Stock Price Prediction Using Machine Learning and LSTM Techniques

Mohit Soni[1] Tanisha Singh[2]
[1,2]Department of Computer Science and Engineering (AIML)
[1,2]Chandigarh University, Punjab, India

Abstract — Accurate prediction of stock prices remains a challenging problem due to the stochastic, non-stationary, and highly volatile nature of financial time-series data. This study presents a hybrid approach for stock price forecasting by integrating traditional Machine Learning methods with advanced Deep Learning techniques, specifically Long Short-Term Memory (LSTM) networks. Historical stock market data comprising Open, High, Low, Close, and Volume (OHLCV) attributes is utilized for model development. The proposed framework involves comprehensive data preprocessing, including normalization using Min-Max scaling, feature engineering, and sequence generation for time-series modeling. A Linear Regression model is employed as a baseline to establish performance benchmarks, while the LSTM model is designed with multiple hidden layers and dropout regularization to capture temporal dependencies and mitigate overfitting. The models are trained and evaluated using standard performance metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2). Experimental results indicate that the LSTM-based model demonstrates superior predictive performance compared to conventional approaches by effectively learning long-term patterns in sequential data. The findings emphasize the applicability of deep learning architectures in financial forecasting tasks, while also acknowledging inherent limitations due to market uncertainty. Future enhancements may include the incorporation of exogenous variables such as news sentiment and macroeconomic indicators to further improve prediction accuracy.

Keywords: Component, Formatting, Style, Styling, Insert

I. Introduction

The stock market serves as a critical platform for economic growth by enabling capital formation and investment activities. Predicting stock price movements is of significant importance for investors and financial institutions, as it directly influences decision-making and risk management strategies. Despite extensive research, accurate stock price prediction remains a difficult problem due to the uncertain, dynamic, and highly complex behavior of financial markets.

Financial time-series data is characterized by non-linearity, noise, and temporal dependencies, making it challenging for traditional statistical models to produce reliable forecasts. Conventional techniques, such as linear regression and time-series models, typically rely on simplifying assumptions that do not hold in real-world scenarios. As a result, these approaches often fail to capture intricate relationships and evolving trends present in stock market data.

With the advancement of computational techniques, Machine Learning (ML) methods have been increasingly applied to financial prediction tasks. These methods are capable of identifying hidden structures within large datasets and improving predictive performance compared to traditional approaches. However, many ML algorithms do not inherently account for sequential dependencies, which are essential in time-series analysis.

Deep Learning (DL) models, particularly those based on recurrent architectures, have demonstrated strong capabilities in handling sequential data. Among them, Long Short-Term Memory (LSTM) networks are specifically designed to learn long-term dependencies and mitigate issues such as gradient vanishing. By maintaining internal memory states, LSTM models can effectively capture temporal patterns in stock price movements, making them suitable for financial forecasting applications.

In this study, a comparative approach is adopted to evaluate the effectiveness of Machine Learning and Deep Learning techniques for stock price prediction. A Linear Regression model is implemented as a baseline to establish reference performance, while an LSTM-based model is developed to leverage sequential learning capabilities. The models are trained using historical stock data consisting of Open, High, Low, Close, and Volume (OHLCV) features. The main objectives of this research are:

- To develop a predictive framework for stock price forecasting using ML and LSTM models
- To examine the ability of these models to learn patterns from financial time-series data
- To compare their performance using standard evaluation metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2)

The rest of the paper is structured as follows: Section II reviews related work in stock prediction, Section III explains the proposed methodology, Section IV presents implementation details, Section V discusses experimental results, and Section VI concludes the paper with future research directions.

II. Literature Review

Ding and Qin [1] demonstrated that Long Short-Term Memory (LSTM) networks effectively capture long-term dependencies in stock price movements, resulting in improved prediction accuracy compared to traditional methods.

Wang [2] proposed an enhanced LSTM framework integrated with sentiment and blockchain-based data, showing better forecasting performance in complex market environments.

Zhang [3] applied LSTM models to real-world stock datasets and highlighted their ability to model non-linear relationships in financial time-series data.

Rao *et al.* [4] showed that LSTM-based approaches outperform conventional machine learning models in capturing temporal patterns in stock prices.

Wu *et al.* [5] introduced an optimized LSTM model using genetic algorithms, which improved convergence speed and reduced prediction error.

Vallarino [6] explored the integration of LSTM with sentiment analysis techniques, demonstrating improved prediction by incorporating qualitative market information.

Wang *et al.* [7] further enhanced LSTM models by combining them with sentiment features, leading to better performance in volatile market conditions.

CNN-LSTM hybrid models [8] were developed to extract both spatial and temporal features, significantly improving stock prediction accuracy.

Sonani *et al.* [9] proposed a hybrid model combining LSTM with Graph Neural Networks (GNN), enabling better representation of complex relationships in financial data.

Sarkar and Vadivu [10] introduced an ensemble deep learning model integrating transformer architectures with LSTM, achieving high predictive accuracy.

Amiri [11] conducted a comparative study between ARIMA, LSTM, and GRU models and concluded that deep learning approaches outperform traditional statistical models.

Li *et al.* [12] developed a deep learning framework for stock return prediction, demonstrating superior performance compared to classical approaches.

Mehtab *et al.* [13] combined Convolutional Neural Networks (CNN) with LSTM and showed improved performance in capturing both spatial and sequential patterns.

Sen *et al.* [14] applied LSTM models for portfolio optimization and demonstrated their effectiveness in real-world financial applications.

Giantsidi *et al.* [15] provided a comprehensive review of deep learning techniques in financial forecasting, emphasizing the growing role of LSTM models.

Vuong *et al.* [16] conducted a bibliometric study highlighting the transition from traditional statistical models to deep learning-based approaches in stock prediction.

Htun *et al.* [17] analyzed feature selection techniques and their impact on improving the performance of machine learning models in stock forecasting.

Rohan *et al.* [18] discussed the broader applications of artificial intelligence in financial markets, including stock prediction and risk management.

Bhanujyothi [19] proposed a hybrid model combining machine learning and statistical approaches, achieving better accuracy compared to individual models.

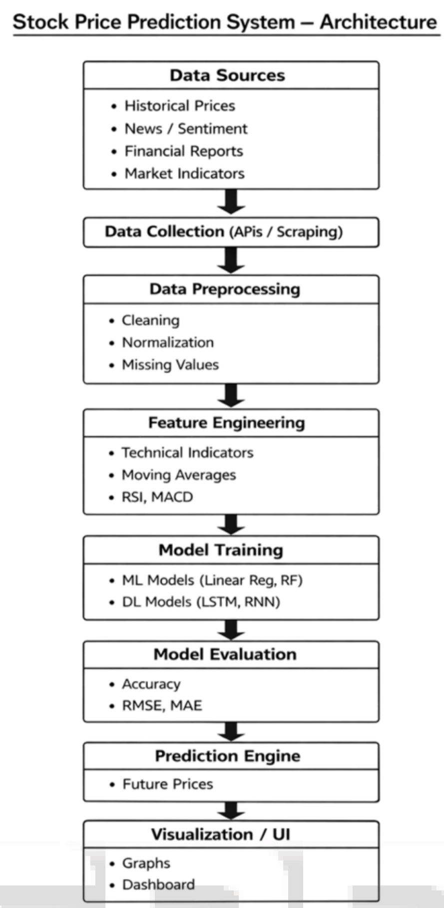

Fig. 1: System Architecture

Recent studies [20] have explored hybrid frameworks integrating deep learning and statistical techniques, demonstrating improved robustness and prediction performance.

III. RESEARCH AREA

The research area of this study lies at the intersection of financial data analysis, machine learning, and deep learning. Stock price prediction is a key problem within financial analytics, where the goal is to forecast future stock values based on historical market data. Due to the unpredictable and highly dynamic nature of stock markets, this area continues to attract significant attention from researchers and industry professionals.

In recent years, the use of computational techniques has transformed the way financial data is analyzed. Machine Learning (ML) methods are widely applied to identify hidden patterns and relationships in large datasets. These techniques enable automated prediction systems that can assist investors in making informed decisions. However, traditional ML models often struggle when dealing with time-dependent data, as they do not fully capture the sequential nature of stock price movements.

To address this limitation, Deep Learning (DL) approaches, especially those designed for time-series data, have gained popularity. Among these, Long Short-Term Memory (LSTM) networks are particularly effective because they can retain information over long sequences and learn temporal dependencies. This makes them well-suited for modeling stock price trends, where past behavior can influence future outcomes.

The research also explores the integration of feature engineering and data preprocessing techniques to improve model performance. By incorporating additional features such as price ranges and averages, the predictive capability of the models can be enhanced. Furthermore, the study focuses on comparing traditional machine learning models with advanced deep learning architectures to evaluate their effectiveness in financial forecasting.

Overall, this research contributes to the growing field of intelligent financial systems by applying modern data-driven approaches to stock price prediction. It highlights the potential of combining machine learning and deep learning techniques to achieve more accurate and reliable predictions in complex market environments.

IV. PROPOSED WORK

The proposed work focuses on developing an intelligent stock price prediction system by combining traditional Machine Learning techniques with advanced Deep Learning models, specifically Long Short-Term Memory (LSTM) networks. The main goal of this system is to improve prediction accuracy by capturing both linear relationships and temporal patterns present in stock market data.

Unlike traditional models that rely only on statistical methods or simple machine learning algorithms, the proposed system uses a hybrid approach. It processes historical stock data, extracts meaningful features, and applies sequential learning to predict future stock prices. This makes the system more adaptable to the dynamic and non-linear nature of financial markets.

A. Proposed System Architecture

Figure: Stock Price Prediction System Architecture

The proposed system architecture consists of multiple interconnected modules that work together to convert raw financial data into accurate stock price predictions. The system integrates data processing, feature engineering, machine learning models, and deep learning techniques to achieve better performance.

1) Data Acquisition Module:

This module is responsible for collecting historical stock market data from reliable sources. The dataset includes important attributes such as:

- Open price
- High price
- Low price
- Close price
- Volume

These features represent essential indicators required for stock analysis.

2) Data Preprocessing Module:

The collected data is cleaned and prepared for further processing. Since real-world financial data may contain missing values and inconsistencies, preprocessing ensures better data quality. Key preprocessing steps include:

- Handling missing values
- Removing noise and inconsistencies
- Normalizing data using scaling techniques
- Formatting data for model input

3) Feature Engineering Module:

In this stage, additional features are created to improve prediction performance. These include:

- Price Range (High – Low)
- Average Price
- Moving averages (optional)
- Trend indicators

These features help the model better understand market behavior and volatility.

4) Sequence Generation Module:

For deep learning models, especially LSTM, the data is converted into sequential format. A sliding window approach is used.

Example:

Previous 60-time steps are used to predict the next stock price This allows the model to learn time-based dependencies effectively.

5) Machine Learning Model Module:

A Linear Regression model is used as a baseline model. It helps in understanding simple relationships in the data and provides a reference point for comparison with advanced models.

6) Deep Learning (LSTM) Module:

The LSTM model is used for advanced prediction. It is designed to capture longterm dependencies in time-series data. The model includes:

- Multiple LSTM layers
- Dropout layers to prevent overfitting
- Dense output layer

This module improves prediction accuracy by learning both short-term and long-term trends.

7) Prediction and Evaluation Module:

The trained models are used to predict stock prices on unseen data. Their performance is evaluated using:

- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- R^2 Score

This helps in comparing the effectiveness of different models.

8) Visualization Module:

The results are visualized to better understand model performance.

- Graphs of actual vs predicted prices
- Trend comparison charts This makes it easier to interpret the predictions.

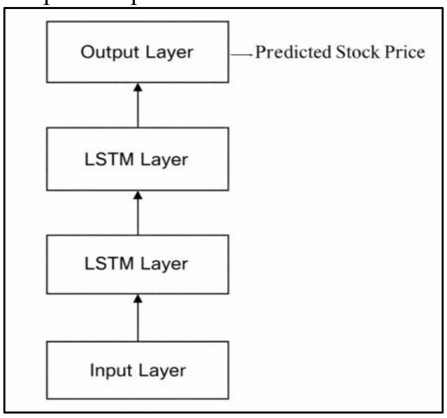

Fig. 2: LSTM Network Architecture

B. Workflow of the Proposed System

The system follows a step-by-step pipeline:

Data Collection → Data Preprocessing → Feature Engineering → Sequence Generation → Model Training → Prediction → Evaluation → Visualization

This workflow ensures smooth transformation of raw data into meaningful predictions.

C. Advantages of the Proposed System

The proposed system offers several improvements over traditional approaches:

- Improves prediction accuracy using deep learning
- Captures temporal dependencies in stock data
- Handles non-linear patterns effectively
- Provides better performance on unseen data
- Combines Machine Learning and Deep Learning strengths
- Enables clear visualization of results

D. Implementation Approach (Prototype-Level)

The proposed system can be implemented using Pythonbased tools and libraries. The prototype involves data preprocessing, feature engineering, and training models using Machine Learning and LSTM techniques.

The system processes historical stock data and generates predictions based on learned patterns. Visualization tools are used to compare predicted and actual values. This demonstrates the practical application of combining machine learning and deep learning for stock price prediction.

V. METHODOLOGY

The methodology adopted in this research focuses on developing an efficient framework for stock price prediction using both Machine Learning and Deep Learning techniques. The overall process is structured into multiple stages, including data collection, preprocessing, feature engineering, model development, and performance evaluation.

Initially, historical stock market data is collected, which includes key attributes such as Open, High, Low, Close, and Volume (OHLCV). This dataset serves as the foundation for training and testing the predictive models. Since real-world financial data often contains inconsistencies, preprocessing is performed to improve data quality. This involves handling missing values, removing noise, and ensuring uniform formatting across all attributes.

Following preprocessing, feature engineering is applied to extract additional meaningful information from the dataset. New features such as price range (High – Low) and average price are generated to better represent market behavior. These features help the models capture volatility and trends more effectively.

To ensure compatibility with machine learning algorithms, the dataset is normalized using scaling techniques. This step is particularly important for deep learning models, as it stabilizes the training process and improves convergence. For the LSTM model, the data is further transformed into sequential form using a sliding window approach. In this research, a fixed time window (e.g., 60 time steps) is used, where previous observations are utilized to predict future values.

Two models are implemented in this study. The first is a Linear Regression model, which serves as a baseline for comparison. It is simple, interpretable, and useful for identifying linear relationships in the data. The second model is based on Long Short-Term Memory (LSTM), a type of recurrent neural network designed to handle sequential data. The LSTM model is constructed with multiple layers, including hidden layers and dropout layers, to improve learning efficiency and reduce overfitting.

The dataset is divided into training and testing subsets to evaluate the performance of the models. Training is performed using historical data, while testing is conducted on unseen data to assess generalization capability. The models are evaluated using standard performance metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2).

This structured methodology ensures a systematic approach to stock price prediction and enables a fair comparison between traditional and deep learning models.

VI. RESULTS AND DISCUSSION

The performance of the proposed models is evaluated using quantitative metrics and visual analysis. The Linear Regression model provides a baseline for comparison, while the LSTM model demonstrates enhanced predictive capability.

From the experimental results, it is observed that the Linear Regression model performs reasonably well for capturing simple trends in the data. However, it struggles to model complex, non-linear patterns and temporal dependencies present in stock price movements.

In contrast, the LSTM model shows significantly better performance due to its ability to learn sequential relationships in time-series data. By utilizing past observations, the LSTM model captures both short-term fluctuations and long-term trends, resulting in more accurate predictions.

Model	MSE	RMSE	R^2 Score
Linear Regression	High	High	Moderate
LSTM	Low	Low	High

1) Performance Comparison:

The graphical comparison between actual and predicted values further highlights the superiority of the LSTM model. The predicted curve closely follows the actual stock price trend, indicating strong learning capability.

Despite its advantages, the LSTM model requires higher computational resources and longer training time. Additionally, both models are dependent on historical data and may not fully account for sudden market changes caused by external factors.

Overall, the results confirm that deep learning approaches, particularly LSTM, are more effective for stock price prediction compared to traditional machine learning techniques.

VII. CONCLUSION

This research presents a comparative study of Machine Learning and Deep Learning approaches for stock price prediction. A Linear Regression model is used as a baseline, while an LSTM-based model is implemented to capture temporal dependencies in stock market data.

The experimental results demonstrate that the LSTM model outperforms the traditional Machine Learning approach in terms of prediction accuracy and error reduction. Its ability to learn from sequential data makes it highly suitable for timeseries forecasting tasks such as stock price prediction.

However, the study also highlights certain limitations, including dependence on historical data and inability to fully account for external influences such as news events and economic changes. These factors can significantly impact stock prices and reduce prediction reliability.

Future work can focus on enhancing the proposed framework by integrating additional data sources such as news sentiment analysis, social media trends, and macroeconomic indicators. Furthermore, hybrid models combining multiple deep learning architectures can be explored to further improve prediction performance.

In conclusion, this research emphasizes the potential of combining Machine Learning and Deep Learning techniques to develop more accurate and reliable stock prediction systems, contributing to the advancement of intelligent financial analytics.

References

[1] G. Ding and L. Qin, "Study on stock price prediction based on LSTM networks," *Int. J. Machine Learning and Cybernetics*, vol. 11, pp. 1307–1317, 2020.
[2] Y. Wang, "LSTM-based stock price prediction with sentiment and blockchain integration," *Journal of Computational Science*, 2025
[3] S. Rajput, "Forecasting stock market prices using LSTM," *IJARIIT*, vol. 11, no. 5, 2025.
[4] S. Rao et al., "Stock price prediction using LSTM neural networks," 2024.
[5] C. Wu et al., "Optimized LSTM model using genetic algorithms for stock prediction," *ACM*, 2025. [6]
[6] D. Vallarino, "AI-enhanced forecasting integrating LSTM and transformer sentiment," *Journal of Economic Analysis*, 2025.
[7] Y. Wang et al., "LSTM model with sentiment integration for stock prediction," 2025.
[8] "CNN-LSTM hybrid model for stock price prediction," *OICC Press*, 2024.
[9] M. S. Sonani et al., "Hybrid LSTM-GNN model for stock prediction," *arXiv*, 2025.
[10] A. Sarkar and G. Vadivu, "Ensemble deep learning model using VAE, Transformer, and LSTM," *arXiv*, 2025.
[11] S. B. Amiri, "Comparison of ARIMA, LSTM, and GRU models for stock prediction," 2025.
[12] Q. Li et al., "Deep learning framework for stock return prediction," *Nature Scientific Reports*, 2024.
[13] S. Mehtab et al., "CNN and LSTM-based models for stock forecasting," *arXiv*, 2020.
[14] J. Sen et al., "LSTM-based stock prediction for portfolio optimization," *arXiv*, 2022.
[15] S. Giantsidi et al., "Deep learning for financial forecasting: A review," *Elsevier*, 2025.[17]
[16] P. H. Vuong et al., "Bibliometric review of stock price forecasting models," *Science Progress*, 2024.
[17] H. H. Htun et al., "Feature selection techniques for stock prediction using ML," *Springer*, 2023.
[18] A. Rohan et al., "Artificial intelligence in financial market prediction," *Frontiers in AI*, 2025.
[19] H. C. Bhanujyothi, "Hybrid model combining ML and statistical techniques for stock prediction," 2025.
[20] "Hybrid deep learning and statistical approach for stock forecasting," *IJARET*, 2025.