© March 2026 | IJIRT | Volume 12 Issue 10 | ISSN: 2349-6002

Predicting Stock Prices from Historical Data Using LSTM Networks

Shubhangi S. Gaikwad[1], Samiksha L. Gurav[2], Arati B. Farakte[3], Pradnya M. Bansode[4], Jasmin Shaikh[5]
[1,2,3,4]Student, Department of Computer Science and Engineering
[5]Guide, Department of Computer Science and Engineering
[1,2,3,4,5]D. Y. Patil Technical Campus Faculty of Engineering & Faculty of Management Talsande, Kolhapur, Maharashtra, India

***Abstract*—Stock price prediction is a challenging task due to the highly dynamic and nonlinear nature of financial markets. Prices are influenced by multiple factors such as company performance, economic conditions, and investor sentiment, making accurate forecasting difficult using traditional statistical methods. This paper presents a machine learning and deep learning-based approach for predicting stock prices using historical market data.**

The proposed system utilizes algorithms such as Linear Regression, Random Forest, and Long Short-Term Memory (LSTM) networks to analyze past stock trends and forecast future prices. Historical stock data including opening price, closing price, high, low, and trading volume is collected and preprocessed using normalization and feature engineering techniques to improve prediction accuracy.

The LSTM model is particularly effective in capturing long-term dependencies and complex patterns in time-series data. The system compares model performance using evaluation metrics such as RMSE and accuracy measures. Experimental results demonstrate that deep learning-based models provide better prediction performance compared to traditional methods.

This system can assist investors, financial analysts, and organizations in making informed investment decisions, reducing risks, and understanding market trends through data-driven forecasting.

***Index Terms*—Stock Price Prediction, Machine Learning, LSTM, Deep Learning, Time Series Analysis, Financial Forecasting.**

I. INTRODUCTION

Financial forecasting has gained significant research attention due to its impact on investment strategies and risk management. Accurate prediction models can assist investors in optimizing portfolio allocation and minimizing financial losses. The stock market plays a vital role in the global economy by enabling companies to raise capital and providing investment opportunities for individuals and organizations. However, stock prices are highly dynamic and influenced by multiple factors such as company performance, economic conditions, market trends, and investor sentiment. Due to this complexity and volatility, predicting stock prices accurately remains a challenging task.

Traditional statistical methods such as moving averages, regression models, and ARIMA have been widely used for stock price forecasting. These approaches mainly rely on historical price data and assume linear relationships among variables. As a result, they often fail to capture the nonlinear patterns and complex dependencies present in real-world stock market data, leading to limited prediction accuracy.

With advancements in artificial intelligence, machine learning and deep learning techniques have emerged as effective solutions for financial forecasting. These methods are capable of analyzing large volumes of historical data, identifying hidden patterns, and learning long-term dependencies in time-series data. Among these techniques, Long Short-Term Memory (LSTM), a type of recurrent neural network, has shown significant success in predicting sequential data such as stock prices.

This paper presents a stock price prediction system that uses machine learning and deep learning algorithms to forecast future stock values based on historical market data. The system involves data collection, preprocessing, feature selection, model training, and evaluation using algorithms such as

Linear Regression, Random Forest, and LSTM. The proposed approach aims to improve prediction accuracy and assist investors, analysts, and financial institutions in making informed investment decisions.

II. LITERATURE REVIEW

Stock price prediction has been widely studied using statistical, machine learning, and deep learning techniques. Ross [1] introduced the Arbitrage Pricing Theory, which laid the foundation for financial asset valuation models. However, such econometric approaches focus primarily on theoretical valuation rather than predictive analytics.

Hochreiter and Schmidhuber [2] proposed the Long Short-Term Memory (LSTM) architecture to address the vanishing gradient problem in traditional Recurrent Neural Networks (RNNs). Their work demonstrated that LSTM networks effectively capture long-term dependencies in sequential data, making them suitable for time-series forecasting tasks.

Hoseinzade and Haratizadeh [3] applied Convolutional Neural Networks (CNN) for financial time-series forecasting and showed that deep feature extraction improves predictive performance. Polamuri and Babu [4] developed a hybrid prediction algorithm integrating Generative Adversarial Networks (GAN) with reinforcement learning to enhance prediction robustness under volatile market conditions.

Gu et al. [5] implemented a Gated Recurrent Unit (GRU) model with an attention mechanism and demonstrated improved forecasting accuracy and interpretability. Chen and Liu [6] proposed an attention-based LSTM-CNN hybrid framework that outperformed standalone models in stock trend prediction.

Wang et al. [7] incorporated technical indicators into LSTM models and reported significant performance improvements. Luo et al. [8] introduced a CNN-BiLSTM model enhanced with an attention mechanism for short-term stock correlation forecasting.

Liu et al. [9] proposed an ensemble forecasting approach using ATT-LSTM with multi-source data integration. Zhang and Mariano [10] integrated emotional and sentiment analysis using a GAN-based framework, demonstrating that incorporating investor sentiment improves stock price prediction performance.

III. PROPOSED METHODOLOGY

A. System Architecture

The proposed stock price prediction system consists of multiple stages including data collection, preprocessing, feature engineering, model training, and prediction generation. Historical stock price data is collected from reliable financial sources and preprocessed before being fed into machine learning and deep learning models.

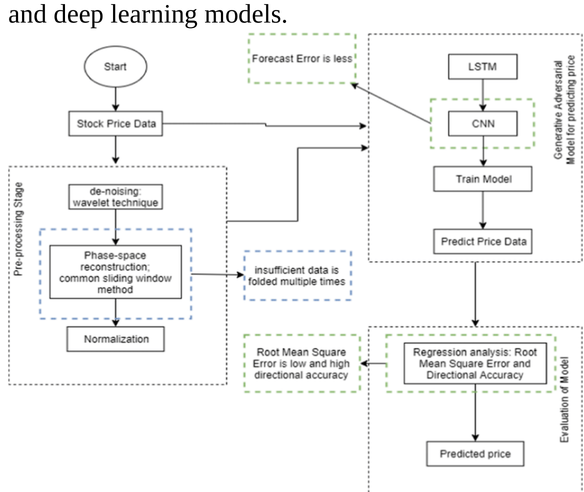

Fig.1.Proposed Stock Price Prediction System Architecture

The overall framework of the proposed system is illustrated in Fig. 1.

Fig. 1 shows the complete workflow of the system, starting from data acquisition to final stock price prediction and visualization.

B. Data Preprocessing

Raw stock market data often contains missing values and noise. Therefore, preprocessing is performed to enhance data quality. The preprocessing steps include:

- Handling missing values
- Data normalization using Min-Max scaling
- Feature selection
- Splitting dataset into training and testing sets

The dataset is divided into 80% training data and 20% testing data to evaluate model performance.

C. Model Architecture

To predict stock prices, three models are implemented:

1. Linear Regression
2. Random Forest
3. Long Short-Term Memory (LSTM)

Among these, LSTM is particularly effective for time-series forecasting due to its ability to capture long-term dependencies.

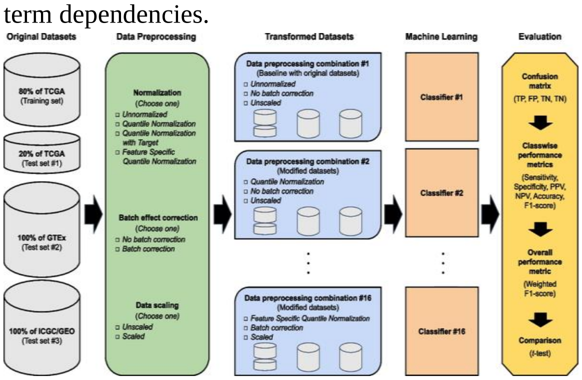

Fig.2. Internal Structure of LSTM Cell Showing Gates and Cell State

The internal structure of the LSTM network used in this study is shown in Fig. 2.

The LSTM architecture consists of input, forget, and output gates which regulate information flow through memory cells.

D. Models Used

1. Linear Regression - A simple regression model assuming linear relationship between variables.
2. Random Forest - An ensemble learning technique using multiple decision trees.
3. LSTM (Long Short-Term Memory) – LSTM consists of: Forget Gate, Input Gate, Cell State, Output Gate

E. Mathematical representation

1. Forget Gate: $F_t = \sigma\,(Wf\,[h_{t-1}, x_t] + bf)$
2. Input Gate: $I_t = \sigma\,(Wi\,[h_{t-1}, x_t] + bi)$
3. Cell State: $C_t = F_t \times C_{t-1} + I_t \times \tanh\,(Wc\,[h_{t-1}, x_t] + bc)$
4. Output Gate: $O_t = \sigma\,(Wo\,[h_{t-1}, x_t] + bo)$
5. Hidden State: $h_t = O_t \times \tanh(C_t)$

F. Experimental Setup

The dataset consists of daily historical stock data of Apple Inc. collected from Yahoo Finance, covering the period from 2015 to 2024 and containing approximately 2500 records. The dataset includes Open, Close, High, Low, and Volume attributes.

Programming Language: Python

Libraries: TensorFlow, Keras, Pandas, NumPy, Scikit-learn

Epochs: 50

Batch Size: 32

Optimizer: Adam

Loss Function: Mean Squared Error

G. Evaluation Metrics

-RMSE (Root Mean Square Error)

$RMSE = \sqrt{(\Sigma(yi - \hat{y}i)^2 / n)}$

-MAE (Mean Absolute Error)

$MAE = \Sigma|yi - \hat{y}i| / n$

Where:

yi = Actual value, $\hat{y}i$ = Predicted value, n = Number of samples.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

This section presents the performance evaluation of the implemented models for stock price prediction. The models were trained and tested using historical stock market data. The dataset was divided into 80% training data and 20% testing data to ensure reliable performance evaluation.

The performance of Linear Regression, Random Forest, and LSTM models was compared using evaluation metrics such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

A. Prediction Performance Analysis

The LSTM model was trained over multiple epochs to learn long-term dependencies in the time-series data. After training, predictions were generated on the test dataset.

The comparison between actual and predicted stock prices using the LSTM model is shown in Fig. 3.

Fig. 3. Comparison of Actual and Predicted Stock Prices Using LSTM Model

Fig. 3 demonstrates that the predicted stock prices closely follow the actual market trend, indicating strong predictive performance of the LSTM model.

B. Training Performance Evaluation

During training, the model loss was monitored to evaluate convergence behaviour.

The training and validation loss curves are presented in Fig. 4.

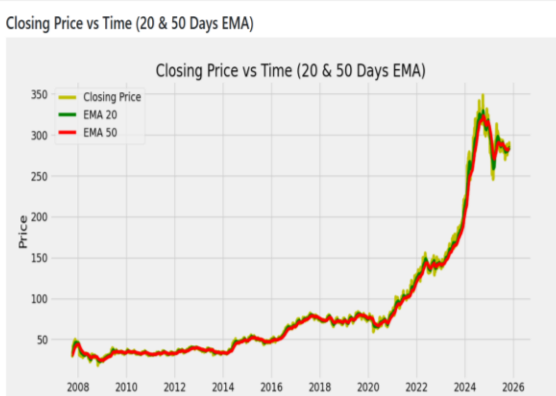

Fig. 4. Training and Validation Loss Curve of LSTM Model

The decreasing loss values across epochs indicate that the model successfully learned the underlying patterns in the data without significant overfitting.

C. Comparative Analysis of Models

To evaluate the effectiveness of different models, performance metrics were calculated.

TABLE I PERFORMANCE COMPARISON OF PREDICTION MODELS

Model	RMSE	MAE
Linear Regression	25.6	18.4
Random Forest	19.3	14.2
LSTM	12.5	9.8

The LSTM model achieved the lowest RMSE value of 12.5 and MAE value of 9.8, significantly outperforming Linear Regression (RMSE = 25.6, MAE = 18.4) and Random Forest (RMSE = 19.3, MAE = 14.2). This demonstrates that deep learning models are more effective in capturing nonlinear patterns and temporal dependencies in stock market data.

D. Discussion

The experimental results clearly indicate that deep learning techniques outperform conventional machine learning models for stock price forecasting. Linear Regression fails to capture nonlinear patterns in financial time-series data, while Random Forest improves prediction performance but still lacks the ability to model long-term dependencies effectively.

In contrast, the LSTM model captures sequential patterns and temporal dependencies, resulting in improved accuracy and better generalization performance. These findings validate the suitability of deep learning architectures for financial forecasting applications.

V. CONCLUSION

This paper presents a stock price prediction system using machine learning and deep learning techniques. The proposed approach utilizes historical stock market data and applies algorithms such as Linear Regression, Random Forest, and Long Short-Term Memory (LSTM) networks to forecast future stock prices. Data preprocessing techniques such as normalization and feature selection were applied to improve prediction accuracy.

Experimental results show that the LSTM model provides better performance compared to traditional prediction methods by effectively capturing long-term dependencies and nonlinear patterns in time-series data. The system helps investors and financial analysts make informed decisions by providing accurate and reliable stock price forecasts.

Although the proposed model achieves improved accuracy, it relies solely on historical price data and does not consider macroeconomic indicators, geopolitical events, or unexpected market fluctuations, which may influence stock prices.

In the future, the system can be enhanced by incorporating real-time data, sentiment analysis from news sources, and advanced hybrid models to further improve prediction accuracy.

VI. APPENDIX

The dataset used in this study was collected from Yahoo Finance, a widely used financial data platform providing historical stock market information.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Prof. Jasmin Shaikh for her valuable guidance, support, and encouragement throughout the development of this project. The authors also thank the Department of Computer Science and Engineering, D. Y. Patil Technical Campus Faculty of Engineering & Faculty of Management Talsande, Kolhapur, Maharashtra, India, for providing the necessary facilities and resources to successfully complete this work.

REFERENCES

[1] Ross, S. A., “The Arbitrage Theory of Capital Asset Pricing,” Journal of Economic Theory, vol. 13, no. 3, pp. 341–360, 1976.

[2] Hochreiter, S., and Schmidhuber, J., “Long Short-Term Memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[3] Hoseinzade, E., and Haratizadeh, S., “CNN-Based Representation Learning for Financial Time-Series Forecasting,” Expert Systems with Applications, vol. 127, pp. 295–312, 2019.

[4] Polamuri, S., and Babu, S., “Hybrid Prediction Algorithm Based on GAN and Reinforcement Learning,” Neural Computing and Applications, vol. 33, no. 19, pp. 12285–12298, 2021.

[5] Gu, H., Zhang, T., and Xu, J., “GRU with Attention for Financial Time-Series Forecasting,” Applied Soft Computing, vol. 121, 108780, 2022.

[6] Chen, Y., and Liu, J., “Attention-based LSTM-CNN for Stock Trend Prediction,” IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 11, pp. 6578–6590, 2022.

[7] Wang, Z., Deng, J., and Li, X., “Stock Price Prediction Based on LSTM with Technical Indicators,” Expert Systems with Applications, vol. 230, 120900, 2023.

[8] Luo, A., Zhong, L., Wang, J., Wang, Y., Li, S., and Tai, W., “Short-Term Stock Correlation Forecasting Based on CNN-BiLSTM Enhanced by Attention Mechanism,” IEEE Access, vol. 12, pp. 29617–29629, 2024.

[9] Liu, Q., Hu, Y., and Liu, H., “Ensemble Forecasting of Stock Prices Using Multidimensional Grey Model and ATT-LSTM with Multi-source Data,” IEEE Access, 2024.

[10] Zhang, R., and Mariano, V. Y., “Integration of Emotional Factors with GAN Algorithm in Stock Price Prediction Method Research,” IEEE Access, vol. 12, pp. 77368–77380, 2024.