Jurnal Ilmiah Informatika dan Komputer (Informatech)
Volume 3, Number 2, December 2026, page 166 - 176
DOI : 10.69533
Website : *https://ejournal.rizaniamedia.com/index.php/informatech*

E-ISSN: 3047-4752

Comparing Random Forest and SHAP Feature Selection for LSTM-Based Stock Price Prediction

Fatimatuz Zahra[1], Umar Zaky[2*]

[1,2*)]Universitas Teknologi Yogyakarta, Jl. Siliwangi, Jombor Lor, Sendangadi, Kec. Mlati, Kabupaten Sleman, Daerah Istimewa Yogyakarta 55285

[1]*email : fatimaif2006@gmail.com*

[*2]*email: umarzaky@uty.ac.id*

(*Article Received: 15 October 2026; Article Revised: 21 November 2026; Article Published: 1 December 2026;*)

ABSTRACT *– Feature selection is often used in stock price prediction, but its results remain mixed. This study analyzes the effect of feature selection based on Random Forest native feature importance (RF Importance) and SHAP on LSTM performance in predicting ADRO, ASII, and TLKM stock prices. Daily data from May 2021 to May 2026 were processed into 20 relative features from price, volume, technical indicators, and the Jakarta Composite Index (JCI), with return as the target. Both methods used the same Random Forest model and a mean-score threshold, while the LSTM was evaluated with five seeds and compared with a naive model. Feature selection reduced the input from 20 to 9 - 12 features (40% to 55%) and the LSTM parameters from 125,761 to 120,129 - 121,665 (3.3% to 4.5%). The MAPE of all schemes ranged from 1.76% to 1.90%, close to the naive model (1.72% to 1.87%), with differences between schemes of no more than 0.07 percentage points. Only for ADRO did the all-features scheme perform clearly better. Both methods retained the volume and JCI features and removed redundant trend indicators, with SHAP selecting fewer features. Therefore, the main benefit of feature selection was a simpler model input, not better prediction accuracy.*

Keywords *- Feature Selection; LSTM; Random Forest; SHAP; Stock Price Prediction*

Perbandingan Seleksi Fitur Random Forest dan SHAP pada LSTM untuk Prediksi Harga Saham

ABSTRAK *– Seleksi fitur sering digunakan dalam prediksi harga saham, tetapi hasilnya masih beragam. Penelitian ini menganalisis pengaruh seleksi fitur berbasis native feature importance Random Forest (RF Importance) dan SHAP terhadap kinerja LSTM dalam memprediksi harga saham ADRO, ASII, dan TLKM. Data harian periode Mei 2021 hingga Mei 2026 diolah menjadi 20 fitur relatif dari harga, volume, indikator teknikal, dan Indeks Harga Saham Gabungan (IHSG), dengan return sebagai target. Kedua metode menggunakan model Random Forest yang sama dan batas rata-rata skor, sedangkan LSTM dievaluasi dengan lima seed dan dibandingkan dengan model naive. Seleksi fitur mengurangi masukan dari 20 menjadi 9 - 12 fitur (40% hingga 55%) dan parameter LSTM dari 125.761 menjadi 120.129 - 121.665 (3,3% hingga 4,5%). MAPE seluruh skema berkisar 1,76% hingga 1,90%, mendekati model naive (1,72% hingga 1,87%), dengan selisih antarskema tidak lebih dari 0,07 poin persentase. Hanya pada ADRO skema seluruh fitur unggul secara jelas. Kedua metode mempertahankan fitur volume dan IHSG serta mengeliminasi indikator tren yang redundan, dengan SHAP memilih fitur yang lebih sedikit. Dengan demikian, manfaat utama seleksi fitur adalah masukan model yang lebih sederhana, bukan akurasi prediksi yang lebih baik.*

Kata Kunci *– Seleksi Fitur; LSTM; Random Forest; SHAP; Prediksi Harga Saham*

1. INTRODUCTION

The capital market plays an important role in the economy, and the number of investors in the Indonesian capital market reached 14.8 million in 2024, an increase of 22.22% from the previous year [1]. However, stock prices are dynamic and volatile, as they are influenced by global economic conditions and market sentiment [2]. Therefore, stock price prediction is needed to support investment decisions. Long Short-Term Memory (LSTM) is widely used because it can capture nonlinear patterns in sequential data, achieving a MAPE below 2% on Indonesian banking stocks [1] and performing well on 727 companies listed on the Indonesia Stock Exchange (IDX) [3].

In addition to the model architecture, LSTM performance is also affected by the input features. Adding technical indicators and Jakarta Composite Index (JCI) data as external features has been reported to improve prediction accuracy [4], [5], [6].

However, adding features does not always lead to better results; for BBCA stock, additional features increased the LSTM MAPE from 1.26% to 2.02% [7]. This indicates that not all features are relevant to the model, so feature selection is needed. One option is to use feature importance scores from a Random Forest model, namely Random Forest native feature importance (hereafter RF Importance) and SHAP (SHapley Additive exPlanations) [8], [9]. However, the results of applying feature selection to stock data remain mixed. Random Forest-based feature selection improved BiGRU performance on ten global stock indices [10], whereas another study found that feature selection improved performance only on some financial datasets [11]. The effectiveness of feature selection methods also varies across sectors [8] and can even reduce BiLSTM performance on BBTN stock [12]. In addition, importance-based methods can outperform SHAP for certain data characteristics [9].

Comparing RF Importance and SHAP matters for more than knowing which method selects fewer features. The two methods measure feature contribution in different ways. RF Importance is calculated from the decrease in impurity at each split, so when several inputs are correlated, their importance is spread among them and becomes difficult to interpret [13]. This condition is common in stock data because many technical indicators are calculated from the same closing price. SHAP measures the contribution of each feature to each prediction based on Shapley values [8], [9], and it has been used to explain which features drive TLKM stock forecasts [14]. As a result, the two methods may retain different features and lead to different conclusions about which market information is relevant. Feature selection is also expected to reduce computational cost and the risk of overfitting [15], and SHAP-based selection has been reported to reduce the number of features without lowering model performance [16]. Comparing the two methods therefore shows which features each method considers important, how much the model input can be simplified, and whether this simplification affects prediction accuracy.

Besides the input features, the choice of prediction target also affects the results. Stock prices are nonstationary, whereas returns are stationary [17]. Therefore, several studies have trained their models on returns [18], [19]. However, comparisons between RF Importance and SHAP for LSTM-based prediction of Indonesian stock prices with returns as the target remain limited. The mixed findings of previous studies also indicate the need for experiments on multiple stocks under an equivalent procedure, namely the same selection rule and target for both methods, with a naive model as a baseline, because complex models often fail to outperform naive forecasts on data with low predictability [20].

Therefore, this study compares three feature schemes, namely all features, features selected by RF Importance, and features selected by SHAP, in an LSTM model for predicting the stock prices of ADRO, ASII, and TLKM from the energy, automotive, and telecommunications sectors. This study aims to determine the effect of RF Importance and SHAP feature selection on LSTM performance in predicting stock prices, to measure how much the model input can be reduced, and to identify the feature patterns retained and eliminated by both methods.

2. LITERATURE REVIEW

Long Short-Term Memory (LSTM) is an extension of the Recurrent Neural Network (RNN) that has a cell state and three gates, namely the forget gate, input gate, and output gate, which regulate the information stored and discarded at each time step. According to Rosyd et al. [21], this mechanism enables LSTM to learn long-term dependencies in time series data such as stock prices. Rosyd et al. [21] applied a two-layer LSTM to BBCA stock and obtained a MAPE of 0.71%, while Sanjaya et al. [2] modified the LSTM architecture for Indonesian stock data and obtained a MAPE of approximately 1.41% to 1.70%. However, both studies used a single variable and therefore did not utilize information from other features.

Technical indicators are commonly used as additional features. They are calculated from historical price and volume data and describe market trends, momentum, and volatility. Dhokane and Agarwal [5] added RSI, Bollinger Bands, and MACD as LSTM features for stocks listed on the Indian stock exchange and obtained a MAPE of 2.206%. For BBRI stock, Zakka and Emigawaty [22] used SMA, EMA, RSI, MACD, and Bollinger Bands in a recursive LSTM and obtained a MAPE of 1.70%, while Aryanto et al. [23] integrated MACD, moving average, EMA, and RSI into a hybrid GRU-LSTM model and obtained a MAPE of 2.7%. In addition to technical indicators, Prasetyo and Hartomo [24] showed that the JCI can be used as an external feature that reflects overall market conditions. Adding these features increases the number of model inputs, although some indicators are derived from the same closing price and may therefore be redundant.

To select relevant features, this study uses two methods for assessing feature importance from a Random Forest model. Talib et al. [25] explained that Random Forest is an ensemble algorithm that combines the results of many decision trees through bagging and random feature selection, making it more resistant to overfitting. However, redundant or noisy features can still reduce model efficiency and

accuracy, so feature selection is still needed [25]. According to Wang et al. [9], RF Importance measures feature importance from the decrease in impurity at the nodes where the feature is used for splitting, so it can be obtained directly from the training process without additional computation. SHAP uses concepts from game theory to calculate the contribution of each feature to a prediction. The global importance of a feature is then obtained from the mean absolute SHAP value across all data, although its computation is more complex [9]. Lee and Cai [8] and Wang et al. [9] classify native feature importance as a model-specific method and SHAP as a model-agnostic method, and therefore both can produce different feature rankings on the same data. Since both methods can be calculated from the same Random Forest model, differences in the selected features reflect differences in how importance is scored rather than differences in the underlying model.

Guo [17] tested stock prices and returns using the Augmented Dickey-Fuller (ADF) test and found that stock prices are nonstationary, but returns are stationary. Stationary data have a mean and variance that do not change over time. A return is the percentage change in the closing price from one day to the next.

3. RESEARCH METHODS

This study consists of several stages: collecting stock price and JCI data, calculating technical indicators and relative features, preprocessing, data splitting, feature selection, feature characteristics analysis, normalization, LSTM training, price reconstruction, and model evaluation. After data splitting, the study is divided into two schemes. The all-features scheme uses 20 features without selection, while the feature selection scheme selects features using RF Importance and SHAP. The features from each scheme then go through the same stages of normalization, LSTM training, price reconstruction, and model evaluation. The research flow is shown in Figure 1.

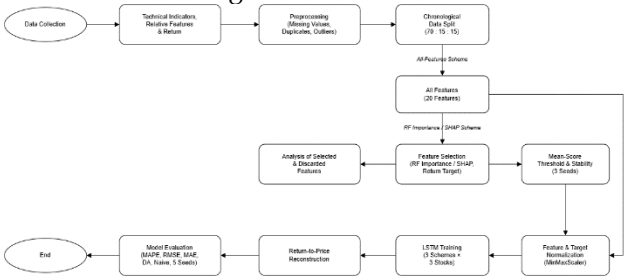


Figure 1. Research Flow

3.1. Data Collection

The data consist of historical daily stock prices of PT Alamtri Resources Indonesia Tbk (ADRO) in the energy sector, PT Astra International Tbk (ASII) in the automotive sector, and PT Telkom Indonesia (Persero) Tbk (TLKM) in the telecommunications sector, as well as the JCI as the market variable. The data were obtained from Yahoo Finance through an Application Programming Interface (API) using the yfinance library in Python, with the ticker symbols ADRO.JK, ASII.JK, TLKM.JK, and ^JKSE for the JCI. The API retrieves data automatically based on the ticker symbol, date range, and interval, so the data are consistent and can be retrieved again using the same parameters. Daily data were collected from May 11, 2021, to May 11, 2026, as prices adjusted for dividends and stock splits. The data for each stock include the open, high, low, and close prices and volume, which were then merged with the JCI data by transaction date, resulting in 1,199 trading days. The entire process was run in Google Colaboratory using Python with the pandas, NumPy, scikit-learn, TensorFlow, and SHAP libraries.

3.2. Technical Indicators

Based on the collected price data, 10 technical indicators were calculated as additional input features. They are divided into three groups: trend indicators (Simple Moving Average (SMA) and Exponential Moving Average (EMA)), momentum indicators (Moving Average Convergence Divergence (MACD), signal line, and Relative Strength Index (RSI)), and volatility indicators (the upper, middle, and lower Bollinger Bands). These indicators have been used as LSTM input features for stock price prediction in the Indonesian capital market [22], [23].

The SMA is calculated over 5- and 20-day periods using Equation (1):

$$SMA_n = \frac{\sum_{i=0}^{n-1} C_{t-i}}{n} \quad (1)$$

Description:

SMA_n : The value of the Simple Moving Average with period n.

C_{t-i} : The closing price on day $t - i$.

n : The number of time periods or sliding window.

t : The current timestep.

The EMA is calculated using periods of 5- and 20-days using Equation (2):

$$\text{EMA}_t = (\text{C}_t \times \alpha) + \left(\text{EMA}_{t-1} \times (1 - \alpha)\right)$$
$$\alpha = \frac{2}{\text{N} + 1} \quad (2)$$

Description:

EMA_t : The value of the Exponential Moving Average on day t.

EMA_{t-1} : The Exponential Moving Average value for the previous day, $t - 1$.

C_t : The closing price on day t.
α : Exponential weighting factor (smoothing constant).
N : The number of days in the period.

The MACD and Signal Line are calculated using Equation (3):

$$MACD = EMA_{12}(C) - EMA_{26}(C)$$
$$Signal\ Line = EMA_9(MACD) \quad (3)$$

Description:
MACD : Moving Average Convergence Divergence value.
Signal Line : Signal line, calculated as the 9-day EMA of the MACD series.
$EMA_{12}(C)$: 12-day Exponential Moving Average of the closing price.
$EMA_{26}(C)$: 26-day Exponential Moving Average of the closing price.
EMA_9 : Exponential Moving Average function with a 9-day period.

The RSI is calculated over a 14-day period using Equation (4):

$$RSI = 100 - \left(\frac{100}{1 + RS}\right)$$
$$RS = \frac{Average\ Gain_n}{Average\ Loss_n} \quad (4)$$

Description:
RSI : Relative Strength Index value (on a scale of 0 to 100).
RS : Relative strength ratio.
$Average\ Gain_n$: Average of the closing price increases when $C_t > C_{t-1}$ over n periods.
$Average\ Loss_n$: Average of the closing price decreases (in absolute value) when $C_t < C_\{t - 1\}$ over n periods.
n : Number of periods (14 days is used as the standard).

Bollinger Bands are calculated over a 20-day period using Equation (5):

$$BB\ Middle = SMA_{20}$$
$$BB\ Upper = SMA_{20} + (2 \times \sigma_{20})$$
$$BB\ Lower = SMA_{20} - (2 \times \sigma_{20}) \quad (5)$$
$$\sigma_{20} = \sqrt{\frac{\sum_{i=0}^{19}(C_{t-i} - SMA_{20})^2}{20}}$$

Description:
BB Middle : Middle line of the Bollinger Bands (the 20-day SMA).
BB Upper : Upper band of the Bollinger Bands.
BB Lower : Lower band of the Bollinger Bands.
SMA_{20} : 20-day Simple Moving Average.
σ_{20} : Standard deviation of the closing price over the last 20 days (price volatility).
C_{t-i} : Closing price on day $t - i$.

3.3. Relative Features and Target Return

Stock prices and the JCI follow long-term trends, so their values in rupiah or points during the validation and test periods may fall far outside the range of the training data. Therefore, all price- and volume-based features are converted into relative values, mostly relative to the closing price on the same day, while the RSI is kept unchanged because it already ranges from 0 to 100. These transformations use only data up to day t, so they do not contain any future information. Details of the 20 features are shown in Table 1.

Table 1. Features used

Group	Features	Description
Stock price	Open, High, Low	Relative difference from the same day's closing price
	Close	Daily return (change from the previous day)
Stock volume	Volume	Relative difference from the 20-day average volume
JCI price	Open, High, Low	Relative difference from the same day's JCI closing price
	Close	Daily JCI return
JCI volume	Volume	Relative difference from the 20-day average volume
Trend indicators	SMA 5, SMA 20, EMA 5, EMA 20	Relative difference from the closing price
Momentum indicators	MACD, Signal Line	Divided by the closing price
	RSI	Unchanged (0-100 scale)
Volatility indicators	BB Upper, BB Middle, BB Lower	Relative difference from the closing price

The model is trained to predict the daily return, which is calculated using Equation (6), while the next day's closing price is stored as the target price for evaluation.

$$Return_t = \frac{(Close_{t+1} - Close_t)}{Close_t} \quad (6)$$

Description:
$Close_{t+1}$: Closing price at time t + 1.
$Close_t$: Closing price at time t.
$Return_t$: Percentage change in the closing price from day t to day $t + 1$.

3.4. Preprocessing and Data Splitting

After the technical indicators and relative features were calculated, the first 19 rows had no indicator

values and the last row had no target. These rows were removed, leaving 1,179 rows for each stock. The data were then checked for missing values, duplicate rows, and duplicate dates, so that each row represented one trading day. The target price in each row was also compared with the closing price of the next day to make sure there was no data leakage. Outliers were detected using the Interquartile Range (IQR) method with a lower bound of $Q1 - 1.5 \times IQR$ and an upper bound of $Q3 + 1.5 \times IQR$. The outliers were not removed because they reflect real market events, such as volume spikes or sharp price changes.

The data were then split chronologically without shuffling in a 70:15:15 ratio, so that the model was not trained on data from periods later than those used for validation and testing. The results of the data split are presented in Table 2.

Table 2. Data Split

Data	Number of Days	Period
Training	825	June 14, 2021 – November 4, 2024
Validation	176	November 5, 2024 – August 11, 2025
Test	178	August 12, 2025 – May 8, 2026
Total	**1,179**	

The training data are used for feature selection and model training. The validation data are used only for early stopping and learning rate adjustment during training, not for feature selection. The test data are used only once, in the final evaluation.

3.5. Feature Selection

Feature selection was performed using a Random Forest Regressor with 500 trees, trained only on the training data to predict returns, the same target as the LSTM. Training was repeated with three seeds (42, 123, and 2024) to reduce the effect of randomness on tree construction. For each model, two feature importance scores were calculated: the RF Importance score from impurity reduction, and the mean absolute SHAP value calculated using TreeExplainer. Because both scores come from the same Random Forest model, differences in the selected features come only from the way feature importance is measured.

The scores of each feature were then averaged across the three seeds. Features with an average score greater than or equal to the mean score of all features were selected as the final features. The mean was used as the threshold because previous studies have used different cutoffs, such as a minimum correlation value or a fixed number or percentage of top-ranked features [15], and no single importance threshold can perfectly separate relevant and irrelevant features [13]. If the threshold were chosen separately for each scheme, for example based on validation performance, each method and stock could end up with a different selection rule, so differences in the results could come from the threshold rather than from the method itself. Because RF Importance scores sum to one while mean absolute SHAP values are expressed in units of return, the scores of each method were expressed as proportions of their total. With 20 features, the mean threshold is therefore $1/20 = 5\%$, which is the contribution each feature would have if all features contributed equally, so the selected features are those with an above-average contribution. This normalization does not change which features are selected. It also does not require choosing the number of features in advance or tuning the threshold on the validation or test data. This rule was applied in the same way to RF Importance and SHAP, and before LSTM training, so the selection did not depend on validation results or test data. The stability of each feature was measured as the number of seeds in which the feature was selected using the same rule, ranging from 0 to 3. Selection was performed separately for each stock, so the number and type of selected features could differ between stocks.

3.6. Analysis of Feature Characteristics

The features selected and discarded by RF Importance and SHAP were analyzed based on their data characteristics to identify the features that each method tended to retain or eliminate. These characteristics were calculated only from the training data, the same data used for feature selection, so no information from the validation or test data was used. The measured characteristics are shown in Table 3.

Table 3. Characteristics of the Analyzed Features

Characteristic	Description
Category	Feature group: stock price, stock volume, JCI price, JCI volume, trend indicators, momentum indicators, and volatility indicators
Correlation with return	Absolute Pearson correlation between the feature and return
Redundancy	Highest absolute correlation between the feature and any other feature
Skewness	Degree of asymmetry of the feature's distribution
Outlier	Percentage of outliers based on the IQR method
Stability	Number of seeds (0 to 3) in which the feature was selected

The values of each characteristic were then averaged for the selected and discarded feature groups, and the two groups were compared. The analysis was conducted separately for each method and each stock to see whether the same pattern

appeared consistently across all three stocks.

3.7. Data Normalization

Features and targets were normalized to a range of 0 to 1 using MinMaxScaler. The scaler was fitted only on the training data and then applied to the validation and test data to prevent data leakage. Normalization was performed separately for each scheme based on its final features, and it was carried out after feature selection because Random Forest and SHAP are not affected by data scale. The original closing prices were stored separately for price reconstruction in the evaluation stage.

3.8. LSTM Architecture and Training

The normalized data were organized into 30-day sliding windows, consisting of the previous 29 days plus the reference day. For the validation and test sets, the last 29 days of the previous period were used only as initial input, so all 176 validation days and 178 test days could be predicted without using targets from other periods. The LSTM architecture is shown in Figure 2.

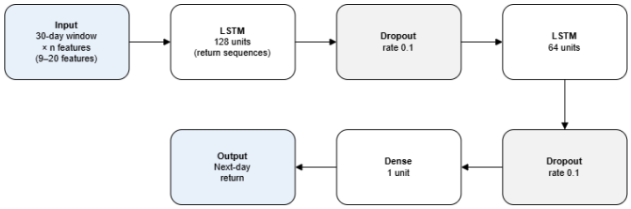


Figure 2. LSTM Architecture

The model was trained with the Adam optimizer (learning rate 0.001), MSE loss, a batch size of 32, and up to 100 epochs. EarlyStopping was used with a patience of 20 and kept the best weights. ReduceLROnPlateau halved the learning rate when the validation loss did not improve for 10 epochs. All three schemes and all three stocks used the same architecture and training settings, so the only difference between them is the input features. The model has 120,129 parameters with 9 features and 125,761 parameters with all 20 features, because only the first LSTM layer depends on the number of inputs. A fixed seed and deterministic operations were used so that the results can be reproduced.

3.9. Model Evaluation

The LSTM outputs are still in the form of returns. They were first converted back to their original scale using the target scaler and then reconstructed into prices using Equation (7). All evaluations were therefore done in price units, because the aim of this study is to predict stock prices.

$$\hat{C}_{t+1} = C_t \times (1 + \widehat{\text{Return}}_t) \quad (7)$$

Description:

$\hat{C}_{t+1}$: Predicted closing price for the next day.

C_t : Actual closing price on day t.

$\widehat{\text{Return}}_t$: Return predicted by the LSTM.

Model performance was evaluated on the test data using four metrics. Mean Absolute Percentage Error (MAPE) measures the average percentage error of the predictions relative to the actual prices. Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) measure the size of the errors in rupiah, with RMSE giving more weight to large errors. Directional Accuracy (DA) measures the percentage of days on which the direction of price movement (up or down) was predicted correctly, that is, when the actual and predicted returns have the same sign. The four metrics were calculated using Equations (8) to (11).

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (8)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (9)$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (10)$$

$$\text{DA} = \frac{100\%}{n}\sum_{i=1}^{n}\mathbb{1}\left[\text{sign}(R_i) = \text{sign}(\hat{R}_i)\right] \quad (11)$$

Overfitting was checked using the gap value, which is the difference between the best validation loss and the best training loss, expressed as a percentage of the best training loss. A naive model was used as a baseline. This model predicts that tomorrow's closing price is the same as today's closing price. The baseline shows whether the LSTM can capture any information beyond the latest price.

To check that differences between schemes were not caused by randomness in training, the final model of each scheme was retrained with five seeds (42, 123, 2024, 7, and 99) using the same architecture, data, and features. The results are reported as the mean of the five seeds, and standard deviation of MAPE is used to compare the schemes. If the difference between two schemes is smaller than the standard deviation across seeds, neither scheme is considered better.

4. RESULTS

This section presents the results of using LSTM to predict the stock prices of ADRO, ASII, and TLKM with three feature schemes: all features, features selected by RF Importance, and features selected by SHAP. The results start with a description of the data,

followed by the feature selection results and the model evaluation results.

4.1. Data Description

The data consist of daily stock prices of ADRO, ASII, and TLKM and the JCI, obtained through the yfinance API for the period from May 11, 2021, to May 11, 2026. All prices are adjusted prices. After each stock's data were merged with the JCI data by date, 1,199 trading days were obtained for each stock. A sample of the merged data is shown in Table 4, with trading volume in millions of shares.

Table 4. Stock Price and JCI Data

Date	Open	High	Low	Close	Vol. (M)
ADRO					
2021-05-11	375.93	383.80	374.36	375.93	78.40
…	…	…	…	…	…
2026-05-11	2,490.00	2,500.00	2,380.00	2,450.00	110.56
ASII					
2021-05-11	3,659.79	3,676.74	3,507.30	3,541.19	49.16
…	…	…	…	…	…
2026-05-11	5,925.00	6,050.00	5,825.00	6,050.00	26.38
TLKM					
2021-05-11	2,256.65	2,278.01	2,242.41	2,263.77	94.28
…	…	…	…	…	…
2026-05-11	2,764.57	2,783.06	2,699.85	2,736.83	58.49
JCI					
2021-05-11	5,965.85	5,966.66	5,911.38	5,938.35	120.81
…	…	…	…	…	…
2026-05-11	6,959.94	7,001.68	6,846.63	6,905.62	353.62

After preprocessing, 1,179 rowa remained for each stock. Table 5 shows the closing price statistics of these rows. It also shows the percentage of test data values outside the range [0, 1] after normalization with the MinMaxScaler fitted on the training data.

Table 5. Closing Price Statistics and Test Data Value Range

Statistic	ADRO	ASII	TLKM
Closing price			
Minimum	379.08	3,151.49	1,962.76
Maximum	2,530.00	7,044.31	3,681.84
Mean	1,296.46	4,554.76	2,897.81
Standard deviation	473.50	733.56	381.96
Std / mean (%)	36.5	16.1	13.2
Test data out of [0, 1] (%)			
Level features	37.87	60.39	15.11
Relative features	0.70	1.15	3.57

ADRO has the highest price variation, with a standard deviation equal to 36.5% of its mean price, followed by ASII at 16.1% and TLKM at 13.2%. The three stocks also have different price scales, so their RMSE and MAE values follow their own price scales. MAPE can be compared across stocks because it is calculated relative to actual prices.

Table 5 also shows that when features are used as absolute price levels, 15.11% to 60.39% of the test data values fall outside the range [0, 1] after normalization, because some prices in the test period are outside the price range of the training data. After the features are converted to relative values, this percentage drops to 0.70% to 3.57%.

4.2. Feature Selection Results

The Random Forest model used for feature selection had a training R^2 of about 0.86, while its out-of-bag (OOB) R^2 ranged from −0.054 to −0.008 for all three stocks. These negative OOB values show that the Random Forest cannot explain the variation in returns on data outside the training sample.

The RF Importance scores and SHAP contributions of each feature are shown in Figure 3.

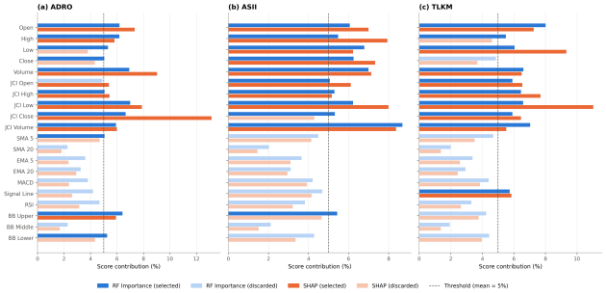


Figure 3. RF Importance and SHAP Scores

The selection threshold of 5% is the mean contribution of the 20 features, so features with contributions at or above this threshold are selected. The feature with the highest score differs for each stock. For ADRO, the highest RF Importance score belongs to JCI Low, and the highest SHAP score belongs to JCI Close. For ASII, both methods rank JCI Volume first, while for TLKM, RF Importance ranks Open first and SHAP ranks JCI Low first. The number of features selected by each method is shown in Table 6.

Table 6. Feature Selection Results

Stock	RF Importance	SHAP	Shared Features
ADRO	12 (9)	9 (9)	8
ASII	11 (10)	9 (8)	9
TLKM	10 (10)	9 (9)	9

The numbers in parentheses show the number of features selected in all three seeds. RF Importance selected 10 to 12 features and SHAP selected 9 features for each stock, which reduced the number of input features by 40% to 55% from the original 20. For ASII and TLKM, all features selected by SHAP were also selected by RF Importance. For ADRO, the two methods had 8 features in common, while JCI Open was selected only by SHAP. Most selected features passed in all three seeds, and for TLKM, all selected features passed in all three seeds for both methods.

The stock volume and JCI volume features were always selected by both methods for all three stocks,

and the JCI price features were almost always selected. In contrast, trend and momentum indicators were mostly not selected. Five features were always selected by both methods for all three stocks: Open, Volume, JCI High, JCI Low, and JCI Volume. Six features were never selected: SMA 20, EMA 5, EMA 20, MACD, RSI, and BB Middle. SMA 20 and BB Middle are identical, because the middle Bollinger Band is the 20-day SMA, so their importance was split between them and both received the lowest scores for all stock [13]. The number of selected features in each category is shown in Figure 4.

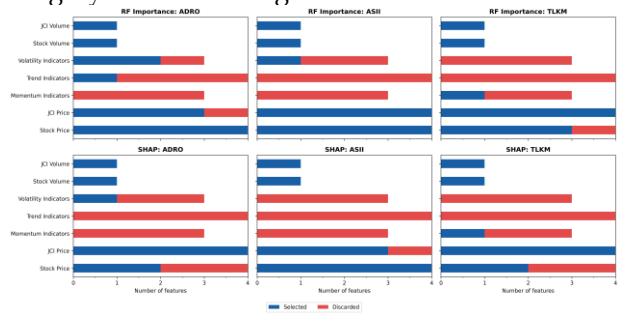


Figure 4. Number of Selected Features per Category

The feature characteristics on the training data show that the discarded features have an average maximum correlation with other features of 0.91 to 0.97, while the selected features have an average of 0.69 to 0.74. The average absolute correlation with returns is 0.037 to 0.098 for the selected features, higher than the 0.018 to 0.052 for the discarded features.

4.3. Model Evaluation Results

Each feature scheme was trained using five seeds (42, 123, 2024, 7, and 99) and then tested on 178 days of test data. The predictions were compared with a naive model that uses today's closing price as the prediction for tomorrow's price. The evaluation results of the three schemes and the naive model are shown in Table 7. Each scheme value is the mean of the five seeds.

Table 7. Model Evaluation Results on Test Data

Scheme	MAPE (%)	RMSE	MAE	DA (%)
ADRO				
All features	1.76	46.67	33.59	**43.93**
RF Importance	1.83	46.96	34.70	**42.47**
SHAP	1.81	47.20	34.41	**41.80**
Naive	1.72	46.03	32.83	-
ASII				
All features	1.86	156.77	110.70	**47.64**
RF Importance	1.86	157.18	110.80	**47.75**
SHAP	1.87	157.67	111.48	**49.44**
Naive	1.85	157.74	110.26	-
TLKM				
All features	1.89	79.85	57.74	**47.08**
RF Importance	1.90	80.22	58.02	**47.19**
SHAP	1.88	80.03	57.52	**49.55**
Naive	1.87	79.83	57.45	-

The standard deviation of MAPE across seeds ranges from 0.01 to 0.02 percentage points for all schemes. Therefore, two schemes are considered equivalent when their MAPE difference is smaller than the standard deviation. For ADRO, the all-features scheme produced the lowest MAPE of 1.76%. The RF Importance and SHAP schemes produced MAPEs that were 0.05 to 0.07 percentage points higher, and these differences were larger than the standard deviations. For ASII, the RF Importance scheme was equivalent to the all-features scheme, while SHAP was slightly higher by 0.01 percentage points but produced the highest DA of 49.44%. For TLKM, the SHAP scheme was equivalent to the all-features scheme, with SHAP giving the lowest MAPE of 1.88% and the highest DA of 49.55%, while RF Importance was slightly higher by about 0.01 percentage points. Overall, the MAPE difference between schemes did not exceed 0.07 percentage points for all three stocks.

Compared with the naive model, the MAPE of all schemes was 0.01 to 0.11 percentage points higher. For ASII, the RMSE of all three schemes was slightly lower than that of the naive model. The DA values of all schemes ranged from 41.80% to 49.55%, all below 50%.

The gap value for ASII is relatively small, from 10.72% to 11.65%, while ADRO and TLKM have much larger gaps, from 134.31% to 135.60% for ADRO and from 118.57% to 125.21% for TLKM. For ADRO and ASII, the three schemes have similar gaps. For TLKM, the RF Importance and SHAP schemes have gaps about 6 percentage points higher than the all-features scheme. Figure 5 compares the actual prices with the predictions of the three schemes on the test data for seed 42.



Figure 5. Actual and Predicted Prices on Test Data

The prediction lines of the three schemes almost

overlap and follow the actual price pattern with a lag of about one day, especially during sharp price changes, such as the rise in ADRO's price in January 2026 and the drop in TLKM's price in late January 2026.

5. DISCUSSION

The feature selection results show a consistent pattern across all three stocks. Both RF Importance and SHAP retained the stock volume, JCI volume, and most JCI price features, while trend and momentum indicators were mostly eliminated. Indicators such as SMA 20, EMA 5, EMA 20, and BB Middle are calculated from the closing price, so their values are very similar to other price features. This can be seen from the average maximum correlations of the eliminated features, which range from 0.91 to 0.97. In Random Forest, the importance scores of correlated features tend to be split among them [13], so each feature receives a small contribution and falls below the selection threshold. This is in line with Talib et al. [25], who stated that redundant features can reduce model efficiency. In contrast, the volume and JCI features carry information that is not contained in the stock's own price, namely trading activity and overall market conditions, and therefore have a higher correlation with returns. Between the two methods, SHAP is more selective because it concentrates the contribution on fewer features; for example, JCI Close accounts for 13.1% of the contribution for ADRO. As a result, for ASII and TLKM, all features selected by SHAP were also selected by RF Importance.

Feature selection reduced the number of input features by 40% to 55% but did not improve LSTM accuracy. The MAPE difference between schemes was no more than 0.07 percentage points, and for ASII and TLKM, most of these differences were within the inter-seed variation. One reason is the base model used to calculate the importance scores. The negative OOB R^2 of the Random Forest shows that the available features can hardly explain daily returns, so the score differences between features are relatively small and no subset of features carries much more information than the others.

This condition can also be seen in the LSTM performance. The MAPE of all three schemes was close to that of the naive model, DA was below 50%, and Figure 5 shows prediction lines shifted by about one day from the actual prices. This means that the LSTM tends to predict returns close to zero, so the reconstructed prices are close to the previous day's prices. Under these conditions, feature selection does not change the predictions much. In addition, the LSTM with all features may still be able to adjust the weights of less relevant features during training. A performance decline larger than the inter-seed variation occurred only for ADRO, the stock with the highest price variation. The large gap values for ADRO and TLKM appeared in all three schemes, which suggests that they were caused more by differences in data characteristics between the training and validation periods than by feature selection.

In terms of efficiency, reducing the features lowered the number of LSTM parameters from 125,761 to 120,129 to 121,665, or only 3.3% to 4.5%, because most parameters are in the connections between LSTM units. Thus, the main benefit of feature selection in this study is a simpler input without a noticeable decrease in accuracy for ASII and TLKM.

These results can be compared with five previous studies. Tjen and Zuhri [7] found that adding features increased the LSTM MAPE for BBCA stock from 1.26% to 2.02%, and this study also shows that more features do not always produce better predictions. Wijaya and Pribadi [12] reported that feature selection reduced BiLSTM performance on BBTN stock, which is similar to the result for ADRO, where both selection methods produced higher MAPE values than the all-features scheme. In contrast, Azman et al. [10] reported that Random Forest-based feature selection improved BiGRU performance on ten global stock indices. This difference may be due to the prediction target, because this study predicts returns, which are stationary [17] and harder to predict than prices. The findings of this study are closest to those of Pabuccu and Barbu [11], who stated that feature selection improves performance only on some financial datasets, as seen in the SHAP scheme for TLKM, which was equivalent to the all-features scheme. The retention of volume features is also in line with Bukhori et al. [14], who found that volume was the main contributor in the SHAP analysis of TLKM stock. However, RSI was also important in that study but was never selected in this study, possibly because that study predicted prices using about one year of data, while this study predicted returns over five years.

These findings have several implications. First, RF Importance and SHAP feature selection should be viewed mainly as a way to simplify the model input and to identify relevant market information, not as a way to improve accuracy, especially when the target is daily return. Similar to Hancock et al. [16], the selected-feature schemes for ASII and TLKM achieved almost the same MAPE as the all-features scheme with 45% to 55% fewer features. Second, the choice between the two methods depends on the goal. SHAP produced the smallest input, with 9 features for all stocks and similar accuracy, while RF Importance retained more features, so SHAP is more suitable when a simpler input is preferred. Third,

feature selection should be applied with caution to stocks with high price variation, because for ADRO both selection methods produced higher MAPE values than the all-features scheme. Fourth, prediction results should always be compared with a naive model. The MAPE of this study, from 1.76% to 1.90%, is in the same range as the 1.70% reported by Zakka and Emigawaty [22] and the 2.7% reported by Aryanto et al. [23], but the naive model in this study also produced a MAPE of 1.72% to 1.87%, so a MAPE below 2% does not by itself show good predictive performance [20]. For investors and practitioners, the consistent retention of stock volume and JCI features suggests that trading activity and overall market conditions are more useful inputs than trend indicators calculated from the same closing price.

This study also has several limitations that can guide future research. This study is limited to three stocks, one-day-ahead prediction, and features derived from price, volume, technical indicators, and the JCI. Only one threshold, the mean score, was tested, so future studies could compare other thresholds such as the median or a fixed number of top-ranked features. The LSTM also used the same architecture and hyperparameters for all schemes, so future studies could tune the hyperparameters for each feature scheme using the validation data. Future research could also add fundamental data and market sentiment, test more stocks from other sectors and longer prediction horizons, and use statistical tests to check whether the differences between schemes are significant.

6. CONCLUSION

RF Importance and SHAP feature selection mainly reduced the model input rather than improving stock price prediction. Both methods reduced the input from 20 to 9–12 features (40% to 55%) while the number of LSTM parameters decreased only slightly, by 3.3% to 4.5%, while the MAPE difference between schemes was no more than 0.07 percentage points. For ASII and TLKM, the MAPE of the selection schemes differed by only about 0.01 percentage points from the all-features scheme, while for ADRO the all-features scheme produced the lowest MAPE of 1.76%. The performance of all three schemes was also close to that of the naive model, with DA below 50%, which suggests that daily returns are still difficult to predict using technical features and the JCI. In the selection process, RF Importance retained 10 to 12 features, while SHAP retained 9 features because it concentrated the contribution on fewer features. Both methods kept the stock volume, JCI volume, and most JCI price features, and removed trend indicators such as SMA 20, EMA 5, EMA 20, and BB Middle, whose correlations with other price features reached 0.98 to 1.00. This pattern was consistent across ADRO, ASII, and TLKM, and for ASII and TLKM, all features selected by SHAP were also selected by RF Importance. Therefore, RF Importance and SHAP are more suitable for simplifying the model input and showing which market information is relevant than for improving prediction accuracy. Future research could test other selection thresholds, tune the hyperparameters for each feature scheme, and add fundamental and market sentiment data.

7. BIBLIOGRAPHY

[1] I. Navila, N. Q. Nada, and R. Renaldy, "Banking Stock Price Prediction Dashboard Using Long Short-Term Memory," *J. Tek. Inform. Jutif*, vol. 7, no. 2, pp. 2008–2022, Apr. 2026, doi: 10.52436/1.jutif.2026.7.2.5802.

[2] D. R. Sanjaya, B. Surarso, and Tarno, "Modification of the LSTM Model in Time Series Data Prediction," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 16, no. 1, pp. 14–26, Oct. 2025, doi: 10.24843/LKJITI.2025.v16.i01.p02.

[3] A. T. Haryono, R. Sarno, and K. R. Sungkono, "Stock Price Forecasting in Indonesia Stock Exchange Using Deep Learning: A Comparative Study," *Int. J. Electr. Comput. Eng. IJECE*, vol. 14, no. 1, pp. 861–869, Feb. 2024, doi: 10.11591/ijece.v14i1.pp861-869.

[4] A. F. Risnandar, F. Supriadi, and D. I. Junaedi, "Prediksi Harga Saham Indofood Menggunakan Long Short-Term Memory," *JATI J. Mhs. Tek. Inform.*, vol. 9, no. 3, pp. 4694–4701, May 2025, doi: 10.36040/jati.v9i3.13321.

[5] R. M. Dhokane and S. Agarwal, "LSTM Deep Learning Based Stock Price Prediction with Bollinger Band, RSI, MACD, and OHLC Features," *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 3, pp. 1169–1176, Mar. 2024.

[6] M. I. Saputra, E. Nurmawati, and R. Abyasa, "Predicting Stock Price Movements with Technical, Fundamental, and Sentiment Analysis Using the LSTM Model," *J. Inform.*, vol. 12, no. 1, pp. 1–9, Mar. 2025, doi: 10.31294/inf.v12i1.22299.

[7] Y. C. Tjen and H. F. Zuhri, "Impact of Volatility and Feature Engineering on the Forecasting Performance of LSTM and XGBoost Models: Evidence from Indonesian Banking Stocks," *Ingénierie Systèmes Inf.*, vol. 31, no. 5, pp. 1389–1402, May 2026, doi: 10.18280/isi.310502.

[8] J. Lee and L. Cai, "Comparing Model-Specific and Model-Agnostic Features Importance Methods Using Machine Learning with Technical Indicators: A NASDAQ Sector-Based Study," *Mach. Learn. Appl.*, vol. 23, p. 100799, Mar. 2026, doi: 10.1016/j.mlwa.2025.100799.

[9] H. Wang, Q. Liang, J. T. Hancock, and T. M. Khoshgoftaar, "Feature Selection Strategies: A

Comparative Analysis of SHAP-Value and Importance-Based Methods," *J. Big Data*, vol. 11, no. 1, p. 44, Mar. 2024, doi: 10.1186/s40537-024-00905-w.

[10] S. Azman, D. Pathmanathan, and V. Balakrishnan, "A Two-Stage Forecasting Model Using Random Forest Subset-Based Feature Selection and BiGRU with Attention Mechanism: Application to Stock Indices," *PLOS One*, vol. 20, no. 5, p. e0323015, May 2025, doi: 10.1371/journal.pone.0323015.

[11] H. Pabuccu and A. Barbu, "Feature Selection with Annealing for Forecasting Financial Time Series," *Financ. Innov.*, vol. 10, no. 1, p. 87, Oct. 2024, doi: 10.1186/s40854-024-00617-3.

[12] J. Wijaya and M. R. Pribadi, "Prediksi Harga Saham Sektor Perbankan LQ45 Menggunakan BiLSTM Dengan Perbandingan Metode Feature Selection," *JATI J. Mhs. Tek. Inform.*, vol. 10, no. 1, pp. 413–419, Jan. 2026, doi: 10.36040/jati.v10i1.16769.

[13] R. Dunne *et al.*, "Thresholding Gini variable importance with a single-trained random forest: An empirical Bayes approach," *Comput. Struct. Biotechnol. J.*, vol. 21, pp. 4354–4360, Jan. 2023, doi: 10.1016/j.csbj.2023.08.033.

[14] H. A. Bukhori, S. Bukhori, S. Anam, F. I. Yusuf, and M. Sari, "GWO-Enhanced Hybrid Deep Learning with SHAP for Explainable TLKM.JK Stock Forecasting," *J. Tek. Inform. Jutif*, vol. 6, no. 4, pp. 2566–2585, Sep. 2025, doi: 10.52436/1.jutif.2025.6.4.5205.

[15] H. H. Htun, M. Biehl, and N. Petkov, "Survey of feature selection and extraction techniques for stock market prediction," *Financ. Innov.*, vol. 9, no. 1, p. 26, Jan. 2023, doi: 10.1186/s40854-022-00441-7.

[16] J. T. Hancock, T. M. Khoshgoftaar, and Q. Liang, "A problem-agnostic approach to feature selection and analysis using SHAP," *J. Big Data*, vol. 12, no. 1, p. 12, Jan. 2025, doi: 10.1186/s40537-024-01041-1.

[17] Z. Guo, "Research on the Augmented Dickey-Fuller Test for Predicting Stock Prices and Returns," *Adv. Econ. Manag. Polit. Sci.*, vol. 44, no. 1, pp. 101–106, Nov. 2023, doi: 10.54254/2754-1169/44/20232198.

[18] E. G. A. Osman and F. A. Otaibi, "Integrating Deep Learning and Econometrics for Stock Price Prediction: A Comprehensive Comparison of LSTM, Transformers, and Traditional Time Series Models," *Mach. Learn. Appl.*, vol. 22, p. 100730, Dec. 2025, doi: 10.1016/j.mlwa.2025.100730.

[19] N. A. Burhani, D. Saepudin, and D. Adytia, "Prediksi Return Saham Berdasarkan Data Histori Dan Data Fundamental Menggunakan LSTM," *Log. J. Penelit. Inform.*, vol. 3, no. 1, Sep. 2025, doi: 10.25124/logic.v3i1.9236.

[20] N. Beck, J. Dovern, and S. Vogl, "Mind the naive forecast! a rigorous evaluation of forecasting models for time series with low predictability," *Appl. Intell.*, vol. 55, no. 6, p. 395, Apr. 2025, doi: 10.1007/s10489-025-06268-w.

[21] A. Rosyd, A. I. Purnamasari, and I. Ali, "Penerapan Metode Long Short Term Memory (LSTM) Dalam Memprediksi Harga Saham PT Bank Central Asia," *JATI J. Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 501–506, Feb. 2024, doi: 10.36040/jati.v8i1.8440.

[22] M. S. Zakka and E. Emigawaty, "Stock Price Prediction Using Deep Learning (LSTM) with a Recursive Approach," *J. Appl. Inform. Comput.*, vol. 9, no. 5, pp. 2468–2477, 2025, doi: 10.30871/jaic.v9i5.10514.

[23] F. H. D. Aryanto, R. H. P. Sejati, and F. I. Sanjaya, "Pengembangan Sistem Prediksi Saham Menggunakan Model Hybrid Gated Recurrent Unit–Long Short-Term Memory Berbasis Integrasi Indikator Teknikal Konvensional," *Terap. Inform. Nusant.*, vol. 6, no. 8, pp. 1296–1307, 2026, doi: 10.47065/tin.v6i8.8988.

[24] E. J. Prasetyo and K. D. Hartomo, "Multi-Industry Stock Forecasting Using GRU-LSTM Deep Transfer Learning Method," *J. INFOTEL*, vol. 15, no. 2, pp. 30–43, May 2023, doi: 10.20895/infotel.v15i2.941.

[25] E. A. H. Talib and A. F. Watratan, "Comparative Analysis of the Application of Feature Selection in Random Forest Regression for Stock Price Prediction," *Nusant. Hasana J.*, vol. 5, no. 3, pp. 334–348, Aug. 2025.