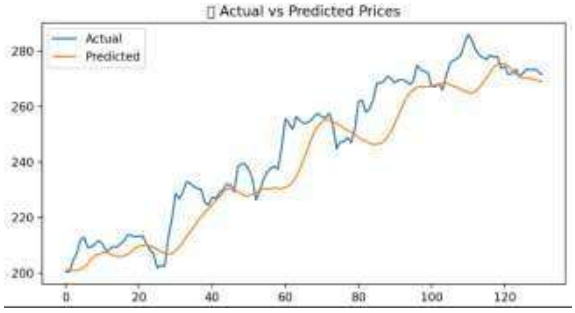
Software Requirements

- Python 3.x
- Streamlit (for dashboard development)
- Pandas (data manipulation)
- NumPy (numerical computations)
- Scikit-learn (data preprocessing and evaluation metrics)
- Keras (LSTM model implementation)
- Matplotlib (data visualization)
- yFinance (stock data collection)
- TextBlob (sentiment analysis)
- Regular Expressions (re) (text cleaning)
- Hardware Requirements
- Processor: Intel i5 or above
- RAM: Minimum 8GB
- Storage: 10GB free space
- Operating System: Windows

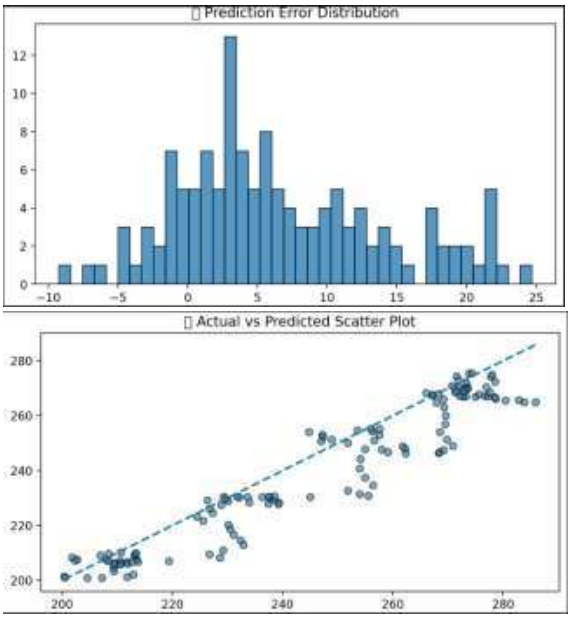
V. RESULTS AND ANALYSIS

The model performance was evaluated using:

- Mean Absolute Error (MAE): 7.53
- Root Mean Square Error (RMSE): 9.86
- Accuracy: 96.97%



© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1719159



VI. ADVANTAGES OF PROPOSED SYSTEM

- Integrates historical stock data with sentiment analysis for improved accuracy.
- Uses LSTM model to capture time-series patterns effectively.
- Considers external factors like financial news and investor sentiment.
- Reduces prediction error compared to traditional methods.
- Scalable and can be extended to real-time and multi-stock prediction.

VII. FUTURE WORK

In future work, the proposed system can be enhanced by incorporating advanced Transformer-based models such as BERT to achieve more accurate and context-aware sentiment analysis of financial text.

Real-time Twitter sentiment integration can also be implemented to capture live market reactions and investor behavior.

Additionally, the framework can be extended to support multi-stock portfolio prediction, enabling broader investment analysis instead of focusing on a single company.

Finally, the system can be deployed as a web-based application with automated data updates and interactive visualizations, making it more practical and accessible for real-world investors and financial analysts.

VIII. CONCLUSION

This research presented a hybrid approach for stock market prediction by integrating machine learning techniques with sentiment analysis. Traditional stock prediction models rely mainly on historical numerical data and technical indicators, which may not fully capture the influence of external factors such as news and public opinion. To address this limitation, the proposed system combines historical stock price data with sentiment scores extracted from financial news headlines.

In this study, stock market data including closing price, trading volume, moving averages, and returns were collected and preprocessed. Sentiment analysis was performed using natural language processing techniques to compute daily polarity scores ranging from negative to positive values. These sentiment scores were merged with technical indicators to form a combined dataset. A Long Short-Term Memory (LSTM) neural network was implemented to model time-series dependencies and predict future stock prices.

Experimental results demonstrate that incorporating sentiment information improves prediction performance compared to models that rely solely on historical price data. The evaluation metrics, including RMSE, MAE, R^2 score, and MAPE, indicate that the proposed hybrid model achieves better accuracy and reduced prediction error. Graphical analysis of actual versus predicted prices further validates the effectiveness of the approach.

Although the stock market remains inherently unpredictable due to its dynamic and volatile nature, the integration of sentiment analysis with deep learning models provides a more comprehensive framework for intelligent financial forecasting. The proposed system can assist investors in understanding market trends and making informed decisions.

Future work may include incorporating real-time social media data, applying advanced transformer-based models such as BERT for improved sentiment extraction, and extending the framework to multi-stock portfolio prediction and real-time deployment.

REFERENCES

[1] X. Zhang, Y. Qu, and S. Li, "Stock Market Prediction via Deep Learning Techniques: A Survey," IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 2, pp. 1510–1525, 2024, doi: 10.1109/TNNLS.2023.3289736

[2] S. Picasso, S. Merello, Y. Ma, L. Oneto, and E. Cambria, "Technical Analysis and Sentiment Embeddings for Market Trend Prediction," Expert Systems with Applications, vol. 135, pp. 60–70, 2019, doi: 10.1016/j.eswa.2019.06.014.

[3] M. Nabipour, P. Nayyeri, H. Jabani, S. Shahab, and A. Mosavi, "Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data: A Comparative Analysis," IEEE Access, vol. 8, pp. 150199–150212, 2020, doi: 10.1109/ACCESS.2020.3015966.

[4] R. Ranco, I. Bordino, G. Bormetti, G. Caldarelli, F. Lillo, and M. Treccani, "Coupling News Sentiment with Web Browsing Data Predicts Intra-Day Stock Prices," PLOS ONE, vol. 11, no. 1, p. e0146576, 2016, doi: 10.1371/journal.pone.0146576.

[5] 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, doi: 10.1016/j.jocs.2010.12.007.

[6] T. H. Nguyen, K. Shirai, and J. Velcin, "Sentiment Analysis on Social Media for Stock Movement Prediction," Expert Systems with Applications, vol. 42, no. 24, pp. 9603–9611, 2015, doi: 10.1016/j.eswa.2015.07.052.

[7] H. Li, Y. Shen, and Y. Zhu, "Stock Price Prediction Using Attention-Based Multi-Input LSTM," in 2018 IEEE Asia-Pacific Conference on Antennas and Propagation (APCAP), Oct. 2018, pp. 1–3, doi: 10.1109/APCAP.2018.8538101.

[8] D. M. Q. Nelson, A. C. M. Pereira, and R. A. de Oliveira, "Stock Market's Price Movement Prediction with LSTM Neural Networks," in 2017 International Joint Conference on Neural Networks (IJCNN), May 2017, pp. 1419–1426, doi: 10.1109/IJCNN.2017.7966019.

[9] P. C. Yauney and P. Rajapakse, "Predicting Stock Prices Using NLP and Deep Learning," in 2019 IEEE Tenth International Conference on Intelligent Computing and Information Systems (ICICIS), Dec. 2019, pp. 113–118, doi: 10.1109/ICICIS46948.2019.9014774.

[10] A. Araci, "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models," arXiv:1908.10063 [cs.CL], Aug. 2019.