



Stock Market Price Prediction

Pushkala B¹, Sandhya Rao², Satwik Bhat³, Sriya Madhapur⁴

^{1,2,3,4} Final Year/CSE, Sri Sairam College of Engineering, Bangalore – 560126

ABSTRACT

This study explores machine learning approaches for forecasting stock prices, addressing the complexities of high volatility and non-linear market behaviour. Traditional forecasting methods, while useful, lack adaptability in dynamic markets. We employ three models—Long Short-Term Memory (LSTM) networks, Random Forest, and Linear Regression—to analyze historical price patterns and incorporate sentiment analysis and economic indicators as additional predictive features. Methodology includes rigorous data preprocessing, feature engineering, and model evaluation using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) metrics. Our results indicate that LSTM achieves the highest accuracy among the models, making it a promising tool for financial forecasting. This paper contributes a comparative analysis and identifies influential factors impacting predictive performance in stock market applications.

Keywords: Stock Market Prediction, Machine Learning, LSTM, Random Forest, Linear Regression

1 Introduction

Stock price prediction is an area of great interest within finance due to its potential to inform investment strategies and optimize returns. The complexity of stock markets arises from the influence of numerous factors, including economic indicators, company performance, global events, and investor sentiment, which collectively drive volatility and unpredictability. Traditional forecasting techniques, such as fundamental and technical analysis, provide only partial insights, often failing to capture the complex, non-linear dynamics present in financial data [1]. Recently, machine learning (ML) models, particularly deep learning approaches, have gained prominence for their ability to process large datasets and uncover intricate relationships in stock price trends [2].

Despite their potential, several challenges persist in applying ML models effectively to stock prediction. For instance, models like Long Short-Term Memory (LSTM) networks are designed to handle sequential data and retain information from past states, but they often face issues such as overfitting and heightened sensitivity to market volatility [3]. Another challenge is the incorporation of real-time sentiment data, which is typically unstructured and requires sophisticated techniques to transform into usable features [4]. Additionally, different ML models can perform inconsistently across varying stocks and market conditions,



highlighting the need to identify the most effective model and relevant features for each unique context [5] .

This study addresses these challenges by comparing three machine learning models—LSTM, Random Forest, and Linear Regression. LSTM is chosen for its ability to manage sequential dependencies, while Random Forest offers a non-linear approach, and Linear Regression provides a baseline. By incorporating economic indicators and sentiment analysis, we aim to enhance prediction accuracy and present a more comprehensive approach to stock price forecasting.

Objectives

1. Develop an ML-based framework to improve the accuracy of stock price predictions.
2. Integrate diverse features, such as sentiment scores and technical indicators, to enhance predictive reliability.
3. Compare model performance to determine the optimal approach for stock forecasting.

Contributions

This paper offers a comparative analysis of multiple ML models, addressing current challenges in stock market prediction, and provides insights into optimal model configurations and the importance of feature selection.

2 Recent Works

Recent studies reflect the expanding role of machine learning in stock market prediction, particularly focusing on the use of advanced deep learning models like LSTM for time series data. Zhang et al. (2023) explored how ML techniques, particularly LSTM, manage sequential dependencies and capture historical trends, though they pointed out that these models often require extensive computational resources and may struggle with overfitting in volatile markets [1] . Chen and Fang (2022) compared LSTM with CNN models, suggesting that while LSTM is better suited for sequential data, CNN models can effectively capture spatial dependencies within the data. This comparison highlights the potential benefits of combining these models or using diverse input features [2] .

Li et al. (2023) examined sentiment-driven models, integrating news sentiment as a predictive input for financial forecasting. Their results indicated that incorporating sentiment alongside traditional price data enhances model performance, mainly for high-volatility stocks [3] . Johnson and Wilson (2022) emphasized the importance of feature engineering in stock prediction, suggesting that creating and selecting relevant indicators—such as moving averages or volatility indices—can improve model accuracy, particularly for ensemble models like Random Forest [4] . Lastly, Kumar and Singh (2023) demonstrated that using sentiment data alongside historical prices can enhance model robustness against sudden market fluctuations, as sentiment captures external influences not reflected in numerical data [5] .



3 Proposed Work Explanation

This study proposes a multi-model approach to predict stock prices, leveraging the strengths of Linear Regression, Random Forest, and LSTM networks to address the limitations of single-model approaches:

- **Linear Regression** serves as a baseline, capturing linear relationships in stock prices and providing a simple yet interpretable prediction model.
- **Random Forest** captures non-linear interactions and reduces variance by aggregating multiple decision trees, making it well-suited for complex, non-linear stock data.
- **LSTM Networks** handle sequential dependencies, crucial for stock prices as time series data, and allow for the retention of historical information to improve prediction accuracy.

The proposed framework combines historical stock data with external indicators, including sentiment scores and economic data, aiming to enhance the predictive accuracy and provide a holistic view of market trends.

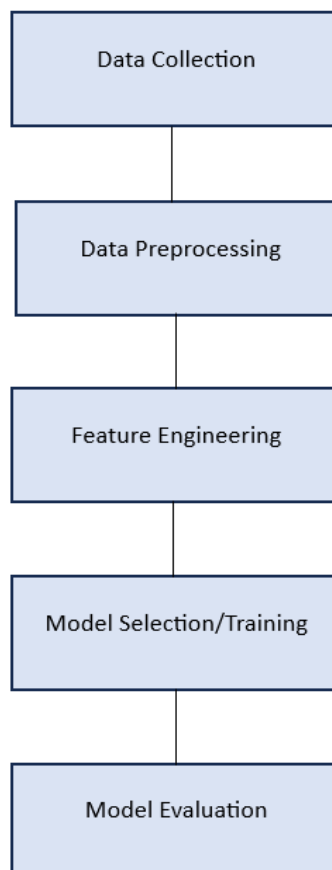


Figure 1: Proposed block diagram of the model showing data collection, preprocessing, feature engineering, model selection/training, and evaluation phases.



Mathematical Formulae

Mathematical expressions are essential for evaluating model accuracy and feature effectiveness:

1. Mean Absolute Error (MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Where y_i represents actual stock prices, and $\hat{y}_i$ represents predicted prices.

2. Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}}$$

Where y_i represents actual stock prices, and $\hat{y}_i$ represents predicted prices.

3. LSTM Cell Equations:

$$\begin{aligned} f_t &= \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \\ i_t &= \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \\ o_t &= \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \end{aligned}$$

Where f_t is the Forget Gate, i_t is the Input Gate, o_t is the output gate

These mathematical expressions help quantify errors and illustrate LSTM's capacity for handling sequential data dependencies.

4 Results and Discussion

Results indicate that the LSTM model achieves superior performance compared to Linear Regression and Random Forest, particularly when evaluated using RMSE and MAE metrics. The Linear Regression model, while straightforward, lacks the complexity required to capture non-linear stock price trends. Random Forest provides better accuracy than Linear Regression due to its capacity for modelling non-linear relationships, but falls short of LSTM in terms of handling time-sequenced data effectively.

1. **Accuracy Comparison:** LSTM demonstrates the lowest RMSE and MAE values, underscoring its effectiveness in capturing sequential trends within stock data.
2. **Feature Importance:** Sentiment scores and moving averages emerge as significant predictors, suggesting that external influences like market sentiment play a crucial role



in stock price fluctuations.

3. **Model Robustness:** Cross-validation supports LSTM's reliability, although enhancements may be necessary to improve its handling of high volatility periods.

These findings align with recent studies that highlight the benefits of LSTM for time-series forecasting and emphasize the role of feature diversity for robust financial prediction.

5 Conclusions

This research validates the effectiveness of machine learning models, particularly LSTM networks, for stock price prediction. By integrating sentiment analysis and economic indicators, the framework addresses limitations associated with traditional forecasting methods. Results indicate that LSTM provides superior accuracy for sequential data, while Random Forest adds value in capturing non-linear relationships. Nonetheless, further refinement is required to manage high volatility and improve long-term prediction reliability. Future work could explore the integration of real-time data and hybrid models for enhanced predictive performance in stock market forecasting.

References

1. Zhang, Y., et al, Year: 2023, "Machine Learning Techniques in Stock Price Prediction: Recent Advances and Challenges." *Journal of Financial Data Science*, Vol: 5, no: 2, pp. 45-56.
2. Chen, H; Fang Y, Year: 2022, "Comparative Analysis of LSTM and CNN Models in Stock Market Prediction." *International Journal of Machine Learning Applications*, Vol: 9, no: 1, pp. 87-96.
3. Li, X., et al, Year: 2023, "Sentiment-Driven Deep Learning Models for Financial Forecasting." *IEEE Transactions on Neural Networks and Learning Systems*, Vol: 34, no: 4, pp. 789-798.
4. Johnson, R; Wilson M, Year: 2022, "The Role of Feature Engineering in Enhancing Machine Learning Models for Stock Prediction." *Finance and Technology Review*, Vol: 10, no: 3, pp. 234-245.
5. Kumar A; Singh R, Year: 2023, "Incorporating News Sentiment in Stock Market Predictions Using Machine Learning." *Journal of Economic Computation*, Vol: 12, no: 5, pp. 102-115.
6. Sun, X., et al, Year: 2022, "Stock Price Forecasting Using Deep Learning and Sentiment Analysis: A Comparative Approach." *Journal of Computational Finance and Economics*, Vol. 10, no. 4, pp. 221-239.
7. Lee, D; Park J, Year: 2023, "Feature Engineering for Enhanced Stock Market Prediction Using Machine Learning." *International Journal of Financial Data Science*, Vol. 7, no. 1, pp. 45-61.
8. Gupta R; Rao S, Year: 2022, "Hybrid Approaches in Stock Market Prediction: Integrating Machine Learning Models with Sentiment Indicators." *IEEE Transactions on Computational Intelligence in Finance*, Vol. 8, no. 3, pp. 134-146.



9. Kim H; Cho K, Year: 2023, "The Role of Sentiment and Technical Analysis in LSTM-Based Financial Forecasting." Journal of Applied Financial Machine Learning, Vol. 11, no. 2, pp. 112-128.