

Multivariate Approach of LSTM Model for Forecasting Tjiwi Kimia Tbk Stock Price

Muhammad Zaki Naufal¹, Sulistyo Dwi Sancoko²

^{1,2}Department of Informatics, University of Technology Yogyakarta, Indonesia

Article Info

Article history:

Received 08 30, 2026

Revised 09 27, 2026

Accepted 10 27, 2026

Keywords:

Prediction,
Stocks,
Price,
Long Short-Term Memory,
Multivariate

ABSTRACT

Stocks are a type of long-term financial instrument found in the capital market. However, stock price movements are volatile and influenced by various factors such as economic conditions, market sentiment, government policies, and global dynamics requiring market participants or investors to conduct analysis before entering the market. Analysis generally falls into two categories: fundamental and technical; yet, even when employing these techniques, investors may still face the risk of failure. To address this issue, this study designs and develops a system to predict Indonesian stock closing prices using the Long Short-Term Memory (LSTM) model, aiming to provide investors with insights for their investment planning. The study utilizes OHLC (Open, High, Low, Close) data for the target stocks and the closing price of the Jakarta Composite Index (JCI) as supplementary data. Evaluation is conducted using RMSE and MAE metrics. The results demonstrate that the prediction system is capable of forecasting closing prices with a Root Mean Square Error (RMSE) of 280 and a Mean Absolute Error (MAE) of 210.55. This model is expected to assist stock market participants and investors in making investment decisions.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Muhammad Zaki Naufal
Department of Informatics
University of Technology Yogyakarta
Yogyakarta, Indonesia
Email: zakinnaufal11@gmail.com
© The Author(s) 2026

1. Introduction

Advancements in information technology and artificial intelligence have progressed rapidly in recent years. Digital transformation has reshaped various sectors, including the financial industry and the stock market. Currently, the capital market plays a vital role in Indonesia's economy [1]. Technological progress and the abundance of available information have paved the way for the development of more sophisticated and efficient analytical methods. In the context of the stock market, the use of artificial intelligence enables the analysis of complex historical stock price patterns, thereby assisting investors in making investment decisions. Models such as Long Short-Term Memory (LSTM) have emerged as widely used algorithms for processing time-series data [2].

The financial world has undergone a major transformation driven by technological advancements, particularly in how investors and traders conduct analysis and make decisions [3]. Stocks are long-term financial instruments traded in capital markets; they are securities issued by companies to represent ownership [1]. Investing in stocks offers two primary types of returns to holders dividends and capital gains. Dividends represent a distribution of corporate profits determined at the General Meeting of Shareholders

(GMS), whereas capital gains are profits realized by investors from selling assets at a higher price in the secondary market. However, alongside this profit potential, stocks also carry significant risks, such as capital losses (losses resulting from a decline in price). Pabrik Kertas Tjiwi Kimia Tbk, traded under the ticker symbol TKIM, is one such company and a leading paper manufacturer in Indonesia. However, stock price movements are volatile and influenced by various factors, such as macroeconomic conditions, market sentiment, government policies, and global dynamics. Consequently, stock price forecasting has become a compelling research topic. Stock price forecasting enables the prediction of future prices based on available historical data, thereby providing insights to support decision-making.

One of the primary challenges in stock investment is accurate price prediction, given that stock prices can fall at any time due to market fluctuations, shifts in investor sentiment, or external factors such as global economic conditions and financial news that influence market sentiment [4]. Without adequate analysis, investment decision-making becomes merely speculative, potentially increasing the risk of financial loss. Analysis is generally categorized into two types: fundamental analysis and technical analysis [5]. Fundamental analysis evaluates a company's financial performance to determine the stock's intrinsic value, whereas technical analysis utilizes historical price and volume data to predict future patterns [6]. Given the non-linear, volatile, and complex nature of the data, various models often struggle to capture non-linear relationships and long-term dependencies within time-series data. Furthermore, low prediction accuracy can lead to errors in investment decision-making. Consequently, there is a need for a more adaptive approach capable of deeply learning data patterns to enhance prediction performance.

Various methods exist for processing time-series data and predicting movements in the Jakarta Composite Index (IHSG), ranging from traditional statistical approaches to more modern machine learning techniques [7]. Machine learning methods are widely employed for stock price prediction [8]. To address this task, the present study utilizes the Long Short-Term Memory (LSTM) model. The LSTM algorithm was selected due to its ability to handle sequential data while retaining long-term information from the dataset. Research [6] indicates that the LSTM model is more effective at modeling complex and dynamic stock price movement patterns. As a variant of the Recurrent Neural Network (RNN) designed to overcome the vanishing gradient problem, LSTM possesses a unique capability to capture long-term dependency patterns in time-series data. Long Short-Term Memory (LSTM) has emerged as one of the most widely adopted model architectures for time-series forecasting, owing to its ability to capture long-term dependencies while mitigating the vanishing gradient issue common in standard Recurrent Neural Networks (RNNs) [9]. Furthermore, the effectiveness of deep learning in supporting data-driven decision-making raises important questions regarding scalability and real-world application [10]. Through a gating mechanism comprising input, forget, and output gates, LSTM can selectively retain or discard information across long data sequences, making it highly effective for analyzing historical data trends such as stock price movements.

Most existing studies on stock price prediction rely on a univariate approach, in which only a single variable typically the closing price serves as the model input. Although computationally simple, this approach often fails to capture the complex interactions among various market factors influencing price movements, such as opening prices, trading volume, and other technical indicators. Previous research by [6] employed an LSTM model to predict the stock prices of BBCA and BBRI using a univariate approach; evaluation results showed that the model achieved a training RMSE of 37.037 and a testing RMSE of 80.128 for BBRI, while for BBCA, it yielded a training RMSE of 36.905 and a testing RMSE of 99.9. These results indicate that, while the LSTM model performed reasonably well during the training phase, the higher RMSE values during the testing phase suggest that the univariate approach still leaves room for improvement, particularly regarding generalization to unseen data.

Based on these findings, this study develops a standard LSTM model by expanding its input structure from a univariate to a multivariate configuration and incorporating additional historical variables believed to influence stock price fluctuations. This study utilizes Open, High, Low, and Close (OHLC) data from selected listed companies as primary features. These four variables comprehensively capture price movements within a single trading period. The "Open" price reflects the initial price established by market sentiment prior to the session's start "High" and "Low" indicate the range of price fluctuations (intra-period volatility) and "Close" represents the final agreed-upon price, which serves as the standard reference for valuation and the target variable for prediction. Compared to relying solely on closing prices, the use of OHLC data provides the model with additional information regarding the direction, range, and intensity of price movements, enabling the model to learn short-term patterns that would otherwise remain undetected when using a single variable. In addition to company-specific data, this study incorporates the closing price of the Jakarta Composite Index (JCI) as an external feature, acknowledging that stock price movements are influenced not only by internal company factors but also by overall market conditions.

It is anticipated that enriching the model with more relevant features will mitigate the performance gap between training and testing phases observed in previous univariate studies, thereby yielding a model

with superior generalization capabilities regarding new data. A primary objective of this research is to develop a deep learning-based stock price prediction model using the Long Short-Term Memory (LSTM) algorithm within a multivariate framework, capable of delivering higher accuracy and better generalization than conventional univariate models [11]. This phenomenon demonstrates that the choice between univariate and multivariate approaches depends not merely on the number of features, but on the extent to which those features genuinely contribute to the dynamics of stock price movements [12]. Through this approach, the study aims to produce a model that effectively leverages diverse historical data sources to identify complex and dynamic patterns in stock price movements. Consequently, this prediction model is expected to provide reliable predictive insights to investors, supporting more informed investment decision-making, reducing uncertainty, and optimizing profit potential in the stock market.

2. Research Method

This research is designed to predict stock prices using deep learning techniques, specifically the Long Short-Term Memory (LSTM) algorithm with a multivariate approach [13]. The design involves collecting historical stock data in Indonesia, processing the data by incorporating relevant variables beyond the closing price, training the model using multivariate input structures, and then evaluating the model's performance based on the resulting prediction accuracy.

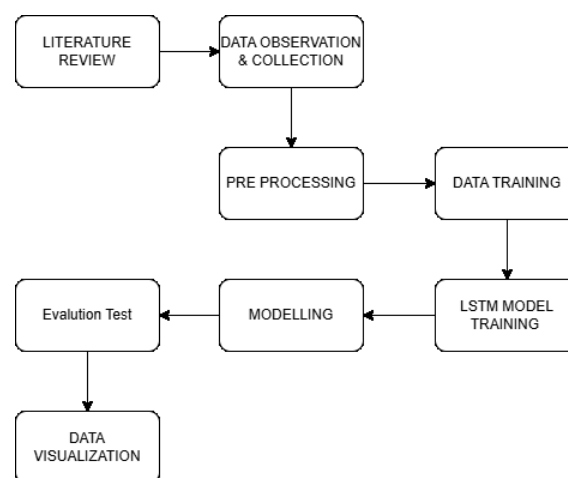


Figure 1. Research Methodology Flow

2.1 Data Collection

Data collection for this study was conducted online, utilizing secondary data sourced from the global financial information portal Investing.com. This platform was selected due to the availability, accuracy, and completeness of its publicly accessible historical stock movement database. The collected historical data utilized a daily timeframe, encompassing attributes such as opening, high, low, and closing prices, as well as trading volume. This daily time-series stock data served as input for the Multivariate Long Short-Term Memory (LSTM) model [13]. The use of daily data enabled the study to capture the dynamics of stock price fluctuations, which are sensitive to daily shifts in market sentiment [14].

The information extraction process was carried out manually by accessing the historical stock data records of the companies under study. Subsequently, the researcher applied a filter to precisely define the observation timeframe, aligning it with the period established within the study's scope. Data collection covered the period from January 1 2021 to January 1 2026, resulting in a time-series dataset comprising 1,205 rows of daily observations. The data were systematically classified into seven key attribute columns reflecting market trading activity, including: trading time parameters; closing stock price at the end of the session; opening stock price; the day's peak price; the day's lowest price; trading volume (the accumulated number of shares traded); and the percentage ratio of price fluctuation relative to the previous trading day.

2.1.1 Preprocessing

In the development of predictive models and time-series analysis for capital markets, data quality serves as the fundamental basis determining system accuracy and reliability. Analyzing the price movements of an individual stock such as PT Pabrik Kertas Tjiwi Kimia Tbk (TKIM) often cannot be done in isolation. Such stock price dynamics are strongly linked to macroeconomic market sentiment, which, in the context of the Indonesian capital market, is represented by the Jakarta Composite Index (JCI). Integrating these two information sources provides a richer data dimension, enabling algorithms to identify historical patterns and market correlations. However, raw financial data is characterized by high volatility and inconsistent

formatting, and is prone to anomalies; consequently, rigorous preprocessing is required before the data can be utilized for modeling.



Figure 2. Preprocessing Flow

The first stage in this information harmonization process is data merging. This step integrates historical records of TKIM stock and JCI movements into a cohesive matrix structure based on a time index. Timestamp alignment is crucial; discrepancies regarding exchange holidays, trading suspensions, or recording differences must be synchronized with precision. Errors during merging can lead to data shifts, which would mislead the algorithm when capturing the causal relationship between general market trends and specific stock reactions on the same day.

Once the dataset has been temporally integrated, the next essential step is data cleaning. Historical financial data often contains structural flaws such as missing values, duplicate rows, or extreme outliers resulting from system glitches [15] [16]. This cleaning process serves to diagnose and rectify such anomalies using techniques such as linear interpolation, forward-filling, or the removal of invalid observations [17]. These measures are implemented carefully to preserve the integrity of the original price trends, ensuring that the resulting model learns from valid market signals rather than computational noise.

The cleaned dataset is subsequently optimized through a data selection process. Not all attributes recorded in raw market data possess high predictive value or analytical relevance [18]. This process filters for key features such as closing (Close), opening (Open), and high/low prices (High/Low) which serve as primary indicators for interpreting market structure [19]. This dimensionality reduction, or the elimination of redundant features, not only significantly alleviates the computational load but also prevents the algorithm from overfitting by memorizing non-essential, secondary variables.

As a final data engineering step prior to consumption by algorithms particularly for machine learning and deep learning architectures that are highly sensitive to numerical values normalization is performed using the Min-Max Scaling method [20]. In their raw state, the value scales of TKIM stock prices and the JCI index points exhibit a significant numerical disparity. This scale imbalance can trigger computational bias, wherein variables with larger absolute values erroneously dominate weight calculations during the model training phase [21]. Min-Max normalization addresses this issue by compressing all values of each variable into a uniform range specifically, between 0 and 1 [6]. This mathematical transformation is executed without altering the original distribution shape or the relative distances between historical data points.

2.1.2 Data Training

In machine learning model development, dataset splitting is an essential step that must be performed before the algorithm enters the learning phase. This step aims to ensure that the model does not merely memorize historical data patterns (overfitting) but possesses robust generalization capabilities when encountering new data in the future. Therefore, the data-splitting strategy must be designed to allow for the objective measurement and validation of the algorithm's predictive performance. The aggregated data is sequentially divided into an 80% training set and a 20% testing set [17].

2.1.3 Training Model LSTM

In this study, the training phase employs a sequential Long Short-Term Memory (LSTM) model. To maintain evaluation integrity and prevent data leakage, normalization and time steps are employed for feature transformation in time-series modeling. Normalization is performed solely on the training data to avoid data leakage. The normalized data is restructured into a sequential format by setting the number of time steps to 15. Historical data (X) spanning 15 days serves as the predictor for the closing price ('Close') on the 16th day (y) a process consistently applied to both the training and testing datasets.

Accurately predict Indonesian technology sector stocks, noting that hyperparameter tuning is required [22]. The model is constructed with multiple LSTM hidden layers, utilizing the tanh activation function to process data. The initial LSTM layer is configured to pass the entire sequence of historical data (return sequences) to the subsequent layer, thereby preserving continuous temporal context. To prevent overfitting where the model essentially memorizes the training data Dropout layers are inserted between the LSTM layers. Finally, a Dense layer serves as the output layer, mapping the extracted temporal features to a single predicted value. Additionally, the Adam optimizer is employed; these hyperparameters were selected following a series of experiments [23].

2.1.4 Model LSTM

Long Short-Term Memory (LSTM) is an evolution of the Recurrent Neural Network (RNN) designed to address the limitations of conventional RNN models, particularly the vanishing gradient problem encountered when processing sequential data [24]. The vanishing gradient is a condition in which the model loses information while processing long sequences of data. The LSTM algorithm was first developed by Hochreiter and Schmidhuber in 1997 [25]. LSTM models are frequently used for time-related prediction and classification tasks.

LSTMs possess the ability to retain information over the long term (long-term dependencies) through a specialized architectural structure comprising three main gates: the forget gate, the input gate, and the output gate [24]. The information processing within an LSTM begins with the forget gate, which determines which past information is no longer relevant and should be removed from the cell state. Next, the input gate is responsible for determining what new information will be stored in the cell state. Finally, the output gate decides which part of the cell state will be output as the hidden state for use in the subsequent time step.

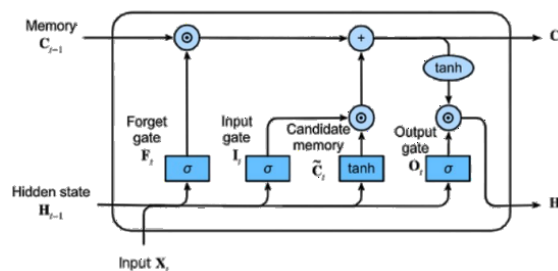


Figure 3. Architecture LSTM

The LSTM structure illustrates how a cell manages information through three primary gates to update the Cell State (C_t) and generate the Hidden State (H_t). The Forget Gate is the first gate, determining which information should be discarded from long-term memory; it receives inputs from the previous hidden state (H_{t-1}) and the current input (X_t). Subsequently, the Input Gate (i_t) works in conjunction with the Candidate Memory ($\tilde{C}_t$) to determine what new information will be stored.

The final stage of this mechanism is generating an output via the Output Gate (O_t). This gate employs a sigmoid function to determine which parts of the cell state will be passed on as output. Before becoming the final result, the updated cell state value is processed through a tanh function to normalize it and is then multiplied by the output of the Output Gate. This process produces a new hidden state (H_t) that carries important information to the next time step.

During the development of deep learning architectures particularly when implementing algorithms for complex time-series forecasting determining hyperparameters is a crucial step that directly impacts performance and prediction accuracy. Designing an artificial neural network requires careful testing to strike an optimal balance between the model's learning capacity and its ability to generalize to new data. To maximize temporal feature extraction while maintaining stability during training, the network structure in this study was designed with a specific, multi-layered configuration. The model incorporates key parameters for the LSTM architecture, utilizing four LSTM layers, with Dropout layers applied to prevent overfitting.

2.1.5 Evaluation Test

The evaluation phase serves as the primary parameter for validating the success of the model architecture. In the context of time-series models, evaluation is conducted by calculating the errors produced by the model [26]. Given the highly volatile and complex nature of stock price movements, it is crucial to determine the extent of the deviation between actual market closing prices and the prices predicted by the model. To mathematically quantify this error magnitude, this study employs two key performance indicators:

Mean Absolute Error (MAE), which represents the overall average absolute error, and Root Mean Squared Error (RMSE), which imposes a heavier penalty on predictions that deviate significantly.

2.1.5.1 Mean Absolute Error (MAE)

Mean Absolute Error (MAE) is a metric that calculates the average absolute difference between actual values and predicted values [27]. The absolute value of each difference is taken, meaning negative values are disregarded. In the evaluation of forecasting models, MAE provides an intuitive measure of the average error across the entire dataset.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

Information :

n : Number of data points or samples

y_i : Actual value

$\hat{y}_i$: Predicted value

2.1.5.2 Root Mean Squared Error (RMSE)

Root Mean Square Error (RMSE) is a metric used to assess the accuracy of an algorithm based on the magnitude of prediction error [27]. the lower the RMSE value (closer to 0), the more accurate the prediction results. RMSE is more sensitive to large errors because it imposes a higher penalty on them by calculating the square of the difference between the predicted and actual values [28].

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

Information :

n : Number of data points or samples

y_i : Actual value

$\hat{y}_i$: Predicted value

2.1.6 Data Visualization

The model's forecasting results are visualized through a user-friendly, web-based interactive dashboard. This dashboard allows users to compare the model's stock price projections against actual prices, enabling a direct evaluation of prediction accuracy. Additionally, the information is further illustrated using visual aids such as graphs and charts to enhance clarity and understanding.

3. Result and Discussion

A prediction system based on the Long Short-Term Memory (LSTM) method is used to forecast stock closing prices which can be recommended to users as investment references by analyzing historical data and market movements. This method operates by processing sequential data and identifying temporal dependency patterns within relevant features, such as OHLC data and the JCI (IHSG) index. The process entails complex computations, as data is evaluated through gating mechanisms (forget, input, and output gates) to retain critical long-term information while discarding irrelevant data. The output from this temporal feature extraction is then passed through multiple hidden layers, ultimately determining the predicted future stock price.

In practice, LSTM delivers more objective and adaptive prediction results by simultaneously considering multiple variables (multivariate analysis); consequently, the resulting price projections accurately reflect actual market dynamics and volatility based on comprehensive data analysis. This method is highly effective in helping users and investors optimally forecast stock price trends amidst market uncertainty, utilizing an in-depth, fully data-driven approach.

3.1.Data Collection Results

In this study, data was obtained from a financial information platform that provides specific market data. The Investing.com website served as the primary data source, providing historical price and transaction volume data for Pabrik Kertas Tjiwi Kimia Tbk (TKIM), as well as closing prices for the Jakarta Composite Index (JCI) as supplementary data. Data collection covered the period from January 1, 2021, to January 1, 2026, resulting in a time-series dataset comprising 1,205 daily observations.

Table 1. Dataset

No.	Date	Close	Opening	Highest	Lowest
0	04/01/2021	10.675	10.150	10.725	9.800
1	05/01/2021	10.700	10.500	11.100	10.375
2	06/01/2021	10.550	10.800	10.850	10.150
...	...	...	...	...	...
1202	24/12/2025	7.250	7.150	7.250	7.000
1203	29/12/2025	7.250	7.250	7.325	7.175
1204	30/12/2025	7.275	7.250	7.400	7.225

The entire dataset is systematically classified into five key attribute columns reflecting exchange trading activities, including: trading time parameters, the closing share price at the end of the session (last), the share price at the market opening (open), the highest price reached during the day (high), and the base price or lowest point (low).

3.2. Model Development

Preprocessing is employed to prepare the data for the modeling stage. The process begins by converting date formats to datetime objects and OHLC (Open, High, Low, Close) values to numeric format, alongside checking these features for noise. Next, feature preparation involves renaming the opening, high, low, and last price columns to OHLC, while the Jakarta Composite Index (JCI) data is reduced to just the closing values, labeled as JCI_Close. Following this renaming, the two datasets are merged into a single dataframe and sorted in ascending order.

After the data undergoes a series of preprocessing stages to prepare it for algorithmic analysis, the model training phase begins by splitting the dataset; this ensures the model can distinguish between the learning and evaluation phases and maintains computational stability. The entire dataset is divided into two primary segments: a training set (allocated 80%) and a testing set (20%). This specific split is performed to evaluate the model's ability to generalize patterns from previously unseen data. Subsequently, to prevent value dominance arising from the varying ranges of stock data and to accelerate the algorithm's convergence, data normalization is applied to all features using the MinMaxScaler method.

1. Min-Max Normalization

Min-Max normalization is a data normalization method that aims to transform original data into a new value range, thereby establishing a balanced benchmark between the pre- and post-process values. The formula for Min-Max normalization is given by the following equation:

$$x^i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (3)$$

Information :

X^i = Normalization results

x_i = I data point

x_{min} = Minimum value data

x_{max} = Maximum value data

The primary objective of this method is to transform original data into a new value range, thereby yielding balanced values on a predefined scale. The value x^i is calculated by taking the difference between the data value to be normalized and the minimum value in the dataset (min), then dividing it by the difference between the maximum value (max) and the minimum value (min) of the entire dataset [6]. Applying this technique compresses the data range into the interval [0, 1], facilitating data interpretation and analysis. Furthermore, this normalization helps mitigate the impact of scale disparities between variables, there by enhancing the performance of machine learning algorithms that are sensitive to data scaling.

The formula will be applied using data from the last five days for a specific issuer, PT Pabrik Kertas Tjiwi Kimia Tbk (TKIM). However, the boundary parameters for the variables must be determined first, the parameters used for this purpose are based on the training data.

Table 2. Constraint Feature

Feature	x_{min}	x_{max}
Open	5.675	16.850
High	5.775	17.250
Low	5.600	16.275

Close	5.675	16.850
JCI_Close	5.760	7.905

In the data transformation process using the Min-Max normalization method, establishing the lower and upper bounds for each feature is a mandatory prerequisite before performing the mathematical calculations. This study first defines the constraint parameters, which consist of the minimum and maximum values for each predictor variable namely Open, High, Low, Close, and JCI_Close. These boundary values are derived directly from the range of the collected historical data. The parameters presented in Table 2 will subsequently serve as the divisor and constant subtrahend in each iteration of the normalization calculation.

$$Open_{scaled0} = \frac{7175-5675}{16850-5675} = \frac{1500}{11175} = 0,1342 \quad (4)$$

$$High_{scaled0} = \frac{7175-5775}{17250-5775} = \frac{1400}{11475} = 0,1220 \quad (5)$$

$$Low_{scaled0} = \frac{7075-5600}{16275-5600} = \frac{1475}{10675} = 0,1381 \quad (6)$$

$$Close_{scaled0} = \frac{7150-5675}{16850-5675} = \frac{1475}{11175} = 0,1319 \quad (7)$$

$$JCI_Close_{scaled0} = \frac{8645,84-5760,58}{7905,39-5760,58} = \frac{1500}{11175} = 1,3452 \quad (8)$$

This was followed by the creation of time steps with a duration of 15 days equivalent to three work weeks. Constructing a deep learning architecture to analyze the complexity of time-series data requires precise hyperparameter tuning to ensure prediction accuracy. An optimal balance between the model's ability to capture information and its reliability when handling unseen data must be achieved through the systematic design of the neural network structure. In this study, this objective was realized through a stacked LSTM architecture comprising four hidden layers with unit distributions of 50, 50, 50, and 60, respectively. To prevent the model from merely memorizing training data patterns (overfitting), Dropout techniques were applied to the network with sequential values of 0.2, 0.1, 0.2, and 0.1. Computationally, the model was optimized using the 'Adam' algorithm and trained over 50 epochs, processing 32 data samples per batch.

3.3. Testing and Evaluation

Once the Long Short-Term Memory (LSTM) architecture has completed its learning iterations on the training data, the subsequent critical phase involves testing and evaluation. This testing phase is carried out using 20% of the total historical dataset, which was specifically designated as the testing set from the outset. Separating the testing data is essential to assess the model's capacity to generalize market fluctuation patterns to previously unseen data. To objectively validate the reliability of the projections, the model's performance is quantitatively evaluated by calculating the deviation between predicted prices and actual closing prices. These error rates are measured using standard statistical metrics Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

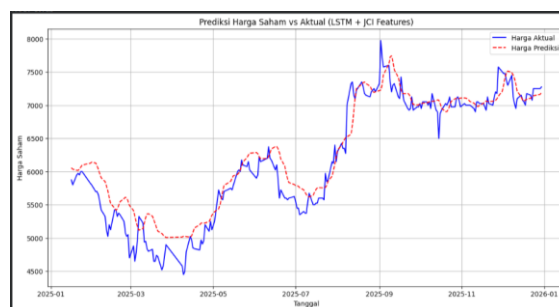


Figure 4. Comparison of predicted vs actual

Visually, the LSTM model demonstrates excellent capability in capturing macro-level price movement patterns or trends. It is evident that whether the actual price undergoes a sharp downward trend in early 2025 or a significant upward rally (uptrend) in mid-2025, the prediction line (red) consistently tracks the direction of these movements.

Table 3. Evaluation Metrics TKIM (OHLC + IHSG)

Evaluation Metrics	Output
Mean Absolute Error (MAE)	210.55
Root Mean Squared Error (RMSE)	280

Based on the conducted tests, the Long Short-Term Memory (LSTM) model demonstrated good performance, achieving a Root Mean Square Error (RMSE) of 280 and a Mean Absolute Error (MAE) of 210.55. Given that the stock price of Pabrik Kertas Tjiwi Kimia Tbk (TKIM) has reached levels as high as 17,000, an RMSE of 280 along with an MAE of 210.55 indicates strong model performance. Furthermore, the prediction model effectively captures the direction of actual price trends, although there are instances of a lag where the prediction line trails slightly behind changes in the actual price.

Table 4. Evaluation Metrics TKIM (OHLC)

Evaluation Metrics	Output
Mean Absolute Error (MAE)	291,68
Root Mean Squared Error (RMSE)	348,17

A comparison of the two scenarios reveals that incorporating JCI closing data as an external feature reduces the model's error rate; the RMSE dropped from 348.17 to 280, while the MAE fell from 291.68 to 210.55. This reduction indicates that information regarding overall market conditions represented by the JCI provides additional context absent in models relying solely on OHLC data, thereby bringing TKIM price predictions closer to actual prices. This aligns with the understanding that stock price movements are influenced not only by a company's internal factors but also by the direction of the broader market. The fact that the reduction in MAE was greater than that in RMSE suggests that the improvement was most pronounced regarding routine prediction errors, whereas significant errors associated with sharp price fluctuations persisted, keeping the RMSE higher than the MAE. Both scenarios were tested using identical data splits, time window lengths, and hyperparameter configurations, ensuring that the differences in results could be attributed to the inclusion of the JCI feature.

Beyond evaluating model performance against historical test data, a practical objective of this study is to implement the validated Long Short-Term Memory (LSTM) architecture for out-of-sample forecasting. At this stage, the optimized model is utilized to project the closing price movements of Pabrik Kertas Tjiwi Kimia Tbk (TKIM) specifically over the next seven trading days. This data extrapolation aims to provide investors with anticipatory insights regarding short-term price trends. Detailed projections of daily closing prices for the period beginning in early January 2026 are comprehensively summarized in Table 5.

Table 5. Prediction results for the next 7 days

Day	Stock Exchange Trading Day	Predicted Price (Closing)
H+1	Friday, 02 Jan 2026	Rp 7.127
H+2	Monday, 05 Jan 2026	Rp 7.147
H+3	Tuesday, 06 Jan 2026	Rp 7.157
H+4	Wed, 07 Jan 2026	Rp 7.158
H+5	Thurs, 08 Jan 2026	Rp 7.157
H+6	Friday, 09 Jan 2026	Rp 7.155
H+7	Monday, 12 Jan 2026	Rp 7.152

Interpreting out-of-sample forecasts often reveals the inherent characteristics of time-series algorithms when they no longer receive fresh data from the exchange. Based on the LSTM model's projections for TKIM stock, price movements over the next seven trading days are predicted to remain largely stagnant (sideways) with very low volatility. This phenomenon occurs because the model iteratively uses the previous day's prediction as the primary input to forecast the subsequent day. Due to the absence of actual market fluctuations and sentiment, data variance progressively diminishes, causing the price movement pattern in the final stages (Day+5 to Day+7) to mathematically tend toward a flat trajectory (flattening).

3.3. Data Visualization

This system is implemented as a web-based application named StockSight. The platform's interactive interface (frontend) is built using the React framework, while computational processes, machine learning model (LSTM) inference, and backend management are handled using the Python programming language.

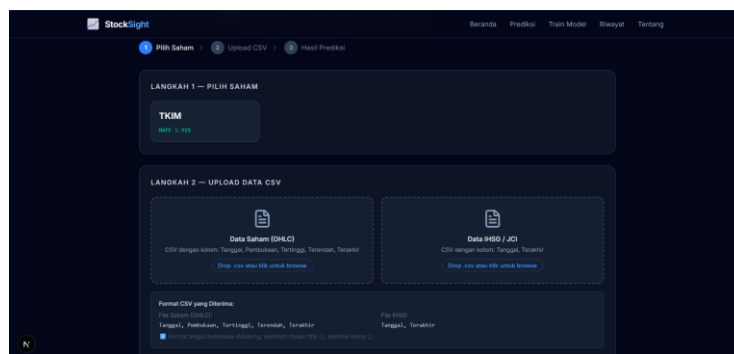


Figure 5. Input Data

The application's initial interface is intuitively designed using a step-by-step (stepper) approach, enabling users to easily input the required parameters. The following image shows the input page where users can select the stock ticker symbol to be analyzed such as TKIM and access an integrated interface for uploading CSV datasets containing historical stock price data (OHLC) and supporting index data (IHSG/JCI).

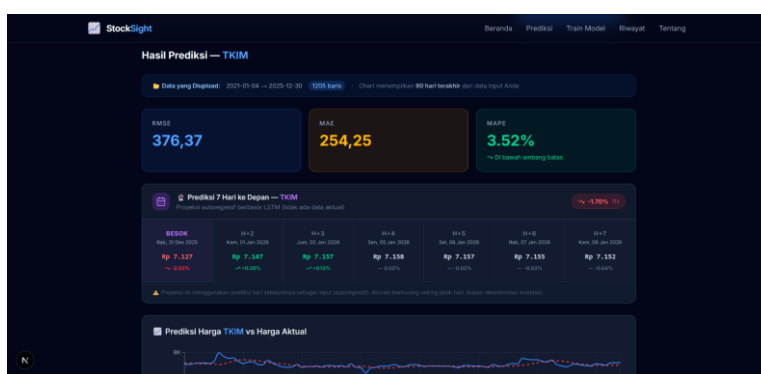


Figure 6. Prediction Result

Once the input dataset is sent to the Python backend for processing and evaluation by the LSTM model, the application renders a comprehensive prediction results dashboard. Figure 6 illustrates the computation results page, which directly presents model performance evaluation metrics such as RMSE, MAE, and MAPE and displays a detailed summary of projected TKIM stock closing prices for the next seven trading days.

4. Conclusion

This study highlights the importance of refining forecasting methods to capture the complex dynamics of capital markets. The developed system demonstrates enhanced analytical capabilities through a shift in data usage moving from approaches that often relied solely on single time-series (univariate) data to multivariate-based modeling. By simultaneously incorporating multiple interrelated variables, the model gains a broader learning dimension; it moves beyond dependence on the trends of a single historical value, becoming capable instead of understanding the interactions among various market indicators.

Based on the results of this study, an application implementing the Long Short-Term Memory (LSTM) model for predicting Indonesian stock prices has been successfully developed. The system produced in this study effectively predicts the closing prices of stocks listed on the Indonesia Stock Exchange by utilizing historical OHLC data and the JCI (Jakarta Composite Index) as multivariate input variables. Variables such as opening, high, low, and closing prices, along with JCI index data, were successfully processed by the model. The LSTM model effectively learned sequential patterns from historical stock price data, enabling it to generate price predictions and a seven-day forecast with a Root Mean Square Error (RMSE) of 280 and a Mean Absolute Error (MAE) of 210.55. This value is lower than that of the model using only OHLC data, which had an RMSE of 348.17 and an MAE of 291.68. These evaluation figures represent a high level of forecasting accuracy. Given that the processed stock prices are in the range of thousands of rupiah, the MAE value indicates that the system's average deviation or prediction error differs from the actual market price by only approximately Rp210. Data for TKIM covering the period from January

2021 to December 2025 (1,205 trading days) shows a price range of Rp4,450 to Rp16,850. If the actual price is Rp10,000, the model's average prediction falls within the range of Rp9,790 to Rp10,210. Furthermore, the RMSE value, being reasonably proportional to the MAE, demonstrates that the model is stable and free from significant extreme prediction errors (outliers); consequently, the projected trend direction is highly suitable for use as a tool to support investment decision-making.

However, the application developed in this study still has several shortcomings. The most significant limitation lies in the scope of model input features, which do not yet account for external macroeconomic factors such as the USD/IDR exchange rate, commodity prices, or benchmark interest rates that demonstrably influence stock performance in the Indonesian capital market. Furthermore, the model training process relies on static hyperparameter configurations, meaning it may not necessarily yield an optimal model architecture for every individual listed company.

References

- [1] M. Rivai and O. S. Nugraha, "Analisis Peramalan Harga Saham Menggunakan Temporal Convolutional Network: Studi Kasus PT Lippo General Insurance Tbk," *J. Probab. Stat.*, vol. 6, no. 2, pp. 48–55, 2025.
- [2] R. S. M. Putra, A. Larasati, T. T. Salsabila, and A. T. Oktavia, "Perbandingan Algoritma Temporal Convolutional Neural (TCN) dan Long Short-term Memory (LSTM) untuk Memprediksi Harga Saham Menggunakan Time Series Data," *J. Rekayasa Teknol. Inf.*, vol. 9, no. 3, pp. 282–292, 2025.
- [3] R. Anwar and L. Rassiyanti, "Analisis Komparasi Model Peramalan Prophet Dan Arima Dalam Memprediksi Harga Saham Penutupan PT ANTM," *Lattice J. J. Math. Educ. Appl.*, vol. 5, no. 1, pp. 57–74, 2025, doi: 10.30983/lattice.v5i1.9478.
- [4] C. Aulia, M. Fakhri, N. Acta, E. Saputra, and M. F. Rahman, "Penerapan Analisis Sentimen Berita Keuangan dari IDX Channel dan Long Short-Term Memory (LSTM) dalam Prediksi Perubahan Harga Saham pada Perusahaan XYZ," *JOISIE (Journal Inf. Syst. Informatics Eng.*, vol. 9, no. 2, pp. 392–402, 2025.
- [5] K. Prayogi, W. Gata, and D. P. Kussanti, "Prediksi Harga Saham Bank Central Asia Menggunakan Algoritma Deep Learning GRU," *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 13, no. 1, p. 647, 2024, doi: 10.35889/jutisi.v13i1.1910.
- [6] N. Zuzzaifa and S. D. Sancoko, "Implementasi Algoritme Long Short-Term Memory untuk Prediksi Harga Saham BBKA dan BBRI," *J. Telemat.*, vol. 19, no. 2, pp. 91–97, 2025, doi: 10.61769/telematika.v19i2.701.
- [7] R. Yafik and M. Azhari, "Analisis Perbandingan Metode LSTM dan BiLSTM untuk Prediksi Harga Saham Menggunakan Alpha Vantage," *J. Komput. Teknol. Inf. Sist. Komput.*, vol. 4, no. 3, pp. 1549–1557, 2026.
- [8] Mushliha, "Implementasi CNN-BiLSTM untuk Prediksi Harga Saham Bank Syariah di Indonesia," *Jambura J. Math.*, vol. 6, no. 2, pp. 195–203, 2024, doi: 10.37905/jjom.v6i2.26509.
- [9] A. P. Meriani and A. Rahmatulloh, "Perbandingan Gated Recurrent Unit (Gru) Dan Algoritma Long Short Term Memory (Lstm) Linear Refression Dalam Prediksi Harga Emas Menggunakan Model Time Series," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 1, 2024, doi: 10.23960/jitet.v12i1.3808.
- [10] K. Khotimah, F. Yudistira, and M. Ardiansyah, "Efisiensi Deep learning untuk Analisis Data dan Pengambilan Keputusan," *J. Insa. Peduli Pendidik.*, vol. 2, no. 2, pp. 79–82, 2024.
- [11] D. Safira, I. A. Sobari, and S. Rahayu, "Prediksi Harga Bbca Menggunakan Lstm Multivariat Dengan Fitur Bi-Rate Sebagai Indikator Eksternal," *Inf. Syst. J.*, vol. 8, no. 02, pp. 119–128, 2025, doi: 10.24076/infosjournal.2025v8i02.2372.
- [12] R. Saputra, Martanto, and R. D. Dana, "Analisis Perbandingan Model LSTM Univariat dan Multivariat untuk Prediksi Harga Saham Nokia (NOK)," *J. Komtika (Komputasi dan Inform.*, vol. 2, no. 9, pp. 205–213, 2025, doi: 10.53564/akademika.v17i2.1692.
- [13] M. C. Lubis, "Implementasi Multivariate Long Short-Term Memory (LSTM) untuk Forecasting Harga Saham ASHA," *Diophantine J. Math. Its Appl.*, vol. 5, no. 1, pp. 34–43, 2026.
- [14] S. Hartanto and A. A. S. Gunawan, "Temporal Fusion Transformers for Enhanced Multivariate Time Series Forecasting of Indonesian Stock Prices," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 7, pp. 140–148, 2024, doi: 10.14569/IJACSA.2024.0150713.
- [15] E. Fitri and D. Riana, "Analisa Perbandingan Model Prediction Dalam Prediksi Harga Saham Menggunakan Metode Linear Regression, Random Forest Regression Dan Multilayer Perceptron," *METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt.*, vol. 6, no. 1, pp. 69–78, 2022, doi: 10.46880/jmika.vol6no1.pp69-78.
- [16] P. Triya, N. Suarna, and N. D. Nuris, "Penerapan Machine Learning Dalam Melakukan Prediksi Harga Saham Pt. Bank Mandiri (Persero) Tbk Dengan Algoritma Linear Regression," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 1207–1214, 2024, doi: 10.36040/jati.v8i1.8958.
- [17] J. Kurniansyah, S. kurnia Gusti, F. Yanto, and M. Affandes, "Implementasi Model Long Short Term Memory (LSTM) dalam Prediksi Harga Saham," *Bull. Inf. Technol.*, vol. 6, no. 2, pp. 79–86, 2025, doi: 10.47065/bit.v5i2.1783.
- [18] R. Zapar, D. Pratama, Kaslani, C. L. Rohmat, and Faturrohman, "Penerapan Model Regresi Linier Untuk Prediksi Harga Saham Bank Bca Pada Bursa Efek Indonesia," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 196–202, 2024, doi: 10.36040/jati.v8i1.8215.
- [19] D. Tu. Saputro and D. Swanjaya, "Analisa Prediksi Harga Saham Menggunakan Neural Network Dan Net Foreign Flow," *Gener. J.*, vol. 7, no. 2, pp. 96–104, 2023, doi: 10.29407/gj.v7i2.20001.
- [20] R. F. Inaku and J. C. Chandra, "Implementasi Data Mining Dalam Prediksi Harga Saham Menggunakan Metode

- Long Short Term Memory (LSTM)," *J. Ticom Technol. Inf. Commun.*, vol. 12, no. 1, pp. 1–7, 2023, doi: 10.70309/ticom.v12i1.99.
- [21] N. W. Saputra, F. Insani, S. Agustian, and S. Sanjaya, "Penerapan Deep Learning Menggunakan Gated Recurrent Unit Untuk Memprediksi Harga Minyak Mentah Dunia," *Build. Informatics, Technol. Sci.*, vol. 5, no. 1, pp. 86–94, 2023, doi: 10.47065/bits.v5i1.3552.
- [22] R. S. Santoso and F. K. S. Dewi, "Konfigurasi Model Prophet Untuk Prediksi Harga Saham Sektor Teknologi di Indonesia Yang Akurat," *J. Buana Inform.*, vol. 15, no. 01, pp. 50–58, 2024, doi: 10.24002/jbi.v15i1.8634.
- [23] I. A. Saputra, A. V. Vitianingsih, Y. Kristyawan, A. L. Maukar, and J. F. Rusdi, "Forecasting Model of Export and Import Value of Oil and Gas Using Gated Recurrent Unit Method," *Teknika*, vol. 13, no. 2, pp. 239–243, 2024, doi: 10.34148/teknika.v13i2.861.
- [24] C. N. Yunantara and A. F. Daru, "Optimalisasi Prediksi Parameter Lingkungan Menggunakan Model LSTM Multivariat dan Univariat," *Bull. Comput. Sci. Res.*, vol. 5, no. 6, pp. 1315–1323, 2025, doi: 10.47065/bulletinsr.v5i6.813.
- [25] R. Al Kiramy, I. Permana, A. Marsal, M. R. Munzir, and M. Megawati, "Perbandingan Performa Algoritma RNN dan LSTM dalam Prediksi Jumlah Jamaah Umrah pada PT. Hajar Aswad," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 4, pp. 1224–1234, 2024, doi: 10.57152/malcom.v4i4.1373.
- [26] R. N. Silalahi and M. Muljono, "Perbandingan Kinerja Metode Linear Regression, LSTM dan GRU Untuk Prediksi Harga Penutupan Saham Coco-Cola," *Komputika - J. Sist. Komput.*, vol. 13, no. 2, pp. 201–211, 2024, doi: 10.34010/komputika.v13i2.12265.
- [27] I. Agustin, U. Khaira, and M. F. Putri, "PENERAPAN MODEL LSTM DAN ANALISIS SENTIMEN UNTUK PREDIKSI HARGA PENUTUPAN SAHAM PT BANK MANDIRI (PERSERO) TBK," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 10, no. 4, pp. 6649–6657, 2026.
- [28] I. Karijadi, Y. Widjaja, D. R. S. Dewi, and L. J. Asrini, "Studi Penerapan Deep Learning untuk Prediksi Pergerakan Indeks Harga Saham Gabungan," *J. Impresi Indones.*, vol. 4, no. 12, pp. 5408–5421, 2025, doi: 10.58344/jii.v4i12.7215.