IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, April 2026
Impact Factor: 8.2

Hybrid Stock Market Analysis Using Sentiment and Technical Indicators with NLP

Sanchit Kharade, Onkar Zagade, Ajit Patil, Prathmesh Bhosale, Sunil Mane
Computer Engineering
Bharati Vidyapeeth College of Engineering, Lavale Pune,
sanchitkharade1120@gmail.com, onkarzagade9822@gmail.com
ajit.r.patil@bharatividyapeeth.edu, prathmeshb275@gmail.com, sunil.mane-coel@bvp.edu.in

Abstract: *Stock returns are hard to forecast, as they are driven not only by past price data, but also by sentiment prevailing in the market. Statistical analysis of financial data largely relies on price, volume, and momentum. These methods may not be fully reliable for capturing sudden reactions to news or the mood of traders. Sentiment-based systems also process textual data but fail to capture long-term price trends. The proposed hybrid prediction framework involves technical indicators on market data and natural language processing (NLP) sentiment analysis on financial news articles and social media data. FinBERT is employed to predict sentiment on the news articles and social media data. Simultaneously, time series features such as Moving Average, Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), volatility etc. are fetched from historical stocks data and provided to a Long Short-Term Memory (LSTM) model for predicting market trends over shorter timeframes. From comparison results, it is concluded that the hybrid model performs better than the above models with a single feature. The model can serve as an investment decision support system for investors and researchers.*

Keywords: Stock prediction, sentiment analysis, FinBERT, LSTM, technical indicators, hybrid model, financial, NLP

I. INTRODUCTION

In modern economies, stock markets are important for many firms, providing them with access to capital and investment opportunities. Forecasting the future price of a stock can prove difficult, as the price is affected by a number of different factors including company and economic performance, political issues, and investor psychology. Customary forecasting methods analyze historical price movements and technical indicators.

Even if historical data has predictive value, a stock's price can rapidly increase or drop in response to new information, such as a positive or negative announcement or a trending topic on social media causing a sudden demand or sell-off of the stock, so sentiment analysis has become more important in financial prediction systems.

On the other hand, recent advancements in NLP have given computers the ability to understand financial text and deep learning models such as LSTMs are more capable of handling time-series data. The purpose of this project is to combine the classes of signals (sentiment and technical indicators) to better predict short-term stock trends.

II. PROBLEM DEFINITION

The majority of existing models capture either the numeric or the sentiment information but ignore their complementary nature. The technical indicator models perform poorly on shock market events. On the other hand, the sentiment-only models have high variance predictions due to the noisy nature of publicly available online data.

DOI: 10.48175/IJARSCT-33669

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 11, April 2026 Impact Factor: 8.2

Review of prior stock price behavior patterns and reading current market sentiment may improve the accuracy of prediction systems, but a unified system is needed for consistency.

III. RELATED WORK AND EXISTING SYSTEM

Multiple studies aim to identify stock price trends using machine learning and deep learning. Earlier research relied on time series models such as ARIMA. Other models such as Support Vector Machines and Random Forest have been proposed as superior classification approaches.

In recent years, researchers have begun to apply deep learning tools such as RNN or LSTM networks, networks that take into account such sequential dependencies, and have found that Twitter message and financial news sentiment can be helpful for improving forecasting results.

Transformer-based language models such as BERT and FinBERT have further improved sentiment classification in the finance domain. Hybrid systems combining sentiment scores with market indicators generally perform better than single-source models.

IV. PROPOSED SOLUTION

This approach combines sentiment analysis and technical analysis indicators using deep learning techniques and consists of:

1. Data Collection Module

- Historical stock data from Yahoo Finance or Alpha Vantage
- Financial news headlines
- Social media postings about the selected companies

2. Sentiment Analysis Module

The text data that is gathered goes through a FinBERT model to produce the output of a series of probabilities for positive and negative, and neutral sentiments.

3. Technical Indicator Module

Historical market data is used to calculate:

- Simple Moving Average
- Exponential Moving Average (EMA)
- Relative Strength Index (RSI)
- MACD
- Bollinger Bands
- Volatility measures

4. Feature Fusion Module

Sentiment scores are combined with daily stock data into a single vector.

5. Prediction Module

The fused features are provided to an LSTM neural network to predict the direction of the next day stock movement/trend.

6. Visualization Dashboard

It provides price charts, sentiment trends, and model forecasts to users.

V. OBJECTIVES AND SCOPE

A. Objectives

To construct a hybrid stock forecasting model that utilizes both technical and sentiment-based characteristics;

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 11, April 2026 Impact Factor: 8.2

- To use Natural Language Processing (NLP) technique to analyze Twitter/News Social Media data for financial sentiment extraction;
- To calculate technical indicators from historical stock price data;
- To Combine two sets of indicators using LSTM-based neural network;
- To Create a dashboard to provide insight into the forecasts produced by the model and trends within the market;

B. Scope

Future versions of this project could add:

- Real time streaming model of predicting markets
- Reinforcement learning model for fully automated trading.
- Multilingual NLP/Sentiment analysis capabilities
- Explainable AI model for additional transparency.
- Highly Scalable solution by utilizing cloud-based infrastructure.

VI. SYSTEM ARCHITECTURE

We've developed a multi-layer architecture-
The Data Layer includes:
- News and tweets in both structured and unstructured formats
- Historical OHLCV stock data
The Processing Layer contains:
- Text Pre-processing
- Sentiment Classification
- Technical Indicator Computation
The Fusion & Modeling Layer includes:
• Feature Alignment
• Hybrid LSTM Network
The Application Layer includes:
• Dashboard Interface
• Visualization and Reporting

VII. DEVELOPMENT METHODOLOGY

Data Collection
• List of text data from Twitter API and News API
• List of historical stock prices from Yahoo Finance or Alpha Vantage
Data Preprocessing
• Tokenize, clean and filter the text data
• Scale and smooth technical data
Sentiment Analysis
The model used will be FinBERT (pretrained language model fine-tuned for financial text) and will predict three classes: Positive; Negative; Neutral (with their respective probability scores)
Technical Indicator Computation
The indicators used will be:
• Simple Moving Average (SMA)
• Exponential Moving Average (EMA)
• 14-Day Relative Strength Index (RSI)

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 11, April 2026 Impact Factor: 8.2

• Moving Average Convergence Divergence (MACD)
• Bollinger Bands
• Volatility

Model Fusion Layer

"To align sentiment windows with technical data timestamps, we use lag-based time mode synchronization."

LSTM Model Training

"We combine the inputs into one feature vector, and pass the feature vector into an LSTM (Long Short Term Memory Neural Network) layer. The LSTM layer is then followed with multiple dense layers that produce a forecast output."

Evaluation Metrics

"The following models will be tested for accuracy: accuracy, precision, recall, f1 score, root mean squared error (RMSE), and a confusion matrix."

VIII. KEY FUNCTIONAL MODULES

A. The Sentiment Analysis Module extracts the sentiment polarity of the financial text and aggregates all sentiment signals.
B. The Technical Processing Module produces a large number of numerical indicators.
C. The Combined Model Module is where the two sets of input data are merged, and the model makes predictions.
D. The Visualization Module shows price charts, sentiment charts, and predicted future values."

IX. EVALUATION AND RESULTS

Model	Accuracy (trend)	Precision	Recall	F1-score
Technical-only LSTM	0.81	0.80	0.79	0.79
Sentiment-only (FinBERT + Dense)	0.80	0.78	0.77	0.77
Hybrid LSTM (ours)	**0.87**	**0.85**	**0.86**	**0.85**

The hybrid model beat out the other two models, either technical-only or sentiment-only. At times of high media coverage, the proportion of weights for sentiment data increases over the proportion of weights for technical data. This shows that combining NLP with modeling is a good idea. The LSTM type of neuron structure was able to better connect the relationship of the time in the data than either a normal approach or a CNN.

X. CONCLUSION AND FUTURE SCOPE

Conclusion:

In this work, we have introduced a hybrid approach that combines sentiment derived from NLP with technical indicators for forecasting short-term directionality of stocks. Experiments show that the resulting hybrid model outperforms pure sentiment-based and technical only models. The approach is useful as decision support tool and can be deployed in trading dashboards and automated strategies with risk mitigating safeguards.

Future Work:

Reinforcement learning for fully automated strategy optimization. Multi-lingual sentiment for global stocks. Alternative modalities: include Alternative data (Google trends, news entity sentiment, insider trading report etc).Explainability & compliance: Integrate XAI tools and audit logs for regulatory needs. Deployment: microservices with streaming ingestion (Kafka + Spark Structured Streaming) and serving of trained models (TensorFlow Serving / TorchServe).

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.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, April 2026
Impact Factor: 8.2

[2] A. K. Nassirtoussi, S. Aghabozorgi, T. Y. Wah, and D. C. L. Ngo, "Text Mining for Market Prediction: A Systematic Review," *Expert Systems with Applications*, vol. 41, no. 16, pp. 7653–7670, 2014.
[3] X. Zhang, Y. Zhang, and T. Wang, "Hybrid Models for Stock Trend Prediction with Deep Learning and Sentiment Analysis," *IEEE Access*, vol. 6, pp. 21013–21022, 2018.
[4] W. Li, F. Wang, and J. Liu, "Stock Movement Prediction from Tweets and Historical Prices Using BERT and LSTM," in *Proc. IEEE Int. Conf. Big Data*, 2020, pp. 3820–3827.
[5] D. Araci, "FinBERT: A Pretrained Language Model for Financial Sentiment Analysis," *arXiv:1908.10063*, 2019.
[6] Alpha Vantage API Documentation, https://www.alphavantage.co/ (accessed 2025).
[7] Yahoo Finance, https://finance.yahoo.com/ (accessed 2025).
[8] T. Mikolov et al., "Efficient Estimation of Word Representations in Vector Space," *arXiv:1301.3781*, 2013.