International Journal of Informatics and Computation (IJICOM)
Vol. 7, No. 2, 2025
ISSN: 2685-8711, E-ISSN: 2714-5263 DOI: 10.35842/ijicom

Deep Learning Approach to Pharmaceutical Stock Forecasting using LSTM Architecture

Indra Irawan[1], M.Hizbul Wathan[2] ,Better Swengky[3] ,Ardi Ramadani[4]

Abstract

Stock price forecasting in the pharmaceutical sector is challenging due to high volatility and nonlinear temporal patterns. Conventional statistical models often fail to capture long-term dependencies in financial time series. This study proposes an optimized Long Short-Term Memory (LSTM) architecture to forecast the closing stock prices of PT Kalbe Farma Tbk (KLBF.JK). Historical daily stock data from 2020 to 2024 were collected from Yahoo Finance and preprocessed using Min–Max normalization. In this study, we evaluate several LSTMs by varying epochs and batch sizes to identify the optimal model. Experimental results show that the proposed LSTM model achieved the lowest Root Mean Square Error (RMSE) of 25.1406 using 100 epochs and a batch size of 5. The configured LSTM demonstrates superior predictive performance in capturing stock price dynamics. The findings confirm that the optimized LSTM architecture is effective for pharmaceutical stock forecasting and can support data-driven investment decision-making.

Keywords:
Prediction, Stock, LSTM, Kalbe Farma, Time Series

1. Introduction

Stock market forecasting is a crucial area in financial data analytics because it helps investors and analysts anticipate market fluctuations and optimize investment strategies. In recent years, stock prices have exhibited increasing volatility due to various internal and external factors, including company performance, macroeconomic conditions, global economic uncertainty, and investor sentiment [1], [3]. This high volatility introduces significant uncertainty and risk, thereby motivating the need for accurate and reliable prediction models that can support effective decision-making in financial markets.

Traditional forecasting methods such as Multiple Linear Regression (MLR) and Ordinary Least Squares (OLS) have long been utilized to estimate stock price trends due to their mathematical simplicity and ease of interpretation [2], [5], [7]. However, these models generally assume linear relationships among variables and therefore fail to capture the nonlinear and time-dependent patterns commonly observed in financial time-series data [4], [8]. As a result, their predictive performance tends to deteriorate when applied to highly volatile and complex stock market environments. This limitation has encouraged researchers to explore more advanced approaches.

Corresponding Author: Indra Irawan (indraa@polman-babel.ac.id)
1 Indra Irawan, Politeknik Manufaktur Negeri Bangka Belitung, indraa@polman-babel.ac.id
2 M.Hizbul Wathan, Politeknik Manufaktur Negeri Bangka Belitung, mhizbul@polman-babel.ac.id
3. Better Swengky, Politeknik Manufaktur Negeri Bangka Belitung, better@polman-babel.ac.id
4. Ardi Ramadani, Politeknik Manufaktur Negeri Bangka Belitung, ardiramadani@gmail.com

Consequently, recent studies have increasingly focused on machine learning (ML) and deep learning (DL) techniques that are capable of modeling complex nonlinear relationships and temporal dependencies in financial data [9], [10], [13]. Among these approaches, deep learning models have demonstrated superior performance in learning patterns from large-scale and sequential datasets. In particular, comparative studies have shown that LSTM and GRU networks outperform conventional regression-based models when dealing with time-dependent financial data due to their ability to retain historical information over long sequences [14], [17].

Among various deep learning architectures, the LSTM network, a variant of the RNN, has demonstrated strong performance in sequential data prediction tasks [6], [12], [14]. LSTM incorporates memory cells and gating mechanisms that enable the model to retain long-term dependencies while mitigating the vanishing gradient problem commonly encountered in traditional RNNs [17]. Previous research comparing regression-based models with deep learning approaches has consistently shown that LSTM achieves superior predictive accuracy in stock price forecasting by effectively capturing temporal dynamics and nonlinear relationships [11], [13], [19].

Unlike traditional regression-based and feed-forward neural network models, LSTM is specifically designed to handle sequential data with long-term dependencies. Stock price movements are strongly influenced by historical patterns over extended periods, making memory-based architectures more suitable for financial time-series forecasting. The gating mechanisms in LSTM allow the model to selectively retain relevant information and discard irrelevant noise, enabling it to capture complex temporal relationships and nonlinear dynamics that conventional machine learning models fail to model effectively. These characteristics make LSTM particularly appropriate for modeling volatile stock price behavior.

This study aims to apply and optimize an LSTM-based deep learning model to forecast the closing stock prices of PT Kalbe Farma Tbk (KLBF.JK), one of Indonesia's largest pharmaceutical companies. The pharmaceutical industry remains a promising investment sector due to its consistent market demand and continuous innovation in healthcare products [18]. However, stock prices in this sector are still subject to significant volatility driven by global market conditions, regulatory changes, supply chain variations, and production cost fluctuations. Using historical daily stock data from 2020 to 2024 obtained from Yahoo Finance, this research seeks to evaluate the effectiveness of an optimized LSTM model in learning temporal dependencies and predicting stock price trends in the pharmaceutical sector.

The novelty of this research lies in its focused and systematic approach to pharmaceutical stock price forecasting in the Indonesian market, using PT Kalbe Farma Tbk (KLBF.JK), a sector that remains underexplored compared to technology and banking stocks in existing literature. This study advances prior work by moving beyond fixed LSTM configurations and implementing a structured hyperparameter tuning strategy to examine the effects of varying epochs and batch sizes on predictive performance. In addition, it provides an explicit mathematical formulation of the LSTM architecture, thereby enhancing model transparency, interpretability, and reproducibility in financial time-series analysis. Finally, the study employs a comprehensive evaluation framework that integrates RMSE-based quantitative assessment with visual trend analysis, demonstrating the capability of LSTM to effectively model nonlinear patterns and temporal dependencies in stock price movements.

2. Related Works

Recent developments in stock price prediction have shown a growing shift from traditional regression models to advanced deep learning architectures. Previous research has widely explored Multiple Linear Regression (MLR) and Ordinary Least Squares (OLS) for financial forecasting due to their interpretability and efficiency in modeling straightforward linear relationships [1], [5]. For instance, Wang et al. [7] applied MLR to forecast the price trends of several major technology companies, achieving moderate predictive accuracy. Similarly, Rusu et al. [4] compared various regression models, including Ridge and LASSO, in predicting Apple Inc.'s stock prices and concluded that while linear regression performs adequately for short-term predictions, it struggles to capture nonlinear trends inherent in financial data.

The limitations of linear models have driven researchers toward machine learning (ML) and deep learning (DL) approaches capable of capturing complex nonlinear relationships. Jin and Yi [3] compared regression models and ML scenarios, showing that tree-based algorithms such as XGBoost and Random Forest yielded better generalization in highly volatile datasets. Li [13] and Zhou [8] emphasized that integrating regression with nonlinear ML techniques significantly improves prediction stability. Recent studies have also proposed LSTM-based stock prediction models and reported improved forecasting accuracy by effectively modeling temporal dependencies in financial time-series data [20]. Other comparative studies, such as those by Yin [14] and Xia [17], demonstrated that LSTM and GRU networks outperform regression-based models when dealing with time-dependent data due to their ability to retain temporal information over long sequences.

Specifically, Zhou [12] and Bhatta et al. [9] highlighted that LSTM is particularly effective in modeling stock price data containing long-term dependencies and irregular fluctuations. Their experiments on companies such as Netflix and General Motors showed that LSTM could capture complex sequential patterns, resulting in lower RMSE compared to regression or feed-forward neural networks. More recent research has explored hybrid deep learning architectures, such as Conv1D–LSTM, for stock price forecasting in the Indonesian market and demonstrated enhanced predictive performance compared to standalone models [21]. Similarly, Rusu et al. [19] conducted a comparative analysis of several models, including SVR, Polynomial Regression, and LSTM, and found that LSTM consistently provided better forecasting accuracy in datasets with nonlinear characteristics.

From a pharmaceutical industry perspective, Xia [6] compared Linear Regression, Recurrent Neural Networks (RNN), and LSTM for predicting Johnson & Johnson's stock price, concluding that LSTM achieved the most stable performance when tested across multiple time intervals. Their study confirms that deep learning methods provide superior adaptability for industries like pharmaceuticals, where stock behavior is influenced by both internal production cycles and external health-related factors.

In the Indonesian context, previous studies primarily focused on regression-based approaches for predicting KLBF's stock prices [3]. Other studies have investigated stock price prediction using technical indicator-based approaches, such as the Moving Average Convergence Divergence (MACD) method, showing that feature engineering and domain-specific indicators can influence forecasting performance [22]. These studies demonstrated the model's potential but revealed limited capability in handling sequential dependencies within time series data. Therefore, this research extends prior works by implementing and optimizing an LSTM deep learning model to forecast PT Kalbe Farma Tbk (KLBF.JK) stock prices, addressing the nonlinear and temporal complexities that earlier regression-based models could not fully capture.

3. Proposed Method

This research applies an LSTM deep learning model to predict the stock prices of PT Kalbe Farma Tbk (KLBF.JK). The proposed method consists of four primary stages: data collection, preprocessing, model development, and model evaluation. The methodological flow is shown in Fig. 1.

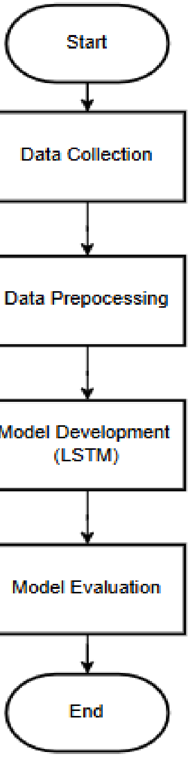

Figure 1. The methodological flow

3.1 Data Collection

The dataset was obtained from the Yahoo Finance website, containing daily historical stock data of PT Kalbe Farma Tbk (ticker symbol: KLBF.JK) from January 1, 2020, to December 30, 2024. The collected attributes include the opening price, high, low, closing price, adjusted close, and trading volume. For this study, the closing price (Close) was selected as the primary target variable since it represents the actual stock price at the end of each trading day and is widely used in financial forecasting models.

3.2 Data Preprocessing

Data preprocessing was performed to ensure data consistency and to improve model performance. Data preprocessing was conducted to ensure data consistency and enhance the predictive performance of the model. Missing values were first examined and removed to preserve the continuity of the time-series sequence. Subsequently, Min–Max normalization was applied to rescale all values into the range [0, 1], as defined in Equation (1).

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

(1)

This normalization process enables stable neural network training by preventing domination from features with larger numerical scales and improving convergence behavior. After normalization, the dataset was split sequentially into 80% training data and 20% testing data to maintain the temporal order inherent in stock price movements. Finally, the data were reshaped into a three-dimensional array format compatible with the LSTM input structure, where each sample represents a time window of historical stock prices, allowing the model to effectively learn temporal dependencies.

3.3 Model Development

LSTM is designed to overcome the vanishing gradient problem when processing long-term sequential data [17]. The model consists of three main gates, including the Forget Gate, Input Gate, and Output Gate, that regulate the flow of information. The LSTM model operates based on the following key components and formulas:

1. **Forget Gate:** Determines which information from the previous state should be forgotten. Table 1 summarizes the mathematics of the **forget gate** in LSTM.

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f \quad (2)$$

Table 1. The Mathematical Notation of the Forget Gate

Notation	Description
f_i	$forget\ gate\ output$
σ	$Sigmoid\ activation\ fuction$
W_f	$Weights\ for\ the\ forget\ gate$
h_{t-1}	$Previous\ hidden\ state$
x_i	$input\ at\ the\ current\ timestep$
b_f	$Bias\ for\ the\ forget\ gate$

2. **Input Gate:** Decides which new information to store in the cell state.

$$\begin{gathered} i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \\ \sim C\ _t = tanh\ (W_c.[h_{t-1}, x_t] + b_c) \\ C_t = f_t.C_{t-1} + i_t.{\sim}C_t \end{gathered} \quad (3)$$

Table 2. The Mathematical Notation of the Input Gate

Notation	Description
i_t	$input\ gate\ output$
$\sim C\ _t$	$Candidate\ cell\ state$
C_t	$Update\ cell\ state$
W_i, W_c	$Weights\ for\ input\ gate\ and\ candidate\ state$
b_i, b_c	$Bias\ terms$

3. **Output Gate:** Determines the next hidden state based on the updated cell memory.
4.

$$\begin{gathered} o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \\ h_t = o_t \times \tanh(C_t) \end{gathered} \quad (4)$$

Table 3. The Mathematical Notation of the Output Gate

Notation	Description
O_t	$Output\ gate\ result$
h_t	$Current\ hidden\ state\ (LSTM\ output)$
W_o	$Weights\ for\ output\ gate$
b_o	$Bias\ term$
C_t	$Current\ cell\ state$

In this research, the LSTM mechanism was implemented to model the sequential behavior of PT Kalbe Farma Tbk (KLBF.JK) stock prices. The network learns to retain relevant long-term dependencies while filtering out noise, enabling more accurate forecasting of stock trends over time.

4.4 Model Evaluation

The model's performance was evaluated using the Root Mean Square Error (RMSE) metric as expressed in Equation (4):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

Where y_i represents the actual stock price, and $\hat{y}_i$ denotes the predicted value. A lower RMSE value indicates better predictive accuracy. Additionally, the predicted results were visualized alongside actual prices to observe how closely the model followed the stock's temporal trend.

4. Experimental Setup

4.1 Experimental Environment

The experiment was conducted to evaluate the predictive performance of the LSTM model for forecasting the closing stock prices of PT Kalbe Farma Tbk (KLBF.JK). The implementation was performed using the Python programming language with the TensorFlow and Keras libraries. The experiment ran on a computer with Windows 1an 1, Intel Core i7 processor, and 16 GB of RAM.

4.2 Dataset Description

The dataset was collected from the Yahoo Finance platform, consisting of daily historical stock data for PT Kalbe Farma Tbk (KLBF.JK) spanning from January 1, 2020, to December 30, 2024. The attributes include: Open, High, Low, Close, Adjusted Close, and Volume. The closing price (Close) was selected as the target variable since it reflects the most representative daily market value and is commonly used for financial forecasting models.

4.3 Parameter Configuration

During the training phase, the dataset was divided into 80% training data and 20% testing data. The LSTM model was configured with multiple layers and trained using the Adam optimizer with Mean Squared Error (MSE) as the loss function. To find the most effective combination of parameters, several hyperparameter tuning experiments were conducted with variations in epochs and batch sizes. The Root Mean Square Error (RMSE) was used as the main evaluation metric, as defined in Equation (2).

Table 4. LSTM Model Parameters

No	Parameter	Description / Value
1	Model Type	LSTM (Sequential Neural Network)
2	Optimizer	Adam
3	Loss Function	Mean Squared Error (MSE)
4	Activation Function	ReLU (hidden layers), Linear (output)
5	Normalization	Min–Max (0–1 range)
6	Evaluation Metric	Root Mean Square Error (RMSE)
7	Data Split	80% Training, 20% Testing
8	Epoch	100, 200, 300
9	Batch Sizes	5,10,15

5. Result and Analysis

5.1 Hyperparameter Tuning Results

The experimental results demonstrate that the LSTM model's performance is highly influenced by hyperparameter selection. As shown in Table 5, smaller batch sizes consistently produced lower RMSE values, indicating smoother gradient updates and better generalization.

Table 5. RMSE Results for Various Hyperparameter Combinations

Epochs	Batch Size	RMSE
100	**5**	**25.1406**
200	5	26.4602
300	5	27.0091
100	10	27.5447
200	10	28.5071
300	10	26.1276
100	15	25.5206
200	15	25.4391
300	15	25.4008

The optimal configuration was achieved with 100 epochs and a batch size of 5, resulting in the lowest RMSE value of 25.1406, which outperformed other tested combinations.

The forecasting accuracy achieved in this study is comparable to results reported in previous LSTM-based stock price forecasting studies. Several earlier works applying standard LSTM architectures reported RMSE values within a similar range when modeling stock price movements in different sectors. In this research, the optimized LSTM configuration achieved an RMSE of 25.1406, indicating that systematic hyperparameter tuning contributes significantly to improving prediction performance. Unlike prior studies that employed fixed model configurations, this study demonstrates that appropriate selection of epochs and batch sizes enhances model stability and generalization, particularly for volatile pharmaceutical stock data.

5.2 Visualization of Predicted vs Actual Prices

Figure 2 illustrates the comparison between actual and predicted stock prices, where the LSTM model closely follows the overall price trend of KLBF JK. Although minor deviations appear during periods of high volatility, the model successfully captures the general movement and turning points of the stock price. Such deviations are common in financial time-series forecasting and indicate the inherent uncertainty of stock market dynamics.

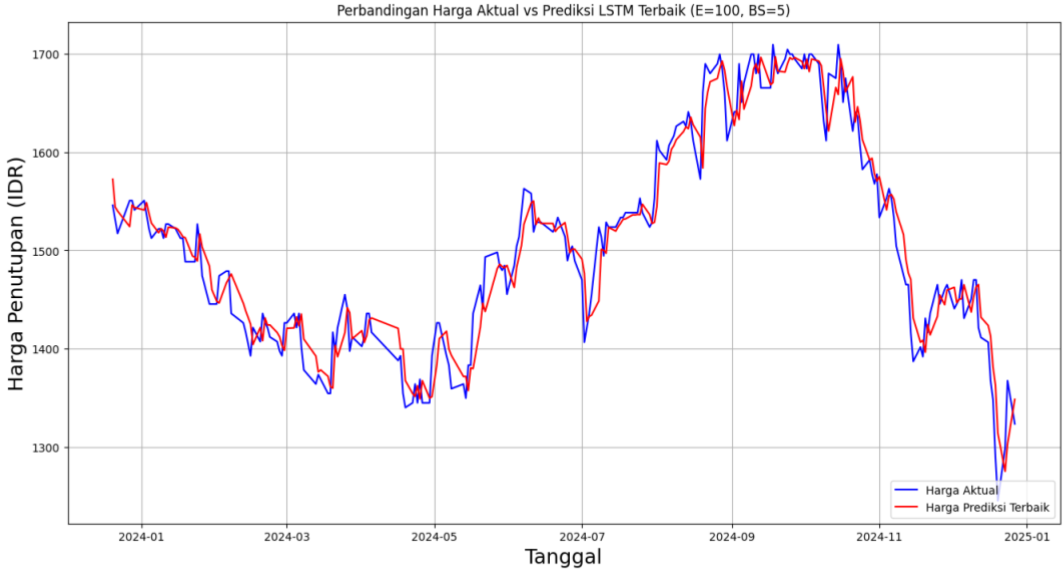

Figure 2. Actual vs Predicted Stock Price of KLBF.JK using LSTM

5.3 Analysis of Final Predictions

Table 6 highlights the differences between actual and predicted closing prices for the last ten trading days, providing insight into the model's predictive accuracy under recent market conditions.

Table 6. Comparison of Actual and Predicted Prices (Last 10 Records)

Date	Actual	Predicted	Difference
2024-12-12	1421.21	1465.27	44.06
2024-12-13	1411.44	1432.01	20.57
2024-12-16	1406.55	1423.59	17.03
2024-12-17	1367.48	1412.99	45.51
2024-12-18	1347.95	1381.20	33.25
2024-12-19	1289.34	1363.04	73.70
2024-12-20	1245.39	1314.00	68.61
2024-12-23	1299.11	1275.23	23.88
2024-12-24	1367.48	1302.99	64.50
2024-12-27	1323.53	1348.44	24.91

The comparison demonstrates that the proposed LSTM model is able to capture the general movement of stock prices with reasonable accuracy. Larger prediction differences occur during periods of high market volatility, suggesting that model performance could be further improved by incorporating external indicators such as trading volume or market sentiment.

6. Conclusion

This study proposed an optimized LSTM model to forecast the closing stock prices of PT Kalbe Farma Tbk (KLBF.JK) using daily historical data from 2020 to 2024. Experimental results demonstrated that the LSTM model effectively captured nonlinear and temporal patterns in stock price movements, achieving the best RMSE value of 25.1406 with 100 epochs and a batch size of 5. These results confirm that the optimized LSTM architecture provides superior predictive capability compared to traditional regression-based models, particularly for volatile financial time-series data in the pharmaceutical sector.

Future research can extend this work by integrating external factors such as macroeconomic indicators, trading volume, and sentiment analysis from financial news or social media. Additionally, hybrid deep learning architectures such as CNN–LSTM or attention-based models may further improve forecasting accuracy and robustness in highly volatile market conditions.

References

[1] H. A. K. Ihlayyel, N. M. Sharef, M. Z. A. Nazri, and A. A. Bakar, “An Enhanced Feature Representation Based on Linear Regression Model for Stock Market Prediction,” Intelligent Data Analysis, vol. 22, no. 1, pp. 45–76, Jan. 2018, doi: 10.3233/IDA-163316.

[2] A. Hayes, “Multiple Linear Regression (MLR): Definition, Formula, and Example,” Investopedia, Oct. 15, 2023. [Online]. Available: https://www.investopedia.com/terms/m/mlr.asp

[3] X. Jin and C. Yi, “The Comparison of Stock Price Prediction Based on Linear Regression Model and Machine Learning Scenarios,” in Proc. 2022 Int. Conf. Bigdata Blockchain Econ. Manag. (ICBBEM 2022), Atlantis Press, Dec. 2022, pp. 837–842. doi: 10.2991/978-94-6463-030-5_82.

[4] M. Rusu, A. Popescu, and D. Ionescu, “Performance Comparison of Various Regression Models in Predicting Apple Inc. Stock Prices,” J. Bus. Res., vol. 14, no. 1, pp. 112–125, 2024.

[5] Q. Wang, C. Xu, and T. Zhou, “Stock Price Prediction Based on Multiple Linear Regression,” BCP Business & Management CMAM, vol. 36, pp. 48–54, Jan. 2023, doi: 10.54691/bcpbm.v36i.3384.

[6] [6] L. Xia, “Comparison of Linear Regression, RNN, and LSTM Models in Predicting Johnson & Johnson Stock Prices,” J. Financial Engineering, vol. 15, no. 2, pp. 89–102, 2023.

[7] Q. Wang, C. Xu, and T. Zhou, “Using Multiple Linear Regression Model to Predict the Trend of Stock Prices,” BCP Business & Management, vol. 38, pp. 142–149, 2023, doi: 10.54691/bcpbm.v38i.3384.

[8] Y. Zhou, “Stock Forecasting Based on Linear Regression Model and Nonlinear Machine Learning Regression Model,” in Advances in Economics, Management and Political Sciences, EWA Publishing, Jan. 2024, pp. 7–13. doi: 10.54254/2754-1169/57/20230364.

[9] S. Bhatta, R. Sharma, and P. Gupta, “Evaluating the Accuracy of Machine Learning Models in Predicting Stock Prices,” Int. J. Mach. Learn. Appl., vol. 8, no. 1, pp. 34–50, 2023.

[10] Y. Zhou, “Research on the Prediction of Stock Price Based on Regression and Deep Learning Models—Taking Netflix as an Example,” Adv. Educ., Manag. & Political Sci., vol. 8119, pp. 1–6, 2024. [Online]. Available: https://www.ewadirect.com/proceedings/aemps/article/view/8119

[11] L. Jin and H. Yi, “Comparative Analysis of Prediction Effects of Linear Regression and Machine Learning Algorithms on Stock Price: A Case Study of LONGi,” in Proc. 2022 6th

Int. Conf. Bus. Inf. Manag. (ICBIM), 2022, pp. 190–193. doi: 10.2991/978-94-6463-011-5_44.
[12] Z. Zhou, “Comparative Study of Linear Regression, LSTM, and GRU Models in Predicting Netflix Stock Prices,” J. Comput. Finance, vol. 12, no. 3, pp. 201–215, 2024.
[13] J. Li, “Ordinary Least Squares (OLS), Ridge, and XGBoost Models for Predicting Stock Prices of Listed Companies,” J. Financial Analysis, vol. 10, no. 3, pp. 78–91, 2023.
[14] Y. Yin, “Comparison of KNN, Linear Regression, and LSTM Models in Predicting GM Stock Prices from 2013 to 2023,” J. Financial Data Sci., vol. 7, no. 2, pp. 56–70, 2024.
[15] Investopedia, “Multiple Linear Regression (MLR) and its Application in Stock Price Prediction,” Oct. 15, 2023. [Online]. Available: https://www.investopedia.com/terms/m/multiple-linear-regression.asp
[16] A. Bhatta, P. Poudyal, D. K. Maharja, and A. Thapa, “Assessing Machine Learning’s Accuracy in Stock Price Prediction,” Int. J. Comput. (IJC), vol. 49, no. 1, pp. 46–63, Sep. 2023. [Online]. Available: https://ijcjournal.org/index.php/InternationalJournalOfComputer/index
[17] B. Xia, “Stock Price Prediction Based on Linear Regression, RNN, LSTM,” BCP Business & Management, vol. 38, pp. 355–362, Mar. 2023, doi: 10.54691/bcpbm.v38i.3715.
[18] Y. Li, “Stock Price Prediction Based on Multiple Regression Models,” in Proc. 2023 Int. Conf. Comput., Mach. Learn., & Artif. Intell. (CMLAI 2023), Highlights in Sci., Eng., & Technol., Apr. 2023, pp. 657–662.
[19] Ș. Rusu, M. I. Boloș, and M. Leordeanu, “Comparative Analysis of Regression Models for Stock Price Prediction: Linear, Support Vector, Polynomial, and LASSO,” J. Financial Studies, vol. 9, no. 17, pp. 143–156, Nov. 2024, doi: 10.55654/JFS.2024.9.17.09.
[20] Diqi, “StockTM: Accurate Stock Price Prediction Model Using Long Short-Term Memory,” International Journal of Informatics and Computation, vol. 6, no. 2, pp. 85–94, 2022.
[21] A. Pratama, R. Nugroho, and S. Raharjo, “Stock Price Prediction in Indonesia’s Mining Sector Using a Hybrid Conv1D–LSTM Model,” International Journal of Informatics and Computation, vol. 8, no. 1, pp. 45–54, 2024.
[22] A. Hamzah and D. Winardi, “Effective Stock Prediction Model Using MACD Method,” International Journal of Informatics and Computation, vol. 6, no. 2, pp. 101–110, 2022.