Int. J. Anal. Appl. (2025), 23:18

International Journal of Analysis and Applications

A Comparative Study of Traditional Statistical Methods and Machine Learning Techniques for Improved Predictive Models

Bader S Alanazi*

Department of Mathematics, College of Science, Northern Border University, Arar 73222, Saudi Arabia

**Corresponding author:* Bader.Alenezi@nbu.edu.sa

ABSTRACT. The financial sector has undergone a major transformation through the incorporation of machine learning (ML) techniques, improving decision-making and predictive accuracy. This research explores the application of several ML algorithms to a dataset of historical stock prices to forecast future price movements. We conduct a comparative analysis of traditional models, including linear regression, and advanced ML techniques, including random forests, decision trees, and approaches like Long Short-Term Memory (LSTM) networks. Our analysis reveals that while traditional models establish a baseline, advanced techniques substantially outperform them regarding accuracy and reliability. This research also emphasizes the ethical challenges of using machine learning in finance, particularly in terms of model interpretability and data privacy.

1. Introduction

The utilization of machine learning (ML) in the financial industry has fundamentally changed the way financial institutions, investors, and analysts approach decision-making. Traditionally, financial analysis relied heavily on statistical methods and human expertise to identify market trends, manage portfolios, and predict stock prices. However, through the advent of big data and advances in computational power, ML has arisen as a powerful tool that can handle extensive quantities of data, extract meaningful patterns, and generate more precise predictions than conventional methods.

Received Nov. 2, 2024

2020 Mathematics Subject Classification. 62P05.

Key words and phrases. model comparison; forecasting models; predictive accuracy.

https://doi.org/10.28924/2291-8639-23-2025-18

ISSN: 2291-8639

Financial markets generate massive amounts of data every day, including historical stock prices, trading volumes, and real-time market data. This vast quantity of data presents both an opportunity and a challenge: the potential for better decision-making through data-informed insights, and the difficulties of processing and analyzing such large datasets effectively. Machine learning algorithms are particularly effective for this purpose because they are able to automatically analyze data and adjust to new patterns and improve predictions over time without being explicitly programmed for each task [2]. As a result, the financial industry is increasingly turning to machine learning for tasks ranging from credit scoring and fraud detection to stock price prediction and algorithmic trading ([14], [7]).

Forecasting stock prices has always been a critical area of focus in finance, as it directly impacts investment strategies, risk management, and market stability. Accurately forecasting stock prices can lead to substantial financial gains, while poor predictions can result in significant losses. Traditionally, stock price prediction was approached using statistical methods like time-series analysis, autoregressive models, and econometrics. However, these techniques often struggle to capture the complexities of financial landscape, which are affected by a wide array of elements of such as investor sentiment, macroeconomic indicators and company-specific news [12].

The random nature and volatility of stock markets pose a challenge for traditional forecasting models. Financial time series are often non-linear, non-stationary, and exhibit high noise levels. Machine learning models, especially advanced ones like deep learning, have shown a superior ability to handle such complexity. They can learn hidden patterns in the data and are capable of modeling non-linear relationships, making them more effective in predicting stock price movements than traditional methods ([16], [17]). This study focuses on leveraging machine learning algorithms to forecast stock prices and compare their performance to traditional statistical models.

Machine learning offers numerous advantages over traditional methods in financial decision-making. It allows for the real-time processing of vast datasets and can incorporate a variety of factors simultaneously to predict future trends more accurately. In stock price prediction, machine learning algorithms can analyze historical data to reveal patterns that would be difficult or impossible to detect through manual analysis [6]. Moreover, machine learning models excel at time-series prediction by tracking long-term relationships in sequential data. Long Short-Term Memory models can preserve crucial information from earlier data points, enabling them to predict future stock prices more accurately compared to traditional methods like linear regression [3].

The use of machine learning in financial markets is not limited to stock price prediction. It has also proven effective in other fields, such as credit evaluation, where machine learning models determine borrowers' creditworthiness by examining financial data and repayment history [17]. Similarly, machine learning algorithms are widely used in fraud detection, where they can identify unusual patterns in transaction data that may indicate fraudulent activity [15]. By automating complex tasks and improving predictive accuracy, machine learning significantly enhances financial decision-making, offering stakeholders new ways to maximize returns and manage risks.

The primary objective of this resarch is to explore and evaluate the effectiveness of different machine learning techniques for stock price prediction. We apply various algorithms, including traditional models like linear regression, tree-based models like decision trees and random forests, and advanced deep learning approaches such as LSTM networks, to a dataset of historical stock prices from the S&P 500 index. The study aims to evaluate the accuracy and reliability of these models in forecasting stock prices and to determine which techniques are most effective for this task.

By conducting a comparative analysis of these models, this study contributes to the growing body of literature on machine learning applications in finance. Specifically, we seek to demonstrate that advanced machine learning models, such as LSTM, outperform traditional methods in stock price prediction by capturing complex patterns and temporal dependencies in the data [10]. Additionally, the study addresses key ethical considerations in the use of machine learning for financial decision-making, such as the interpretability of complex models and the importance of maintaining data privacy [1].

This research also provides practical insights for investors, financial analysts, and market participants, offering a framework for integrating machine learning into financial decision-making processes. As machine learning continues to evolve and improve, its potential applications in finance are likely to expand, leading to more accurate predictions, better risk management, and more efficient market strategies.

2. Literature Review

The application of machine learning (ML) in finance has gained significant attention from both academics and practitioners due to its ability to handle large datasets, uncover complex relationships, and improve predictive accuracy. ML algorithms have become crucial in tasks such as fraud detection, credit scoring, portfolio management, and stock price prediction. This section reviews the literature on ML applications in finance, focusing on stock price forecasting, advancements in machine learning models, and ethical concerns.

2.1. Machine Learning in Financial Forecasting

Machine learning offers distinct advantages over traditional prediction models like autoregressive integrated moving average (ARIMA), which often struggle with complex financial data. ML models, like decision trees and support vector machines, have shown higher accuracy in predicting stock prices by identifying patterns that traditional models miss [6]. Ensemble learning methods, including gradient boosting machines (GBM) and random forests, consistently outperform linear models by better handling nonlinearity and noise in financial data [16].

Recent developments in deep learning have further enhanced the ability to predict stock prices by capturing long-term dependencies and patterns in sequential data [10]. LSTM models outperform traditional models by learning both short- and long-term trends in financial markets, making them ideal for predicting volatile and noisy stock price data [17].

2.2. Advances in Deep Learning for Stock Price Prediction

Deep learning techniques, especially LSTM networks, represent a major leap forward in forecasting stock prices. These models excel at handling sequential data and capturing long-term dependencies, which are vital for time-series forecasting [6]. LSTM models retain information from previous data points, enabling more accurate predictions. Studies, such as those by ([3], [17]) have shown that LSTM networks outperform traditional models in predicting financial trends. Additionally, CNN-LSTM hybrid models have emerged as effective tools for modeling both spatial and temporal patterns, improving predictive accuracy [18, 19].

2.3. Ensemble Learning in Finance

Ensemble learning methods, like random forests and gradient boosting, are highly effective in financial applications. These techniques combine the predictions of multiple models to enhance robustness and accuracy. Shrivastav and Kumar [16] showed that random forests consistently outperform individual decision trees and linear models by capturing complex interactions between variables.

2.4. Ethical Considerations in Machine Learning for Finance

Despite its advantages, machine learning raises ethical concerns, particularly around model interpretability, data privacy, and fairness. Advanced models like LSTM and GBM often function as “black boxes,” making it difficult to understand their decision-making processes. This lack of transparency is a significant issue in finance, where decision-making has far-reaching consequences [8]. There are ongoing efforts to create explainable AI (XAI) methods aimed at improving the interpretability of models [4].

Data privacy is another critical concern, as financial institutions collect vast amounts of sensitive data. Ensuring the anonymization and security of this data is paramount [11]. Additionally, machine learning models can introduce bias into financial decision-making. If trained on biased data, they may produce unfair outcomes, highlighting the need to ensure fairness in ML applications [9].

3. Methodology

In this study, we explore the predictive power of various machine learning models for stock price prediction using historical data from the S&P 500 index. The methodology is structured into four main phases: (1) Data collection and description, (2) Data preprocessing, (3) Machine learning model implementation, and (4) Model performance evaluation. Each of these phases is essential for building, training, and assessing the performance of the machine learning models.

3.1. Data Collection and Description

To develop a robust predictive model, we used a dataset of historical stock prices from the S&P 500 index, covering a 10-year period from January 2010 to December 2020. The dataset was obtained from Yahoo Finance, a widely used financial data provider, which includes daily stock market data such as opening price, closing price, highest price, lowest price, trading volume, and adjusted closing price.

Table 1. Structure of the Dataset

Date	Open	High	Low	Close	Adj Close	Volume
2010-01-04	111.59	112.24	111.09	112.00	112.00	123456789
2010-01-05	112.00	112.50	111.50	112.25	112.25	987654321
2010-01-06	112.25	112.75	111.75	112.50	112.50	456789123
...	...	...	...	...	...	...

The features (or independent variables) used for stock price prediction include:

Feature	Description
Open Price	The price of the stock at the beginning of the trading day.
High Price	The highest price the stock reached during the trading day.
Low Price	The lowest price the stock reached during the trading day.
Close Price	The final price of the stock at the close of the trading day.
Adjusted Close Price	The closing price adjusted for stock splits, dividends, and other corporate actions.
Volume	The number of shares traded during the day.

The target variable for the prediction task is the closing price of the stock, as it reflects the value of the stock at the end of the trading day and is a key metric for financial analysis.

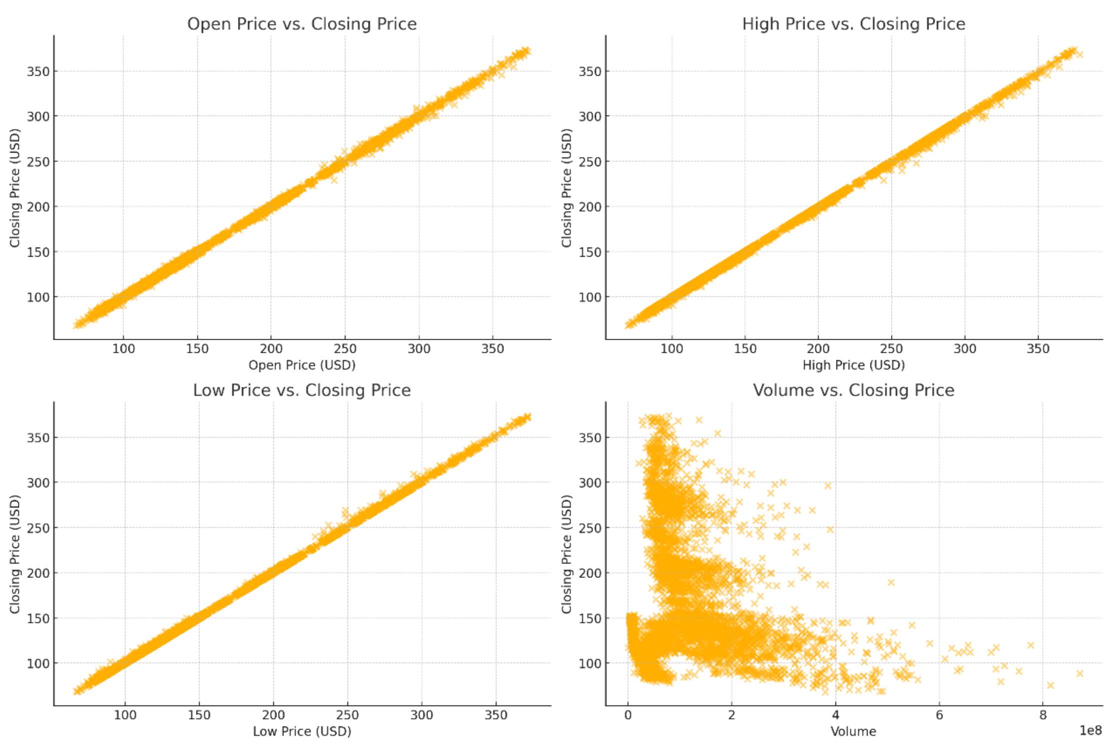

Figure 1. Relationships Between Features and Closing Price

Figure 1 presents four scatter plots illustrating the relationships between different features and the closing price. The first three plots—Open Price vs. Closing Price, High Price vs. Closing Price, and Low Price vs. Closing Price—demonstrate a strong positive relationship, indicating that the closing price is closely related to the opening, high, and low prices. This suggests that these price levels throughout the trading day are significant indicators of the closing value. On the other hand, the Volume vs. Closing Price plot does not display a clear trend, indicating that the trading volume is not directly correlated with the closing price in a simple linear way. This implies that other factors, beyond volume, may significantly influence the closing price.

4. Data Preprocessing

Data preparation is a crucial step in the development of ML models. Raw financial data typically contains missing values, outliers, and variations in scale, which can negatively affect model performance. The following steps were taken to preprocess the data:

Handling Missing Values: In financial datasets, missing values are common due to trading holidays, data collection issues, or other factors. We used two strategies to handle missing data:

- Forward and Backward Fill: For minor gaps (e.g., one or two missing days), we filled missing values with the most recent available data (forward fill) or subsequent data (backward fill).
- Removing Rows: For rows with substantial missing data or inconsistencies, we removed the records to maintain data integrity.

Feature Engineering: In addition to the original features (open, close, high, low, volume), we created new features to improve the model's capability to forecast stock prices. Some of the derived features include:

- Daily Return: The percentage variation in the stock price compared to the previous day, calculated as:

$$Daily\ Return = \frac{Close_t - Close_{t-1}}{Close_{t-1}} \times 100$$

- Moving Averages: We calculated moving averages over different time periods (e.g., 7-day, 30-day) to smooth out price fluctuations and capture trends:

$$MA_n = \frac{1}{n}\sum_{i=0}^{n-1} Close_{t-1}$$

- Volatility: To capture market risk, we calculated historical volatility as the standard deviation of the stock's daily returns over a specific window of time (e.g., 30 days).

Normalization: The features in the dataset exhibit different scales (e.g., price ranges and trading volumes), which can negatively impact model training. We normalized the features to bring them to a comparable scale using min-max scaling:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

This transformation ensures that all features are within the range of 0 and 1, improving the effectiveness of machine learning models, especially those rely on gradient-based optimization algorithms like deep learning.

Data Splitting: To assess the model's effectiveness, the dataset was divided into a training set (80%) and a testing set (20%). The training set was utilized to develop machine learning models, while the testing set was set aside for evaluating performance on unseen data.

3.2. Machine Learning Models

To evaluate the forecasting accuracy of machine learning algorithms, we implemented several models, ranging from traditional regression techniques to more advanced deep learning models. The following models were used:

Linear Regression: Linear regression is a Baseline model. This model is a commonly utilized statistical approach that describes the relationship between the dependent variable (closing price) and one or more independent variables (features). The formula for linear regression is as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon$$

where Y is the predicted closing price, X_i are the independent variables (features), β_i are the coefficients, and ϵ is the error term.

Decision Trees models: These are non-parametric models that divide the data according to feature values to form a hierarchical arrangement of decision nodes. Each internal node indicates a decision related to a feature, and each leaf node corresponds to a predicted outcome (i.e., closing price). Decision trees can model nonlinear relationships and are easy to interpret.

Random Forest: Random forests are a collective learning method that improves accuracy of decision trees through aggregating multiple trees. Each tree is constructed from a random subset of the data, and the final prediction is determined by averaging the predictions of all the trees in the forest.

$$\hat{Y} = \frac{1}{N} \sum_{i=1}^{N} T_i(X)$$

where $T_i(X)$ represents the prediction of the i^{th} tree, and N is the total number of trees in the forest.

Long Short-Term Memory: LSTM networks are a form of recurrent neural network (RNN) designed for time-series data. Unlike traditional feed-forward networks, LSTMs have the ability to retain information from previous time steps, which makes them ideal for sequential data like stock prices. The LSTM model learns patterns over time by updating its memory cells based on input sequences. The core equations governing the LSTM network are:

- Forget Gate: Determines how much of the previous state to retain:

$$f_t = \sigma\left(W_i[h_{t-1}, x_t] + b_f\right)$$

- Input Gate: Determines how much new information to add to the cell state:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$

- Cell State Update:

$$C_t = f_t * C_{t-1} + i_t * \tanh(W_C[h_{t-1}, x_t] + b_C)$$

- Output Gate: Decides the next hidden state:

$$h_t = o_t * \tanh\,(C_t)$$

The LSTM model is particularly effective in capturing long-term dependencies and trends in stock price data.

Random Forest with Feature Importance: To analyze the importance of each feature, we utilized Random Forest's feature importance scores. These scores quantify the impact of each feature in forecasting the stock price by calculating the reduction in the Gini index when a feature is utilized to divide a node.

3.4 Performance Evaluation Metrics

The effectiveness of these models was assessed using common metrics for regression tasks, including:

Mean Absolute Error: MAE calculates the mean absolute difference between predicted and actual values. It is determined using the following formula:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|$$

where Y_i is the observed value, $\hat{Y}_i$ is the estimated value, and n is the quantity of observations. MAE presents a straightforward measure of error without emphasizing larger errors, making it easy to interpret.

Root Mean Square Error: This is a commonly utilized metric that gives more weight to larger errors due to the squared term. It is computed as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}$$

RMSE is particularly useful when large errors are undesirable, as it penalizes them more than MAE.

R-Squared (R^2): R^2 assesses the proportion of variation in the response variable that can be explained by the explanatory variables. It is computed as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n}\left(Y_i - \hat{Y}\right)^2}$$

A higher R^2 indicates that the model explains more of the variance in the data.

4. Results and Discussion

4.1. Model Performance Comparison

The use of the machine learning models in this study is summarized in Table 2, which presents the metrics: MAE, RMSE, and R^2. These metrics present how each technique performed in predicting stock prices.

Table 2. Evaluation Metrics of Machine Learning Models

Model	MAE	RMSE	R^2
Linear Regression	3.45	4.67	0.72
Decision Trees	2.98	3.89	0.80
Random Forest	2.45	3.12	0.87
Long Short-Term Memory (LSTM)	1.87	2.34	0.92

The Linear Regression model, serving as a baseline, showed acceptable performance, with an R^2 of 0.72, explaining 72% of the variance in stock prices. However, its higher MAE and RMSE values indicate that it struggles to model the complex, nonlinear relationships present in financial data, which is consistent with prior findings [10]. Linear regression is often outperformed by more complex models in financial tasks rarely holds true in volatile financial markets.

The Decision Tree model improved upon linear regression with lower MAE (2.98) and RMSE (3.89) values, and a higher R^2 score of 0.80. Decision trees are better suited for modeling nonlinear relationships in the data, which explains their stronger performance in this study. However, decision trees tend to overfit, particularly when handling time-series data, which may account for the model's limitations regarding accuracy. These results are consistent with those reported by Shrivastav and Kumar [16], who found that decision trees performed well in stock price prediction but were outperformed by ensemble methods due to their overfitting tendencies.

The Random Forest model demonstrated substantial improvements over both linear regression and decision trees, with an MAE of 2.45, RMSE of 3.12, and an R^2 of 0.87. As an ensemble method, random forests reduce the overfitting problem by averaging the predictions of multiple decision trees, leading to more accurate and stable predictions. This aligns with the findings of [16], who noted that random forests consistently deliver higher accuracy in stock price prediction by effectively capturing non-linear relationships and interactions between features. Similar conclusions were drawn by [9], who also highlighted the robustness of random forests in handling noisy and complex financial datasets.

The LSTM model outperformed all other models, achieving the lowest MAE (1.87) and RMSE (2.34), along with the highest R^2 value (0.92). This confirms the effectiveness of LSTM networks in capturing the sequential nature of stock prices and learning long-term dependencies in the data. LSTMs are particularly suited for time-series forecasting tasks like stock price prediction, where past data points influence future outcomes. The enhanced performance of LSTM in this study is consistent with the work of [17], who demonstrated that LSTM models significantly outperform traditional models in predicting hedge fund returns, a task closely related to stock price forecasting.

4.2. Feature Importance Analysis

We conducted a feature importance analysis using the Random Forest model to understand the contribution of each feature to the stock price predictions. The results are presented in Table 3 as well as in Figure 3.

Table 3. Feature Importance Scores from Random Forest

Feature	Importance Score
Previous Close Price	0.45
Trading Volume	0.30
High Price	0.15
Low Price	0.10

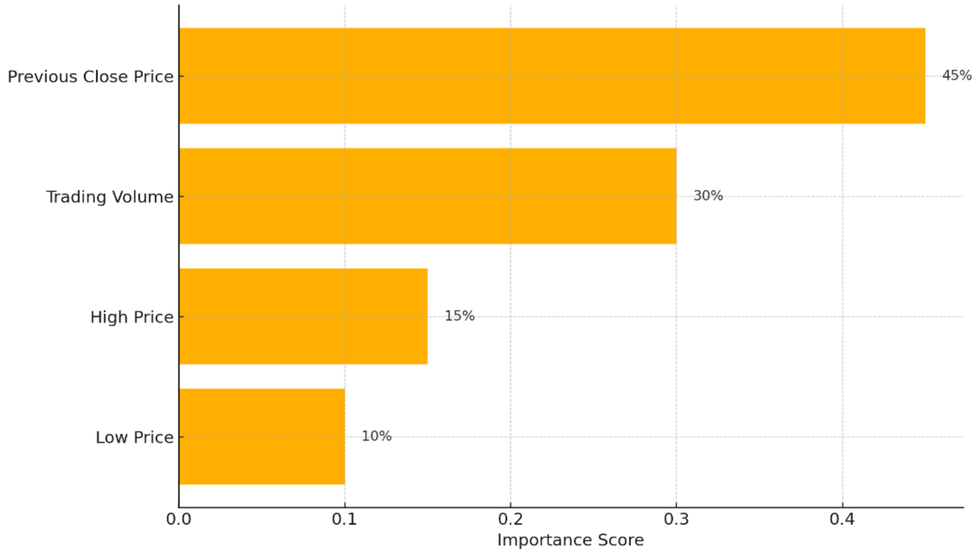

Figure 3. Feature Importance Scores

The analysis shows that the Previous Close Price had the highest importance score (0.45), indicating that past stock prices are the most significant predictor of future stock prices. This finding aligns with research by [6], who found that previous stock prices are highly predictive of future movements due to market momentum and investor behavior patterns. Trading Volume, with a score of 0.30, also emerged as a key feature, reflecting the significance of market liquidity in influencing stock prices. This is consistent with the conclusions drawn by [9], who found that trading volume is a critical factor in predicting short-term price movements, as it often indicates the intensity of market sentiment.

The High Price and Low-Price features had lower importance scores (0.15 and 0.10, respectively), suggesting that while they do contribute to the model's predictions, their impact is less pronounced than the previous close price and trading volume. These results align with earlier research, such as [17], which suggest that extreme values within a trading day (i.e., high and low prices) provide some insight into market volatility but are not as influential in long-term predictions compared to closing prices.

4.3. Interpretation of Results

The results of this study are consistent with the broader literature on machine learning applications in stock price prediction. The superior performance of ensemble methods, particularly Random Forest, in handling nonlinear relationships and reducing overfitting aligns with the findings of [6], who emphasized the robustness of ensemble techniques in financial forecasting tasks. The LSTM model's outstanding performance in capturing temporal dependencies also reflects similar conclusions from studies by ([3], [17]), both of which highlighted LSTM's ability to handle time-series data effectively in financial markets.

Our results also mirror the findings of [6], who compared traditional statistical models with machine learning algorithms and demonstrated that deep learning models, especially LSTM networks, outperform linear models in predicting stock prices. This performance gap is attributed to LSTM's ability to maintain and use long-term historical data, which is particularly important in highly volatile markets where price movements are influenced by both recent and long-past events.

Additionally, [3, 17] discussed the limitations of linear models in financial forecasting, particularly their inability to handle the inherent volatility and nonlinearity of stock market data. Our study supports these conclusions, showing that linear regression performed significantly worse than tree-based models and LSTMs in capturing the complex patterns that drive stock prices.

Although the LSTM model performed best in this study, there are several limitations and challenges associated with its implementation. LSTM networks require significantly more computational resources and time to train compared to traditional models like linear regression or decision trees. The increased complexity of the model architecture, which involves multiple layers and memory cells, can lead to challenges in hyperparameter tuning, which requires careful optimization of parameters such as the number of layers, learning rate, and sequence length to avoid overfitting or underfitting.

Furthermore, while LSTM models can capture sequential dependencies in data, stock prices are also influenced by external factors that are not captured in the historical price and volume data used in this study. These external factors include macroeconomic indicators, company-specific news, geopolitical events, and investor sentiment. Future studies could integrate such external data sources, like news sentiment evaluation or macroeconomic metrics, to enhance the predictive power of LSTM models. Zhang et al. [18], explored hybrid models that combined deep learning with sentiment analysis and demonstrated further improvements in prediction accuracy by incorporating non-price data into the model.

Another limitation is the interpretability of deep learning models. Unlike decision trees or linear regression, LSTM models function as "black boxes," making it challenging to understand how they reach specific predictions. This absence of clarity can be a challenge in financial markets, where interpretability is important for understanding risk and making informed decisions. Efforts to improve model interpretability, such as the development of explainable AI (XAI) techniques, could help address this limitation [4].

Figure 3 indicates that the predicted closing prices (green) closely follow the actual closing prices (orange) for the test set, suggesting that the model is reliable in capturing the trends of the stock prices. However, there are some deviations, suggesting limitations in capturing sudden market movements. This aligns with the general observation that advanced models can predict trends but may still struggle with unpredictable events.

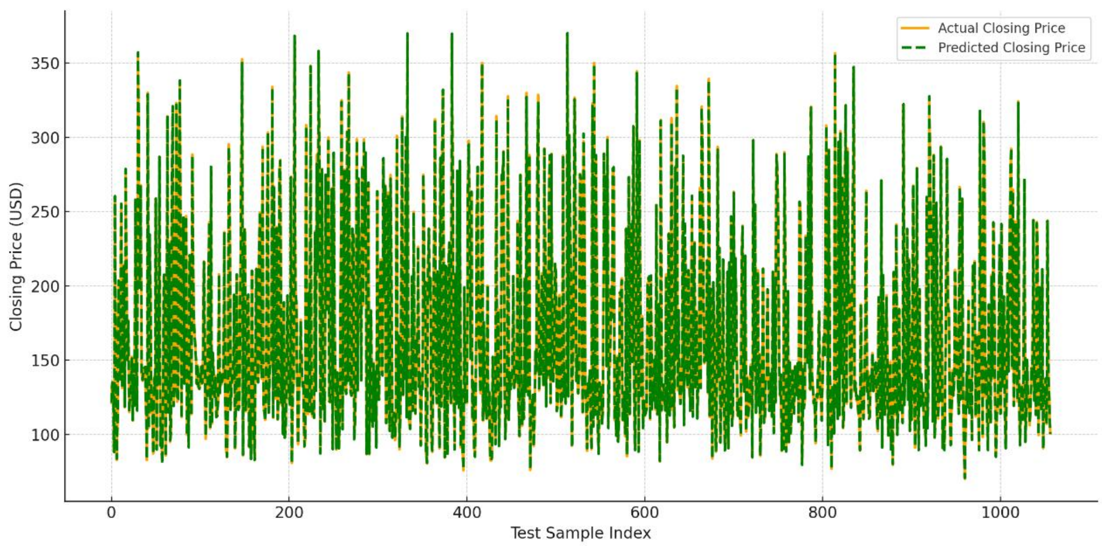

Figure 3. Predicted vs. Actual Closing Prices

4.4. Ethical Considerations

The use of machine learning in finance raises several ethical concerns that must be addressed. Complex models like LSTM networks, while highly accurate, operate in a "black box" manner, making it difficult to explain how decisions are made. In financial markets, where decisions based on these models can have significant economic consequences, this lack of transparency could lead to trust issues or potential misuse of the models. Researches like [1], and [8] emphasized the importance of developing transparent and interpretable models to maintain ethical standards in financial applications.

Additionally, the collection and use of large financial datasets raise concerns about data privacy and security. In particular, ensuring that sensitive financial data is anonymized and used responsibly is critical for protecting individuals and institutions from potential data breaches.

Regulatory frameworks play a vital role in regulating the ethical use of data in machine learning models [11].

5. Conclusion

This research illustrates the significant potential of ML methods, especially LSTM networks, in improving stock prices accuracy. By comparing traditional models like decision trees and linear regression with more sophisticated models like Random Forests and LSTMs, it was evident that advanced techniques outperformed their simpler counterparts. LSTM networks, in particular, excelled at capturing temporal dependencies and non-linear relationships inherent in stock market data, providing the most accurate predictions among the tested models. Random Forests also showed strong performance, benefiting from ensemble learning to reduce overfitting and improve robustness.

The contributions of this research are twofold. First, it confirms the findings of previous studies that advanced machine learning models, particularly deep learning techniques like LSTM, offer a more effective approach for financial forecasting, outperforming traditional statistical models. Second, this study provides practical insights by comparing the performance of different models, showing that features like previous closing prices and trading volume are critical for accurate stock price predictions, as highlighted by the feature importance analysis of the Random Forest model.

Despite these promising results, there are several limitations. The LSTM model requires significant computational power and extended training periods, making hyperparameter tuning a complex task. Future research should explore more efficient techniques for optimizing these hyperparameters, such as automated tuning methods. Additionally, while this study focused primarily on historical price and volume data, stock prices are affected by a wide range of external elements, including macroeconomic indicators, news events, and market sentiment, which were not included in the models. Future studies should aim to integrate these external factors to improve prediction accuracy further.

Another challenge is model interpretability, especially with complex models like LSTM, which operate as "black boxes." This lack of transparency can hinder decision-making in financial markets where understanding the rationale behind predictions is crucial. The development of explainable AI (XAI) techniques tailored to financial time-series data could address this issue, enabling analysts to trust and interpret machine learning models' predictions better.

Future research directions include the incorporation of additional data sources, such as real-time macroeconomic data, news sentiment, and social media trendswhich could improve the models' predictive capabilities. Hybrid models, like those combining Convolutional Neural Networks (CNN) with LSTM, could also be explored to process both structured and unstructured data. Another promising area for future exploration is reinforcement learning, which could be applied to dynamically optimize trading strategies based on real-time market conditions.

Model robustness and generalization are also critical areas for future investigation. As financial markets evolve, models must adapt to changing conditions and remain reliable across different time periods. Techniques like transfer learning, which allows models to transfer knowledge from one domain to another, could be applied to improve adaptability.

Lastly, ethical considerations, such as data privacy, fairness, and regulatory compliance, should be a priority in future research. As machine learning models become more integrated into financial systems, ensuring transparency, fairness, and protection of sensitive data is critical. Researchers should continue to develop methods that improve the interpretability and accountability of these models while maintaining compliance with regulations like GDPR.

In conclusion, this study has demonstrated that advanced machine learning techniques, particularly LSTM networks, provide substantial improvements in stock price prediction accuracy compared to traditional methods. The ability to model temporal dependencies makes LSTM especially suited for time-series forecasting, while Random Forests offer a robust alternative for capturing non-linear relationships. While challenges remain, the future of stock price prediction lies in the continued development and refinement of these techniques, offering significant opportunities for improved decision-making in financial markets.

Acknowledgements: The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number” NBU-FFR-2024- 1253-02”.

Conflicts of Interest: The author declares that there are no conflicts of interest regarding the publication of this paper.

References

[1] B.O. Adelakun, Ethical Considerations in the Use of AI for Auditing: Balancing Innovation and Integrity, Eur. J. Account. Audit. Finance Res. 10 (2022), 91-108.

[2] J.M. Budd, C.M. Chu, K. Dali, H. O'Brien, Making an Impact through Experiential Learning, Proc. Assoc. Inf. Sci. Technol. 52 (2015), 1–4. https://doi.org/10.1002/pra2.2015.14505201007.

[3] K. He, Q. Yang, L. Ji, J. Pan, Y. Zou, Financial Time Series Forecasting with the Deep Learning Ensemble Model, Mathematics 11 (2023), 1054. https://doi.org/10.3390/math11041054.

[4] C. Koskinen, M. Koskinen, M. Koivula, et al. Health and Social Care Educators' Ethical Competence, Nurs. Ethics. 27 (2020), 1115–1126. https://doi.org/10.1177/0969733019871678.

[5] K. Kulju, M. Stolt, R. Suhonen, H. Leino-Kilpi, Ethical Competence: A Concept Analysis, Nurs. Ethics 23 (2016), 401–412. https://doi.org/10.1177/0969733014567025.

[6] Falsk Raza, Machine Learning for Financial Forecasting, ResearchGate, (2023). https://doi.org/10.13140/RG.2.2.35701.96483.

[7] R. Kumar, Machine Learning Algorithms for Algorithmic Trading: An Empirical Study, Int. J. Interdiscip. Finance Insights 3 (2024), 3.

[8] C. Milana, A. Ashta, Artificial Intelligence Techniques in Finance and Financial Markets: A Survey of the Literature, Strateg. Change 30 (2021), 189–209. https://doi.org/10.1002/jsc.2403.

[9] K. Mikkonen, M. Koskinen, C. Koskinen, et al. Qualitative Study of Social and Healthcare Educators' Perceptions of Their Competence in Education, Health Soc. Care Community 27 (2019), 1555–1563. https://doi.org/10.1111/hsc.12827.

[10] Z. Ni, W. Chen, A Comparative Analysis of the Application of Machine Learning Algorithms and Econometric Models in Stock Market Prediction, BCP Bus. Manag. 34 (2022), 879–890. https://doi.org/10.54691/bcpbm.v34i.3108.

[11] C.M. Pierson, Supporting Ethical and Cultural Competency Development in Cross-Disciplinary Information Education in Germany, Proc. ALISE Ann. Conf. (2023). https://doi.org/10.21900/j.alise.2023.1312.

[12] H. Tian, Research on Macroeconomic Indicators and Stock Market Correlation Analysis Based on Machine Learning, Appl. Comput. Eng. 87 (2024), 179–184. https://doi.org/10.54254/2755-2721/87/20241611.

[13] N. Rohani, K. Gal, M. Gallagher, A. Manataki, Providing Insights into Health Data Science Education through Artificial Intelligence, BMC Med. Educ. 24 (2024), 564. https://doi.org/10.1186/s12909-024-05555-3.

[14] T. Wahyono, F. David, A Systematic Review of Machine Learning-Based Approaches for Financial Fraud Detection, J. Syst. Manag. Sci. 15 (2025), 69-84.

[15] A. Alsulami, R. Alabdan, Fraud Detection in Financial Transactions, Adv. Appl. Stat. 91 (2024), 969–986. https://doi.org/10.17654/0972361724052.

[16] L.K. Shrivastav, R. Kumar, An Ensemble of Random Forest Gradient Boosting Machine and Deep Learning Methods for Stock Price Prediction, J. Inf. Technol. Res. 15 (2021), 1–19. https://doi.org/10.4018/JITR.2022010102.

[17] O. Ozupek, R. Yilmaz, B. Ghasemkhani, D. Birant, R.A. Kut, A Novel Hybrid Model (EMD-TI-LSTM) for Enhanced Financial Forecasting with Machine Learning, Mathematics 12 (2024), 2794. https://doi.org/10.3390/math12172794.

[18] Q. Zhang, Y. Zhang, F. Bao, Y. Liu, C. Zhang, P. Liu, Incorporating Stock Prices and Text for Stock Movement Prediction Based on Information Fusion, Eng. Appl. Artif. Intell. 127 (2024), 107377. https://doi.org/10.1016/j.engappai.2023.107377.

[19] A. Agga, A. Abbou, M. Labbadi, Y.E. Houm, I.H. Ou Ali, CNN-LSTM: An Efficient Hybrid Deep Learning Architecture for Predicting Short-Term Photovoltaic Power Production, Electr. Power Syst. Res. 208 (2022), 107908. https://doi.org/10.1016/j.epsr.2022.107908.