International Journal Artificial Intelligent and Informatics
Vol. 2, No. 2, July 2024, pp. 47~55
ISSN: 2622-626X, DOI: http://doi.org/10.33292/ijarlit.v2i2.36

Enhancing Stock Price Predictions: Leveraging LSTM for Accurate Forecasting of Ecopetrol's Stock Performance

Jaminton Muñoz[1*], K Castaño[1]
[1]Universidad Tecnológica de Pereira, Colombia
Email: jamintonmunoz@gmail.com

Article Info

Article history:

Received: September 20, 2023
Revised: December 12, 2023
Accepted: March 30, 2024
Available Online: July 30, 2024

Keywords:

Stock Price Prediction
Long Short-Term Memory (LSTM)
Time Series Forecasting
Financial Data Analysis
Ecopetrol Stock
Emerging Market

ABSTRACT

Accurate stock price prediction remains a significant challenge in financial forecasting, particularly for emerging market stocks. This study investigates the efficacy of Long Short-Term Memory (LSTM) networks in forecasting the stock prices of Ecopetrol (EC), Colombia's largest oil and gas company. Using historical stock data from Yahoo Finance spanning September 18, 2018, to September 18, 2023, we developed an LSTM model to capture complex temporal patterns in the stock market. The model's performance was evaluated using a range of metrics including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Scaled Error (MASE). The results demonstrate that the LSTM model achieves low MAE (0.2509652), MSE (0.11678666), and RMSE (0.34174064), alongside a MAPE of 2.071206, indicating high accuracy and reliability in predicting stock prices. Although the MASE of 1.125679 suggests that the model performs similarly to a naive forecasting approach, it still provides valuable insights into stock price movements. This study highlights the effectiveness of LSTM in handling sequential data and capturing intricate stock price patterns, while suggesting that future improvements could be made by optimizing the model further and integrating additional relevant features.

1. INTRODUCTION

The stock market is a fundamental pillar of the global economic system, playing a key role in shaping investment strategies, government policies, and corporate decision-making [1]. Predicting stock price movements has long been a focus of financial research, as accurate forecasts can significantly influence risk management and resource allocation [2]. One stock of particular interest is Ecopetrol (EC), Colombia's largest oil and gas company, which is critical not only to the Colombian economy but also to global energy markets. Given the company's central role in the oil and gas sector, predicting its stock price fluctuations could provide valuable insights for investors, policymakers, and other stakeholders navigating the volatile energy landscape [3]. However, the inherent complexity and volatility of stock price movements make accurate forecasting a formidable challenge[4].

Traditional methods for stock price prediction, such as technical analysis and fundamental analysis, have long been employed in financial markets [5], [6]. Technical analysis relies on historical price and volume data to identify patterns, while fundamental analysis examines financial statements, economic indicators, and market conditions to assess a stock's intrinsic value [7]. While these methods offer useful insights, they are limited in their ability to capture the dynamic and non-linear nature of stock price movements [8]. Specifically, these traditional approaches often struggle to account for complex temporal dependencies in time series data, which are critical for accurate predictions in volatile markets such as oil and gas. This has spurred the adoption of machine learning models, which are better equipped to handle the complexity of financial time series.

One such machine learning method that has gained traction in recent years is Long Short-Term Memory (LSTM), a type of recurrent neural network (RNN) designed to learn long-term dependencies in sequential data [9], [10]. LSTM models are particularly effective at overcoming the vanishing gradient problem that plagues traditional RNNs, enabling them to capture long-range correlations in time series data[11]. This capability makes LSTM especially useful for forecasting stock prices, which are influenced by a wide range of factors over varying time horizons [12]. Several studies have demonstrated the efficacy of LSTM in stock price prediction, showing its superiority over conventional methods when dealing with the non-linearity and volatility of financial data [13].

Despite the proven success of LSTM models in stock markets, most of the research has focused on developed markets with relatively efficient and stable financial systems [14]. Emerging markets, such as Colombia, present a different set of challenges due to higher volatility, market inefficiencies, and susceptibility to external shocks. In the case of Ecopetrol, which operates in an industry subject to global oil price fluctuations, domestic political conditions, and regional market dynamics, traditional forecasting methods may fall short in providing accurate predictions. To date, there has been limited exploration of LSTM's potential in predicting stock prices in such high-risk, high-reward environments, indicating a clear research gap.

This study seeks to address this gap by applying LSTM models to predict the stock price of Ecopetrol, with a focus on capturing the unique characteristics of the Colombian market [15]. We aim to leverage LSTM's ability to model complex, non-linear relationships and long-term dependencies in time series data to improve the accuracy of stock price forecasts. By applying LSTM to Ecopetrol's stock, this research not only contributes to the growing body of literature on machine learning in finance but also provides valuable insights for investors and analysts interested in emerging markets [16]. Understanding how LSTM models perform in less efficient and more volatile markets is crucial for advancing both academic research and practical applications in stock price forecasting [17].

To ensure the robustness of our predictions, this study employs a range of performance metrics to evaluate the LSTM model's accuracy, including Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) [18]. These metrics are commonly used in time series forecasting to measure the difference between predicted and actual values, offering a comprehensive evaluation of the model's predictive performance [19]. By rigorously assessing the model's output using these metrics, we aim to provide a reliable framework for stock price forecasting in emerging markets, where traditional methods often fall short [20]. This approach underscores the importance of using advanced machine learning techniques in complex, real-world financial scenarios.

In conclusion, this study fills a critical gap in the research by applying an LSTM model to predict the stock price of Ecopetrol, a major player in the Colombian oil and gas industry. The findings will not only contribute to the growing interest in machine learning for financial forecasting but also provide practical insights for investors and decision-makers operating in volatile and unpredictable markets. The implications of this research extend beyond Ecopetrol and Colombia, offering a broader perspective on how advanced forecasting models can be applied in other emerging markets and sectors characterized by complexity and uncertainty.

2. METHOD

2.1. Data Collection

The data used in this study comprises historical stock prices of Ecopetrol (EC), obtained from Yahoo Finance, a widely recognized and reliable source for financial market data. Ecopetrol is the largest oil and gas company in Colombia and one of the most significant players in Latin America's energy sector. The dataset includes daily stock price data from September 18, 2018, to September 18, 2023, providing a substantial period for training and testing the model. This time range is chosen to cover various market conditions, including periods of high volatility due to global events such as oil price fluctuations and geopolitical tensions, as well as more stable periods.

The specific stock data used for the prediction model is the closing price, which represents the final price of the stock at the end of each trading day. Closing prices are widely used in financial forecasting due to their stability and reflection of market sentiment at the close of each trading session. By focusing on the closing price, this study aims to capture the daily stock price trends, reducing the noise associated with intraday price fluctuations, which may obscure long-term trends.

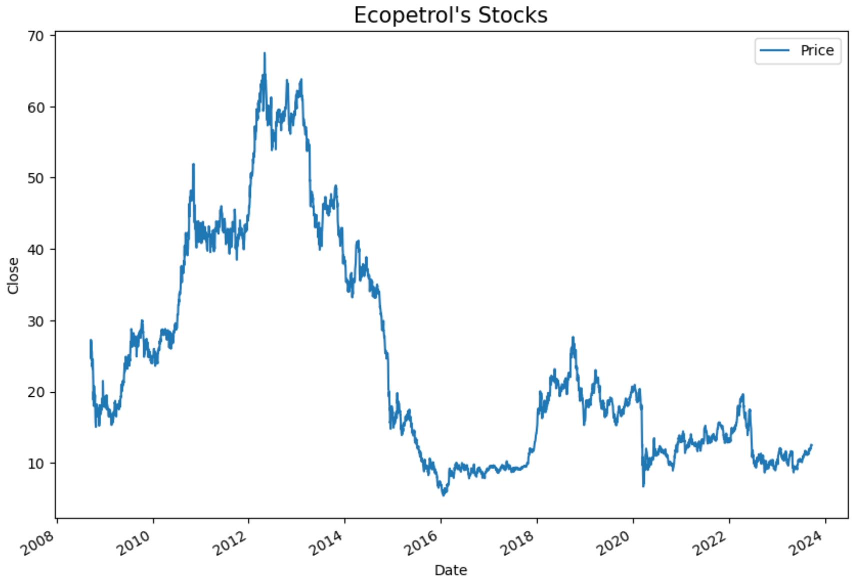

Figure 1. Ecopetrol stock chart

2.2. Data Preprocessing

Data preprocessing is a critical step in preparing the raw stock price data for input into the Long Short-Term Memory (LSTM) model [21]. This phase ensures that the data is clean, structured, and suitable for training and evaluation, thereby enhancing the accuracy and reliability of the predictive model [22]. The preprocessing steps involve several stages: handling missing data, data normalization, and splitting the dataset into training and test subsets using a time series split method. Each stage is essential for ensuring the robustness of the model and its ability to generalize well to unseen data.

2.2.1. Handling Missing Data

In financial datasets, missing values can arise due to various reasons such as market holidays or data recording errors [23]. To maintain the integrity of the time series data, it is essential to address these gaps effectively. In this study, missing data points are handled using interpolation techniques. Specifically, linear interpolation is applied to estimate missing values based on the available data points before and after the gaps. Linear interpolation assumes that the change between two known points is linear and fills in the missing values accordingly. This method is chosen for its simplicity and effectiveness in maintaining the continuity of the time series. The integrity of the dataset is preserved by ensuring that all time points are accounted for without introducing significant biases or discontinuities.

2.2.2. Data Normalization

Normalization is a crucial preprocessing step that transforms the data to a common scale, which is particularly important for neural networks like LSTM [24]. Raw stock price data can vary significantly in magnitude, which can adversely affect the training process and model performance. To address this, the stock prices are normalized using min-max scaling.

Min-max normalization is chosen because it preserves the relationships between data points and ensures that all features contribute equally to the model's training process. This transformation is essential for the LSTM model to learn effectively, as it helps prevent issues related to differing scales of input features and accelerates convergence during training.

2.2.3. Time Series Split

After preprocessing the data, the next step is to split it into training and test subsets. Unlike random splitting, which may disrupt the temporal order of data, a time series split method preserves the sequence of observations. This approach is critical for time series forecasting as it maintains the chronological order and ensures that the model is trained on past data and evaluated on future data.

In this study, the dataset is divided into 80% training data and 20% test data. This ratio is chosen to provide a substantial amount of data for training the LSTM model while retaining enough data for robust evaluation. The training set includes data from September 18, 2018, to a specified cutoff date, and the test set comprises data from the cutoff date to September 18, 2023. This split ensures that the model is trained on historical data and evaluated on a separate, subsequent period, reflecting real-world forecasting scenarios where future data is unavailable during model training.

A time series split is visualized as follows:

1. Training Period: This segment contains 80% of the dataset, used to train the LSTM model. The training set includes a sequence of daily closing prices, allowing the model to learn the underlying patterns and dependencies in the time series.
2. Test Period: This segment contains the remaining 20% of the dataset, reserved for evaluating the model's performance. The test set includes data following the training period, which is used to assess how well the model generalizes to unseen data.

The time series split is crucial for evaluating the model's ability to make accurate predictions on future data, ensuring that the performance metrics reflect the model's predictive capability in practical scenarios.

In summary, the data preprocessing phase in this study involves meticulous handling of missing values, normalization of data to ensure consistency, and a time series split to maintain chronological order. These steps are integral to preparing the data for LSTM modeling, enabling accurate and reliable stock price predictions for Ecopetrol and contributing to the overall robustness of the forecasting model.

Figure 2. Graph of train and test data split

2.2.4. Horizon and Window Definition

Long Short-Term Memory (LSTM) networks are designed to handle sequential data by learning patterns and dependencies over time. To effectively train an LSTM model for stock price prediction, the input data must be organized into sequences that capture temporal relationships. This involves defining two key parameters: horizon and window. These parameters help structure the time series data in a way that the LSTM can leverage its memory capabilities to make accurate predictions.

Horizon refers to the number of future timesteps that the model is expected to predict. In this study, the horizon is set to 1, meaning that the model aims to predict the stock price for the next day based on past data. This choice of horizon reflects a short-term forecasting approach, which is common in financial applications where daily price movements are of primary interest.

Window denotes the number of past timesteps used to make a prediction for the horizon. In this study, the window is set to 7, indicating that the model will use stock price data from the previous 7 days to forecast the price for the following day. This configuration aligns with the notion that recent historical data can provide relevant context for short-term predictions, capturing recent trends and patterns in stock price movements.

2.2.5. Data Sequencing Process

The process of creating sequences involves transforming the raw time series data into a format suitable for LSTM input. This is accomplished through the following steps:

1. Data Organization: The raw stock price data is organized into a series of input-output pairs, where each input sequence consists of a window of past timesteps and the corresponding output is the stock price for the horizon. For example, if the input window is the 7-day period from January 1 to January 7, the output would be the stock price on January 8.
2. Sequence Generation: To generate sequences, a sliding window approach is employed. The window slides over the time series data, creating multiple input-output pairs. Each pair consists of a window of 7 consecutive days of stock prices and the price on the following day. This approach ensures that the model is exposed to various sequences of data, enabling it to learn patterns and dependencies across different time periods.
3. Data Structuring: The generated sequences are then structured into the format required by the LSTM model. Specifically, the data is organized into a 3D array with dimensions corresponding to the number of sequences, the window size, and the number of features (in this case, one feature representing the stock price). This format allows the LSTM model to process the sequential input data effectively, leveraging its memory cells to capture temporal dependencies.
4. Normalization: After generating the sequences, the data is normalized to ensure consistency and improve model performance. Normalization is applied to both the input sequences and the target values, scaling them to a range between 0 and 1. This step is crucial for maintaining the stability of the LSTM training process and ensuring that the model can learn effectively from the sequences.

By using a horizon of 1 and a window of 7, this study aims to provide the LSTM model with relevant historical data to make accurate short-term forecasts of Ecopetrol’s stock price. The chosen parameters are designed to balance the need for capturing recent trends while avoiding the potential for overfitting to highly specific past patterns. This approach facilitates the model's ability to generalize and make reliable predictions based on recent historical data.

In summary, creating sequences for LSTM modeling involves defining the horizon and window parameters, organizing the data into input-output pairs, and structuring it in a format suitable for the LSTM network. This process ensures that the model can effectively learn from historical stock price data and make accurate predictions about future price movements.

2.3. Building the LSTM Model

The construction of the Long Short-Term Memory (LSTM) model is pivotal for predicting Ecopetrol's stock prices accurately. This section details the architecture and training process, ensuring the model is well-tuned and capable of handling sequential financial data effectively.

2.3.1. Model Architecture

1. LSTM Layer: The initial LSTM layer is designed to capture temporal dependencies and complex patterns in the time series data. This layer comprises a specified number of neurons, which are hyperparameters optimized during the model development process. Each neuron in the LSTM layer includes memory cells and gating mechanisms that manage the flow of information over time, enabling the model to learn and retain long-term dependencies. The choice of the number of neurons is critical, as it directly influences the model's capacity to generalize from historical data and predict future trends.
2. Dropout Layer: To combat overfitting and enhance the model's generalization ability, a dropout layer is included following the LSTM layer. The dropout rate, a hyperparameter, determines the fraction of neurons to be randomly deactivated during each training iteration. This technique prevents the model from becoming overly reliant on specific neurons, promoting the learning of more robust and generalized features from the training data.
3. Dense Output Layer: The final dense layer produces the model's predictions for the stock price. This layer is fully connected to the preceding LSTM and dropout layers, transforming the learned representations into a single output value. The dense layer's configuration ensures that the output is a continuous value, reflecting the forecasted stock price for the next timestep.

2.3.2. Training Process

The LSTM model is trained using the Adam optimization algorithm, which is well-regarded for its adaptive learning rate capabilities and efficiency in managing large datasets. Adam integrates the strengths of AdaGrad and RMSProp, providing robust gradient updates and facilitating rapid convergence during training. To measure the performance of the model, the Mean Squared Error (MSE) loss function is employed. MSE is selected for its ability to emphasize large deviations, making it particularly suitable for regression tasks where minimizing prediction errors is critical.

During the training phase, several hyperparameters are tuned to optimize the model's performance. The number of epochs is a crucial hyperparameter, determining how many times the entire training dataset is processed. Training continues until convergence is achieved, which is indicated by a plateau or minimal decrease in the loss function. Another important hyperparameter is the batch size, which specifies the number of samples processed before the model's parameters are updated. The choice of batch size impacts both the stability of training and memory usage. Additionally, the number of neurons in the LSTM layer is adjusted to enhance the model’s ability to capture complex temporal patterns in the data. Optimal values for these hyperparameters are identified through rigorous experimentation and cross-validation.

The training process is meticulously monitored to ensure that the model converges appropriately and to prevent overfitting. Techniques such as early stopping are utilized to terminate training when the validation loss no longer improves, thus balancing model performance with computational efficiency.

2.4. Model Evaluation

The evaluation phase is crucial for assessing the performance of the LSTM model, ensuring its predictive accuracy and reliability. To achieve this, several metrics are utilized. The Mean Absolute Error (MAE) measures the average magnitude of errors in predictions, offering an intuitive understanding of the model’s accuracy. MAE is robust to outliers, providing a clear and straightforward interpretation of prediction performance. Complementing MAE, the Mean Squared Error (MSE) calculates the average squared difference between predicted and actual values. MSE is particularly sensitive to large errors, making it effective for identifying significant deviations in predictions and highlighting areas where the model may be underperforming.

Additionally, the Root Mean Squared Error (RMSE), which is the square root of MSE, provides a measure of errors in the same units as the stock prices. This metric is valuable for understanding the average magnitude of errors and emphasizing larger deviations. The Mean Absolute Percentage Error (MAPE) expresses prediction errors as a percentage of the actual values. MAPE is particularly useful for evaluating the relative accuracy of predictions and allows for comparison of performance across different scales and datasets.

Lastly, the Mean Absolute Scaled Error (MASE) compares the model’s errors to those of a naive forecasting method, offering a scale-independent measure of accuracy. MASE is beneficial for evaluating how well the model performs relative to simple benchmark methods. Together, these evaluation metrics provide a comprehensive assessment of the LSTM model’s performance, ensuring that it delivers accurate and reliable stock price forecasts for Ecopetrol. Utilizing multiple metrics allows for a thorough evaluation of the model’s strengths and weaknesses, supporting rigorous scientific analysis and interpretation.

3. RESULTS AND DISCUSSION

3.1. Results

Table 1. Evaluation result table of LSTM model

Evaluation Method	Value
MAE	0.2509652
MSE	0.11678666
RMSE	0.34174064
MAPE	2.071206
MASE	1.125679

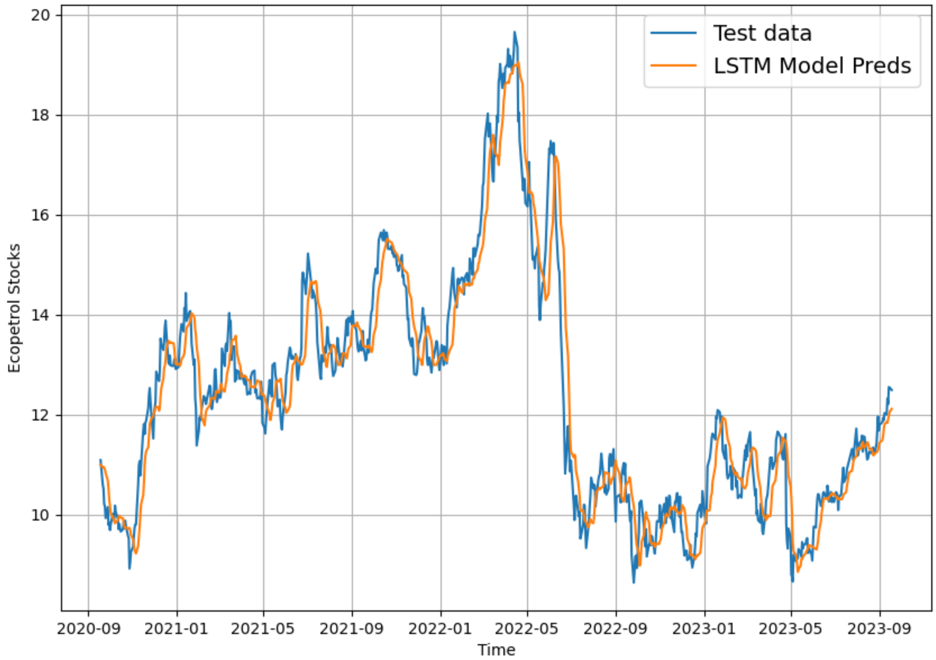

Figure 3. Comparison graph between test data and LSTM prediction model

Following the training of the Long Short-Term Memory (LSTM) model for predicting Ecopetrol stock prices, the model's performance was evaluated using several standard metrics. This section provides a detailed analysis of the results obtained and discusses their implications in the context of stock price forecasting.

The model achieved a Mean Absolute Error (MAE) of 0.2509652. MAE measures the average magnitude of the errors in a set of predictions, without considering their direction. It is calculated as the mean of the absolute differences between predicted and actual values. A lower MAE signifies a more accurate model. In this study, an MAE of approximately 0.25 stock price units indicates that the LSTM model's predictions are generally close to the actual prices, reflecting high prediction accuracy. MAE is particularly valuable because it provides a clear and intuitive understanding of model performance. Its robustness to outliers ensures that the reported error is representative of typical prediction accuracy, thus affirming the model's reliability in real-world applications.

The model's Mean Squared Error (MSE) is 0.11678666. MSE quantifies the average squared difference between predicted and actual values, emphasizing larger errors more than smaller ones. This metric is particularly useful in detecting significant deviations because it penalizes large errors more severely. The relatively low MSE in this study suggests that the LSTM model performs well overall, with few significant prediction errors. However, because MSE squares the errors, it can be heavily influenced by outliers or large deviations. This sensitivity is both a strength and a limitation; while it highlights significant discrepancies, it may obscure the model's performance if large errors disproportionately impact the metric.

The Root Mean Squared Error (RMSE) for the model is 0.34174064. RMSE is derived from MSE by taking the square root, which brings the error measure back to the original units of the data, in this case, stock price units. RMSE thus provides a measure of the average magnitude of prediction errors that is directly comparable to the scale of the actual data. The relatively low RMSE indicates that the LSTM model effectively predicts stock prices with minimal average error. RMSE is a more interpretable metric compared to MSE because it reflects the error magnitude in the same units as the data, facilitating a more straightforward assessment of the model's predictive accuracy.

The Mean Absolute Percentage Error (MAPE) of the model is 2.071206. MAPE expresses the prediction error as a percentage of the actual values, offering a relative measure of accuracy. This metric is advantageous for understanding the error in the context of the magnitude of the actual values, making it easier to compare performance across different scales and datasets. The low MAPE value of approximately 2% indicates that the LSTM model's predictions are highly accurate relative to the actual stock prices. This high level of accuracy reflects the model’s robustness across varying market conditions and its effectiveness in delivering precise forecasts.

The Mean Absolute Scaled Error (MASE) for the model is 1.125679. MASE compares the MAE of the LSTM model to that of a naive forecasting method, providing a scale-independent measure of forecasting accuracy. A MASE value greater than 1 indicates that the LSTM model performs slightly worse than the naive model, which uses a simple heuristic to make predictions. Although the MASE value suggests that the LSTM model does not outperform the naive model significantly, it is important to note that this metric assesses relative performance rather than absolute accuracy. The MASE value highlights areas where the model may be improved, particularly in handling data with seasonal or trending components more effectively.

3.2. Discussion

The evaluation results demonstrate that the LSTM model performs well in predicting Ecopetrol's stock prices, as evidenced by the low MAE, MSE, and RMSE values. These metrics collectively indicate that the model provides accurate and reliable predictions, with minimal errors. The low MAPE further confirms the model's effectiveness, reflecting its high relative accuracy. Despite the generally positive performance, the higher MASE value suggests that there is room for improvement, especially in terms of outperforming naive forecasting methods.

The results underscore the LSTM model's ability to handle complex temporal patterns inherent in stock price data. However, the higher MASE value indicates potential limitations, particularly in addressing seasonal variations or trends. Future research could explore enhancements to the model, such as incorporating additional features, adjusting model parameters, or employing alternative architectures to improve performance further. Overall, the LSTM model shows considerable promise for stock price forecasting, providing valuable insights for investors and decision-makers while highlighting areas for future development.

4. CONCLUSION

The evaluation of the Long Short-Term Memory (LSTM) model for predicting Ecopetrol stock prices reveals strong performance across several metrics. The model achieved low values in Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), indicating high accuracy in its predictions. The small MAE and MAPE suggest minimal prediction errors, both in absolute terms and relative to the actual stock prices, while the low RMSE confirms that the errors are well within acceptable ranges. Although the Mean Absolute Scaled Error (MASE) is slightly above 1, showing that the LSTM's performance is comparable to a naive forecasting method, the model still proves to be a reliable tool for stock price prediction. Overall, the LSTM model effectively handles sequential data and captures complex patterns, demonstrating its value in financial forecasting. Further improvements could be made by optimizing the model and incorporating additional relevant features.

REFERENCES

[1] P. Khuwaja, S. A. Khowaja, I. Khoso, and I. A. Lashari, "Prediction of stock movement using phase space reconstruction and extreme learning machines," *J. Exp. Theor. Artif. Intell.*, vol. 32, no. 1, pp. 59–79, 2020.

[2] A. Doshi, A. Issa, P. Sachdeva, S. Rafati, and S. Rakshit, "Deep Stock Predictions," *Research Papers in Economics*. 2020.

[3] A. Kaur, M. Bhadauria, and M. Monika, "Heuristic approach for forecasting stock price using LSTM and technical indicators," 2022, pp. 1–6. doi: 10.1109/AIST55798.2022.10064776.

[4] B. J. Galli, "Enhanced Stock Market Prediction Using Hybrid LSTM Ensemble," 2022, pp. 49–61. doi: 10.1007/978-981-16-9416-5_4.

[5] I. Kalbandi, A. Jare, O. Kale, H. Borole, and S. Navsare, "Stock Market Prediction Using LSTM," pp. 123–128, 2021, doi: 10.48175/IJARSCT-877.

[6] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon, and K. P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," in *2017 international conference on advances in computing, communications and informatics (icacci)*, 2017, pp. 1643–1647.

[7] J. Yin and Z. Chuanrui, "Stock price prediction method and system based on improved LSTM." 2019.

[8] L. Brown, "Predicting Stocks with LSTM-based DRNN and GAN." 2022. doi: 10.31979/etd.fjsz-y926.

[9] M. Hiransha, E. A. Gopalakrishnan, V. K. Menon, and K. P. Soman, "NSE Stock Market Prediction Using Deep-Learning Models," *Procedia Comput. Sci.*, vol. 132, no. Iccids, pp. 1351–1362, 2018, doi: 10.1016/j.procs.2018.05.050.

[10] V. Deshpande, "Implementation of Long Short-Term Memory (LSTM) Networks for Stock Price Prediction," *Res. J. Comput. Syst. Eng.*, vol. 4, no. 2, pp. 60–72, 2023.

[11] Q. Wu, J. Lu, and H. Zhang, "An improved LSTM stock price prediction model based on multiple basis function expansion and two-layer Bagging algorithm," *Highlights Business, Econ. Manag.*, vol. 10, pp. 227–235, 2023, doi: 10.54097/hbem.v10i.8043.

[12] X. Liu, "A modified LSTM neural network model for stock price prediction based on auxiliary information and model average," *Highlights Sci. Eng. Technol.*, vol. 22, pp. 92–100, 2022, doi: 10.54097/hset.v22i.3296.

[13] A. Lawi, H. Mesra, and S. B. H. Amir, "Implementation of Long Short-Term Memory and Gated Recurrent Units on grouped time-series data to predict stock prices accurately," *J. Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00597-0.

[14] A. W. Li and G. S. Bastos, "Stock Market Forecasting Using Deep Learning and Technical Analysis: A Systematic Review," *IEEE Access*, vol. 8, pp. 185232–185242, 2020, doi: 10.1109/ACCESS.2020.3030226.

[15] F. Kamalov, "Forecasting significant stock price changes using neural networks," *Neural Comput. Appl.*, vol. 32, no. 23, pp. 17655–17667, 2020, doi: 10.1007/S00521-020-04942-3.

[16] G. Li, M. Xiao, and Y. Guo, "Application of Deep Learning in Stock Market Valuation Index Forecasting," 2019. doi: 10.1109/ICSESS47205.2019.9040833.

[17] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "A novel multi-source information-fusion predictive framework based on deep neural networks for accuracy enhancement in stock market prediction," *J. Big Data*, vol. 8, no. 1, pp. 1–28, 2021, doi: 10.1186/S40537-020-00400-Y.

[18] J. Zhao, D. Zeng, S. Liang, H. Kang, and Q. Liu, "Prediction model for stock price trend based on recurrent neural network," *J. Ambient Intell. Humaniz. Comput.*, vol. 12, no. 1, pp. 745–753, 2021, doi: 10.1007/S12652-020-02057-0.

[19] P. Mehta, S. Pandya, and K. Kotecha, "Harvesting social media sentiment analysis to enhance stock market prediction using deep learning," *PeerJ*, vol. 7, 2021, doi: 10.7717/PEERJ-CS.476.

[20] S. Garcia-Vega, X.-J. Zeng, and J. A. Keane, "Stock returns prediction using kernel adaptive filtering within a stock market interdependence approach," *Expert Syst. Appl.*, vol. 160, 2020, doi: 10.1016/J.ESWA.2020.113668.

[21] A. Peivandizadeh *et al.*, "Stock market prediction with transductive long short-term memory and social media sentiment analysis," *IEEE Access*, doi: 10.1109/access.2024.3399548.

[22] X. Zhang and Y. Tan, "Deep Stock Ranker: A LSTM Neural Network Model for Stock Selection," 2018, pp. 614–623. doi: 10.1007/978-3-319-93803-5_58.

[23] Z. Ying *et al.*, "Predicting stock market trends with self-supervised learning," *Neurocomputing*, 2023, doi: 10.1016/j.neucom.2023.127033.

[24] Z. An, Y. Wu, F. Hao, Y. Chen, and X. He, "A novel hierarchical feature selection with local shuffling and models reweighting for stock price forecasting," *Expert Syst. Appl.*, 2024, doi: 10.1016/j.eswa.2024.123482.