# Hybrid Support Vector Regression-Genetic Algorithm Model for Forecasting Stock Prices

[1]Muhammad Ulil Albab, [2*]Taghfirul Azhima Yoga Siswa, [3]Rofilde Hasudungan
[1,2,3]Department of Computer Engineering, Universitas Muhammadiyah Kalimantan Timur
Email: [1]2211102441104@umkt.ac.id , [2]tay758@umkt.ac.id , [3]rh219@umkt.ac.id

**Article Info**
***Article history:***
Received Jan 27th, 2026
Revised Feb 16th, 2026
Accepted Mar 05th, 2026

***Keyword:***
Genetic Algorithm
Machine Learning
Prediction
Stock
Support Vector Regression

**ABSTRACT**
The stock market exhibits a high level of volatility, which often leads to significant price fluctuations and increases the risk of financial losses for investors. Therefore, stock price prediction is an important tool to support investment decision-making, particularly for PT Aneka Tambang Tbk (ANTM.JK). This study aims to predict ANTM stock prices using Support Vector Regression (SVR) optimized with a Genetic Algorithm (GA). The data used in this study consist of 1202 historical stock price data of ANTM from September 11, 2020 to September 11, 2025, obtained from Investing.com, and the data are normalized using the Min-Max normalization method. The dataset is divided into training data and testing data using an 80:20 ratio, where 80% of the data are used for training and 20% for testing. The SVR model is constructed using the Radial Basis Function (RBF) kernel, while the GA is employed to optimize the SVR parameters in order to obtain the optimal parameter combination, with main GA parameters including population size of 50, 30 generations, crossover rate of 0.8, and mutation rate of 0.1. Model performance is evaluated by comparing the prediction results of SVR without optimization and GA-optimized SVR using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The experimental results indicate that applying GA improves the model's predictive performance. The SVR model without optimization produces RMSE, MAE, and MAPE values of 85.48, 59.02, and 2.62%, respectively. After parameter optimization using GA, the model performance improves as indicated by reduced error values, with RMSE of 75.97, MAE of 52.42, and MAPE of 2.42%.



***Corresponding Author:***
Taghfirul Azhima Yoga Siswa,
Department of Computer Engineering,
Universitas Muhammadiyah Kalimantan Timur,
Jl. Ir. H. Juanda No.15, Sidodadi, Samarinda Ulu District, Samarinda City, East Kalimantan 75124
Email: tay758@umkt.ac.id

**DOI: http://dx.doi.org/10.24014/ijaidm.v9i1.39057**

## 1. INTRODUCTION

The stock market is a collection of exchanges and trading venues that bring together buyers and sellers to trade shares of publicly listed companies and constitutes an integral part of the free-market economic system by providing equal access to trading and capital exchange activities for all types of investors [1]. Stock market activities span various major industrial sectors, including property, industry, communication, technology, healthcare, consumer goods, energy, finance, and materials [2]. In Indonesia, several stocks such as PT Aneka Tambang Tbk (ANTM), PT Merdeka Copper Gold Tbk (MDKA), and PT United Tractors Tbk (UNTR) are recorded as the three issuers with the highest net foreign buying value. Among them, PT Aneka Tambang Tbk (ANTM) ranks first, as the company's performance is projected to drive net profit growth of 43–51% and offer dividend potential of up to 12%, thereby strengthening its attractiveness to foreign investors [3].

In January 2025, the number of Indonesian capital market investors reached 15.16 million Single Investor Identification (SID) accounts, with 99.7% coming from the younger generation [4]. Nevertheless, the experience during the 2020 pandemic demonstrated that the Jakarta Composite Index (JCI) could decline by more than 5% in a single day and fall by up to 36% on a year-to-date basis [5]. Furthermore, macroeconomic conditions in early 2025, such as low inflation of 0.76% in January, 0.09% in February, and 1.03% in March 2025 [6], along with a reduction in the BI Rate to 5.75%, were insufficient to maintain market stability [7]. These conditions indicate increasing stock market instability and heightened investment risk, thereby necessitating reliable prediction methods to reduce uncertainty prior to investment decisions [8].

Initially, stock market prediction was conducted using traditional methods based on fundamental analysis and price movement patterns, and later evolved into modern technology-based approaches [9]. With the advancement of artificial intelligence, machine learning (ML) and deep learning techniques have been widely adopted due to their ability to process both historical and real-time data, thereby assisting investors in making informed decisions to reduce risk while increasing potential returns [10]. ML enables systems to automatically learn from data and has been extensively applied in pattern recognition, classification, regression, and prediction tasks [11]. Several ML algorithms that have been employed in previous stock prediction studies include Support Vector Machine (SVM), Decision Tree (DT), Logistic Regression (LR), Naïve Bayes (NB), K-Nearest Neighbor (KNN), AdaBoost, XGBoost, Artificial Neural Network (ANN), Hidden Markov Model (HMM), and Support Vector Regression (SVR) [12], [13].

A study on the MSI 20 index demonstrated that SVR achieved the best predictive performance, with a Mean Absolute Error (MAE) of 3.092, Root Mean Square Error (RMSE) of 3.993, and Mean Absolute Percentage Error (MAPE) of 0.368. These results outperformed Multilayer Perceptron (MLP) with MAE of 3.101 and RMSE of 4.018, Long Short-Term Memory (LSTM) with MAE of 5.065 and RMSE of 6.322, and XGBoost with MAE of 9.165 and RMSE of 13.515 [14]. Another study on stock price prediction of BBCA using SVR showed the best performance with an RMSE of 4.79% and MAE of 3.52% compared to Random Forest (RF), which produced an RMSE of 6.80% and MAE of 5.35%, and K-Nearest Neighbors (KNN) with an RMSE of 6.98% and MAE of 5.21% [15].

SVR is an extension of SVM designed for continuous value prediction using the ε-insensitive loss function, which ignores small errors and only penalizes data points that lie far from the predicted function [16]. The SVR algorithm maps data into a high-dimensional feature space and seeks an optimal hyperplane, enabling accurate target estimation, effective handling of non-linear problems, strong generalization capability, and robustness to noise and outliers [17]. However, using SVR with default parameters still has limitations in capturing complex nonlinear patterns. This is evidenced by a study on stock data, where SVR performed worse, with RMSE 0.1575 and MAE 0.1364, compared to LSTM, which achieved RMSE 0.0316 and MAE 0.0240. Based on these results, the study recommended the exploration of hybrid models and the integration of optimization techniques to improve SVR performance [18]. A similar conclusion was reported by on cryptocurrency data, where SVR obtained an MAPE of 0.07772, prompting the authors to suggest the development of optimization methods to improve the consistency of SVR performance [19].

SVR also suffers from several limitations, including long training time for large datasets, difficulty in modeling complex and highly dynamic data, challenges in selecting appropriate hyperparameters, and sensitivity to hidden noise in the input data [20]. To address these issues, various metaheuristic algorithms such as Artificial Bee Colony (ABC), Genetic Algorithm (GA), Grey Wolf Optimizer (GWO), Particle Swarm Optimization (PSO), and Water Cycle Algorithm (WCA) have been employed for hyperparameter optimization. GA is an optimization technique inspired by natural selection and biological evolution, utilizing crossover, mutation, and selection processes to solve complex optimization problems. It has demonstrated strong capability in exploring large solution spaces and effectively solving complex optimization tasks [21]. A study showed that applying GA to optimize LSTM parameters for stock data reduced RMSE from 64.97 to 18.03 and MAE from 35.46 to 14.98 [22] . Similarly, GA-based optimization of SVR parameters significantly improved energy consumption prediction performance, reducing RMSE from 6.50 to 1.53 and MAE from 5.02 to 1.18 [23]. However, most previous studies have focused on stock indices, other companies, or different financial datasets, while the application and evaluation of the SVR model optimized using a GA (SVR-GA) specifically for ANTM.JK stock prediction remain limited. This indicates a research gap that still needs to be further examined to evaluate the effectiveness of the SVR-GA approach on ANTM.JK stock data.

## 2. RESEARCH METHOD

PT Aneka Tambang Tbk (ANTM) is a state-owned mining company engaged in the management and sale of mineral resources, such as nickel, gold, and bauxite. ANTM is part of Indonesia's state-owned mining holding company, MIND ID. The company's main activities cover the entire mining value chain, ranging from exploration and extraction to processing and marketing of mining products. ANTM operates its business through three main segments: the nickel business, which is one of its flagship products, the gold business and precious metal refining services; and the bauxite business, which is processed into alumina.

As a commodity-based company, ANTM's stock price is highly influenced by global market fluctuations, which result in high volatility and non-linear data patterns. These are the characteristics which present a significant challenge for traditional orecasting methods. Therefore, this study is set to employ a hybrid SVR-GA approach, in which the GA is specifically utilized to optimize SVR hyperparameters. Moreover, this study aims to ensure the SVR-GA model is applicable to capture the complex trends in ANTM's historical stock data effectively.

### 2.1 Research Data

The data used in this study are derived from the historical stock data of PT Aneka Tambang Tbk (ANTM), which include the opening price (open), highest price (high), lowest price (low), closing price (close), percentage change (change %), and trading volume. The dataset was obtained from Investing.com and covers a period of the last five years. A sample of daily data for one week is presented in Table 1.

**Table 1.** ANTM Stock Data Sample

| Date | Close | Open | High | Low | Vol. | Change % |
|---|---|---|---|---|---|---|
| 11/09/2025 | 3.390 | 3.530 | 3.570 | 3.380 | 161,96M | -3,42% |
| 10/09/2025 | 3.510 | 3.580 | 3.600 | 3.470 | 192,13M | -4,10% |
| 09/09/2025 | 3.660 | 3.680 | 3.930 | 3.590 | 514,31M | 1,39% |
| 08/09/2025 | 3.610 | 3.470 | 3.710 | 3.450 | 400,45M | 6,49% |
| 04/09/2025 | 3.390 | 3.470 | 3.470 | 3.370 | 142,47M | -2,59% |
| 03/09/2025 | 3.480 | 3.520 | 3.580 | 3.360 | 393,01M | 0,29% |
| 02/09/2025 | 3.470 | 3.300 | 3.480 | 3.290 | 441,94M | 8,44% |

Table 2.1 presents a sample of the historical stock data used in this study. This research uses daily time series data of PT Aneka Tambang Tbk stock downloaded from Investing.com, covering the period from September 11, 2020 to September 11, 2025. The total number of data entries obtained is 1,202 records, including the attributes date, closing price (close), opening price (open), highest price (high), lowest price (low), trading volume (volume), and price change (change %).

### 2.2 Research Flow

This research was conducted through a series of interrelated stages forming a clear research workflow. The stages of the research process are illustrated in Figure 1.

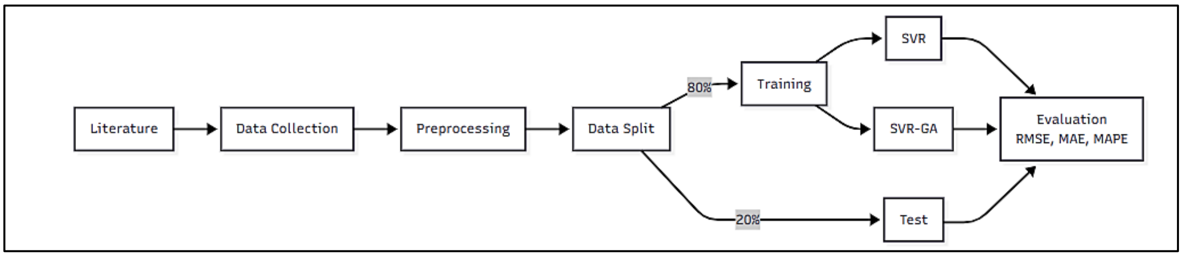


**Figure 1.** Research Flow

In accordance with Figure 1 which presents research flow, this study was conducted through a systematic workflow which designed to optimize the prediction accuracy. First, the process was begun with a Literature Review to identify the research gap and establish the theoretical foundation for stock price forecasting. Seconf, the Data Collection stage was performed by gathering historical stock data of ANTM from Investing.com. Third, the raw data entered the preprocessing stage where it was cleaned and then transformed, in this stage, feature selection and data normalization were conducted to ensure its suitability for machine learning analysis. Fourth, the Data Split stage was conducted to divide the dataset into two parts, mainly 80% of the data was allocated for Training while the remaining of 20% data was reserved for the Test set to evaluate the model's performance on unseen data. Fifth, during the modeling phase, the training data were processed using two different approaches including a standard SVR and an optimized SVR-GA. The GA itself was specifically worked within the SVR-GA block to find the most optimal hyperparameters. Sixth, both models were validated against the Test Data in the Evaluation Stage. In this stage, the accuracy was measured and compared using RMSE, MAE, and MAPE metrics to determine the effectiveness of the proposed hybrid model.

### 2.3 Data Preprocessing

Data preprocessing is the initial step in machine learning that transforms raw data into a structured format through cleaning, integration, transformation, and reduction processes, thereby improving data quality and helping the model better understand features to produce more accurate predictions [24]. Preprocessing also includes preparing and organizing data to make it more suitable for analysis, ensuring that the data quality is maintained and ready for the next stage [25].

#### 2.3.1 Data Cleaning

Data cleaning is an essential step to eliminate missing values or inconsistent data [26]. In this stage, the cleaning process for the ANTM dataset which consists of 1.202 records involved three stages mainly identification, handling missing values, and removing duplicates. Through identification, the data structure was reviewed to ensure consistency across 1.202 rows. Through handling missing values, the dataset was then inspected for null entries. However, No. missing values were found as it meant that the data integrity for the 5-year-time-series were well-maintained. Through removing duplicates, a verification for duplicate rows was conducted to prevent data reduncancy and ensured each trading date was unique.

#### 2.3.2 Data Transformation

Data transformation is a stage performed to convert or adjust data prior to its use in analysis or machine learning modeling, with the objective of ensuring that the data meet the requirements of the analysis [27]. For the ANTM dataset, symbols (%), text notations (M), and varying price scales were then converted into pure numerical values as shown in Table 2. This stage involved the simplification to numerical constants in order to ensure consistency across all features.

**Table 2.** Data Transformation

| Feature | Before Transformation | After Transformation |
|---|---|---|
| Date | 28/08/2025 | 28/08/2025 |
| Last Price | 2.98 | 2980 |
| Opening Price | 3.00 | 3000 |
| Highest Price | 3.02 | 3020 |
| Lowest Price | 2.92 | 2920 |
| Volume | 121.88M | 121,880,000 |
| Change (%) | 0.68% | 0.68 |

#### 2.3.3 Data Selection

Data selection is a dimensionality reduction technique aimed at selecting relevant features by eliminating redundant and irrelevant features in order to improve the performance of machine learning algorithms and simplify the model [28]. In this study, the features used consist of open, high, low, and close (OHLC) prices as input variables, while the closing price is also employed as the target variable. Other features, such as trading volume and percentage change, are excluded from the modeling process, as they are considered less relevant for closing price analysis. Therefore, the model is focused solely on OHLC price data [29]. The features selected for this study are presented in Table 3.

**Table 3.** Data Selection

| Last (Close) | Open | High | Low |
|---|---|---|---|
| 3390 | 3530 | 3570 | 3380 |
| 3510 | 3580 | 3600 | 3470 |
| 3660 | 3680 | 3930 | 3590 |
| 3610 | 3470 | 3710 | 3450 |
| 3390 | 3470 | 3470 | 3370 |
| 3480 | 3520 | 3580 | 3360 |
| 3470 | 3300 | 3480 | 3290 |

Table 3 displays the features used in this study, specifically the Last (Close), Open, High, and Low prices as input variables, with the Last (Close) price also serving as the target variable. Other features such as Volume (M) and Percentage Change (%) were excluded from the modeling process. This decision aligns with the earlier study, which emphasizes that these features are less relevant for predicting the final closing price and allows the model to focus purely on core price-action data [30].

#### 2.3.4 Lagged Features

Lagged features are characteristics formed by shifting time series data points forward or backward to introduce historical values as new features. This technique is important in autoregressive modeling because it captures the influence of past data on current conditions and improves prediction accuracy [31]. The computation of lagged features is shown in Equation 1.

$$\mathrm{X}_{\mathrm{t}}^{(\mathrm{lag\ k})} = \mathrm{X}_{\mathrm{t-k}} \tag{1}$$

To capture the temporal dependencies inherent in stock price movements, this study implemented a lag-1 configuration. In this process, the historical values from the previous days were transformed into new input features, where the lagged feature value at time *t* was computed as $X_t^{(lag\ k)}$ with k = 1. Specifically, the values of $Open_{(t-1)}$, $High_{(t-1)}$, $Low_{(t-1)}$, and $Close_{\ (t-1)}$ were utilized as inputs to predict the current day's closing price, $Close_{(t)}$. Consequently, the lagging process removed the first data record, reducing the total dataset from 1.202 to 1.201 records.

#### 2.3.5 Data Normalization

Data normalization is the process of transforming data to a uniform scale, typically 0 to 1, to facilitate analysis and improve the performance of machine learning models that are sensitive to data distances. One commonly used method is Min-Max Normalization [32]. The computation of Min-Max Normalization is shown in Equation 2.

$$\mathrm{X}' = \frac{\mathrm{X}-\mathrm{X}_{\min}}{\mathrm{X}_{\max}-\mathrm{X}_{\min}} \tag{2}$$

X' is the value of Min-Max Normalization which is achieved. While X represents the original value, $X_{min}$ and $X_{max}$ denote the minimum and maximum values within the feature set, respectively. This formula ensures that high-magnitude features do not dominate the learning process during SVR training [33].

#### 2.3.6 Data Split

Data split is a common approach in model validation by dividing data into training data and testing data. Training data are used to build the model, while testing data are used to evaluate prediction performance [34]. Previous studies have shown that an 80:20 split ratio provides optimal performance for stock index prediction [35].

#### 2.3.7 Data Denormalization

Data denormalization is the process of converting model output values back to their original range before normalization [36]. The computation of data denormalization is shown in Equation 3.

$$\mathrm{x_i} = \mathrm{x}'(\mathrm{x_{max}} - \mathrm{x_{min}}) + \mathrm{x_{min}} \tag{3}$$

$x_i$ is the value of data denormalization which is achieved through formulation. While x' denotes the data to be normalized, and $x_{max}$ also $x_{min}$ correspond to the maximum and minimum values of the original data. This step is crucial to return the predicted stock prices to the Indonesian Rupiah (IDR) scale so they can be compared directly with actual datmarket prices [37].

### 2.4 Support Vector Regression (SVR)

Support Vector Regression (SVR) is a high-performance machine learning algorithm from the SVM family. Unlike conventional regression models, SVR focuses on controlling the error margin, enabling it to handle complex and nonlinear relationships between input and output. Through kernel functions, SVR maps input features into higher-dimensional space, making it effective for nonlinear data and robust to outliers [38]. The linear regression function in Support Vector Regression (SVR) is defined as follows in Equation 4.

$$f(\mathrm{x}) = w \cdot x + b \tag{4}$$

Where f(x) is the predicted value, w represents the weight vector that determines the orientation of the hyperplane, and b denotes the bias or intercept.

The objective of SVR is to minimize the model complexity by minimizing the norm of the weight vector, which is expressed as Equation 5.

$$\min \frac{1}{2} ||w||^2 \tag{5}$$

Where Min is the minimization operation, $\frac{1}{2}$ is a constant used to simplify the differentiation process, and $||w||^2$is a squared norm of the weight vector.

SVR introduces an e-insensitive loss function, which constrains the prediction error to lie within a predefined tolerance margin. The mathematical representation of this loss function is defined in Equation 6.

$$|y_i - (w \cdot x_i + b)| \leq \epsilon \tag{6}$$

Where $y_i$ is the actual value of the *i-th* data point, $x_i$ is the input value of the *i-th* data point, and $\epsilon$ is the acceptable error margin.

To account for data points outside the e-insensitive boundary, the model incorporates slack variables, which are defined by the following constraints in Equation 7.

$$\epsilon_i = \max(0, |y_i - (w \cdot x_i + b)| - \epsilon) \tag{7}$$

Where $\epsilon_i$ is the deviation outside the $\epsilon$ margin and max (0,n) is a function that ensures only errors exceeding the margin are penalized.

The primary objective of SVR is to minimize model complexity while penalizing predictions that fall outside the tolerance margin. This is achieved by minimizing the norm of the weight vector combined with an $\epsilon_i$ loss function. Based on [39], the optimization problem is formulated as equation 8.

$$\min \frac{1}{2} |w|^2 + C \sum_{i=1}^{N} \epsilon_i \tag{8}$$

Subject to the constraint (equation 9).

$$\begin{cases} y_{i\,-} - (w.x_i + b) \leq \epsilon + \epsilon_i \\ (w.x_i + b) - y_i \leq \epsilon + \epsilon_i^* \\ \epsilon + \epsilon_i^* \geq 0 \end{cases} \tag{9}$$

Where C is the regularization parameter that controls the trade-off between model complexity and penalty for prediction errors, $\epsilon_i$ is the total prediction error outside the $\epsilon$ margin, and N is the number of data samples.

### 2.4.1 Kernel Function-RBF

The Gaussian Radial Basis Function (RBF) kernel is the most widely used kernel in SVM because it can implicitly map data into a high-dimensional space [40]. The RBF kernel function is formulated as shown in Equation 10.

$$K_{\gamma,d}(x, y) = \exp\left(-\frac{||x-y||^2}{\gamma^2}\right), \quad x, y \in R^d \tag{10}$$

Where $K_{\gamma,d}(x, y)$ represents the Gaussian Radial Basis Function (RBF) Kernel value between data points x and y, where both x and y are input vectors within a d-dimensional feature space ($x, y \in R^d$). The term $||x-y||^2$ denotes the squared Euclidean distance between these data points, while $\gamma^2$serves as the Kernel parameter that controls the width of the Gaussian function.

Through systematic testing of various ranges, this study determined the optimal $\gamma$ value to be 0.001 to ensure the best generalization performance.

### 2.4.2 Sequential Minimal Optimization (SMO)

Sequential Minimal Optimization (SMO) is an efficient algorithm used to solve the complex quadratic programming (QP) optimization problems that arise during the training of SVR. SMO breaks the large optimization problem into the smallest possible sub-problems, which are then solved analytically. The dual optimization problem for SVR solved by SMO is expressed as follows [41].

$$\text{Minimize: } \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\propto_i^* - \propto_i)(\propto_i^* - \propto_i)K(x_i, x_j) + \sum_{i=1}^{N} y_i(\propto_i^* - \propto_i) - \sum_{j=1}^{N} \in (\propto_i^* - \propto_i) \tag{11}$$

Subject to the following constraints

$$\sum_{i=1}^{N}(\alpha_i^* - \alpha_i) = 0, \quad 0 \le \alpha_i \le C,\ i = 1,2,\ldots,N \tag{12}$$

Where $\alpha_i^* - \alpha_i$ represents the Lagrange multipliers optimized through the SMO algorithm, while $y_i$ denotes the actual target value of the *i-th* training data sample. The term $K(x_i, x_j)$ refers to the kernel function used to map data into a higher-dimensional feature space, and $\in$ indicates the acceptable error margin or tolerance level. Furthermore, C serves as the regularization parameter that limits the influence and contribution of each individual data point to the solution. Finally, the linear equality constraint $\sum_{i=1}^{N}(\alpha_i^* - \alpha_i)$ ensures the bias term b is correctly calculated [41].

In this study, Sequential Minimal Optimization (SMO) is employed to solve this quadratic optimization efficiently by breaking it into smaller sub-problems. Based on the experimental testing, the optimal parameters for the SVR model were found to be C = 400, an epsilon value of 0.0001, and a maximum of 1000 iterations.

### 2.5 Genetic Algorithm (GA)

Genetic Algorithm (GA) is a metaheuristic optimization method based on the principles of natural selection and genetics used to find high-quality solutions for complex problems. In this study, GA is utilized to optimize the SVR hyperparameters, specifically the regularization parameter (C), epsilon (ε), and gamma (γ). Each candidate solution is represented as an individual or "chromosome" within a population. The quality of each individual is evaluated using a fitness function, which is defined as the Root Mean Square Error (RMSE) of the SVR prediction on the training data. The objective is to minimize the fitness value to achieve the most accurate prediction model [41]. The fitness function used to evaluate each candidate solution is defined in Equations 13 and 14 [42].

$$\text{fitness} = \sum_{i=1}^{n} f_{S_{target}, S_i} - \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} f_{S_i, S_j} \tag{13}$$

Where n represents the total number of selected features, and y denotes the target feature, which corresponds to the predicted class or output. The terms $f_i$ and $f_j$ refer to the *i-th* and *j-th* features being evaluated, respectively. Additionally, I is defined as the feature relationship score, which quantifies the dependency or mutual information between the variables.

$$f_{S_i, S_j} = MI_{S_i, S_j} + F_{S_i, S_j} + C_{S_i, S_j} \tag{14}$$

Where S represents the combined relationship score between features, which integrates multiple statistical measures. The term F denotes the F-statistic value used to evaluate the significance of the features, while r indicates the Pearson Correlation Coefficient value, which measures the linear strength and direction of the relationship between variables.

### 2.6 Evaluation

The evaluation metrics used in this study include MAE, MAPE, and RMSE to evaluate and compare the prediction accuracy of the developed models. The mathematical formulations for these performance indicators are presented in Equations 15-17.

#### 2.6.1 Mean Absolute Error (MAE)

Mean Absolute Error (MAE) is a metric used to measure the average magnitude of errors in a set of predictions by calculating the mean of the absolute differences between the forecasted values and actual targets. The MAE is computed using the equation 15 [43].

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |A_i - F_i| \tag{15}$$

Where n represents the total number of observations, $A_i$ denotes the actual value (target) for the *i-th* data point, and $F_i$ signifies the predicted or forecasted value for the same point.

#### 2.6.2 Mean Absolute Percentage Error (MAPE)

The Mean Absolute Percentage Error (MAPE) is a measure of prediction accuracy expressed as a percentage that compares the absolute difference between predicted and actual values to the actual observation. This metric indicates the relative error scale and is calculated through equation 16 [44].

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{A_i - F_i}{A_i}\right| \times 100\% \tag{16}$$

Where n is the number of data points, $A_i$ is the actual value, and $F_i$ is the predicted value for each observation i.

### 2.6.3 Root Mean Squared Error (RMSE)

Root Mean Square Error (RMSE) is a standard evaluation metric that measures the average magnitude of error by taking the square root of the average squared difference between predicted and actual observations. RMSE effectively penalizes larger errors and is defined by the formula 17 [45].

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_i - F_i)^2} \tag{17}$$

Where n reprsents the total number of data observations, while $A_i$ and $F_i$ represent the actual and forecasted values at index I, respectively. In this study, RMSE also serves as the fitness function for the Genetic Algorithm to optimize the SVR parameters.

## 3. RESULTS AND ANALYSIS

This section presents the experimental results of the hybrid SVR-GA model for PT stock price forecasting. Aneka Tambang Tbk (ANTM.JK). The analysis focuses on parameter optimization and performance comparison between standard SVR and hybrid SVR-GA models.

### 3.1 Data Preprocessing

The dataset consists of historical stock price data including Open, High, Low, Close (OHLC) prices. Preprocessing involved data cleaning (no missing values detected), lag feature engineering (lag=1), and MinMaxScaler normalization. The dataset was split 80:20 (training:test), resulting in 960 training samples and 241 test samples with 4 input features (previous day's OHLC values) to predict the current day's closing price.

### 3.2 SVR Model Parameter Optimization

A systematic parameter sensitivity analysis was conducted to identify optimal hyperparameters for the SVR model. Four critical parameters were evaluated: iteration count, regularization parameter C, epsilon, and gamma.

**Table 4.** Optimal SVR Parameters

| Parameter | Optimal Value | Range Tested | Performance Impact |
|---|---|---|---|
| Iterations | 1,000 | 100-1,000 | RMSE: 118.12 (100 iter) to 77.73 (1000 iter) |
| C | 400 | 100-500 | Balances complexity and generalization |
| Epsilon | 0.0001 | 0.0001-1.0 | Narrow tolerance for precision |
| Gamma | 0.001 | 0.000001-1.5 | Moderate RBF kernel sensitivity |

Table 4 presents optimal SVR hyperparameters. Iteration testing showed 34.2% RMSE improvement from 100 to 1,000 iterations, indicating substantial iterations are required for convergence.
Iteration count testing revealed significant performance improvement with increased iterations. Performance improved from RMSE 118.12 at 100 iterations to 77.73 at 1,000 iterations, representing a 34.2% reduction. This indicates that SVR optimization requires substantial iterations to converge for this dataset. Parameter C testing across ranges 100-500 identified C=400 as optimal, providing balance between model complexity and generalization. Epsilon testing indicated that a narrow tolerance margin (0.0001) is appropriate for financial forecasting precision. Gamma testing showed that moderate kernel sensitivity (0.001) provides optimal performance, with extreme values degrading accuracy.

### 3.3 Hybrid SVR-GA Model Parameter Optimization

The hybrid SVR-GA model employs a Genetic Algorithm to automatically optimize SVR hyperparameters. Comprehensive testing was performed on both SVR hyperparameters and GA control parameters. Optimal SVR-GA Parameters can be view Table 5.

**Table 5.** Optimal SVR-GA Parameters

| Parameter Type | Parameter | Optimal Value | Range Tested |
|---|---|---|---|
| SVR Hyperparameters | C | 445.32 | 50 – 600 |
| | Epsilon | 0.0078 | 0.00001 – 1.5 |

| Parameter Type | Parameter | Optimal Value | Range Tested |
|---|---|---|---|
| GA Parameters | Gamma | 0.0038 | 0.000001 – 1.5 |
| | Population Size | 50 | 30 – 100 |
| | Generations | 30 | 30 – 150 |
| | Crossover Rate | 0.8 | 0.5 – 0.9 |
| | Mutation Rate | 0.1 | 0.01 – 0.3 |
| | Selection Rate | 0.5 | 0.3 – 0.7 |

Table 5 shows the optimal parameters obtained for the hybrid SVR–GA model. The Genetic Algorithm identifies optimal SVR hyperparameters, with C ranging from 400 to 445.32, epsilon from 0.0001 to 0.0078, and gamma from 0.001 to 0.0038, demonstrating the capability of automated parameter optimization. Population size testing (30, 50, 70, and 100) indicates that 50 individuals provide sufficient genetic diversity. Generation testing shows that 30 generations achieve efficient convergence. A crossover rate of 0.8 supports effective genetic material exchange, while a mutation rate of 0.1 enables adequate exploration without disrupting convergence. In addition, a selection rate of 0.5 maintains a balance between exploration and exploitation.

### 3.4 Model Performance Comparison

Model performance was evaluated using RMSE, MAE, and MAPE on the test set. Table 6 presents the comparative results.

**Table 6.** Model Performance Comparison

| Model | RMSE | MAE | MAPE (%) |
|---|---|---|---|
| SVR (Best Parameters) | 85.48 | 59.02 | 2.62 |
| Hybrid SVR-GA | 75.97 | 52.42 | 2.42 |

Table 6 reports the forecasting performance of the baseline SVR model and the proposed hybrid SVR–GA model on the test set consisting of 241 samples. All metrics are computed on denormalized predictions. The hybrid SVR–GA model consistently outperforms the baseline SVR across all error measures. In particular, RMSE decreases from 85.48 to 75.97 (an improvement of 11.1%), MAE from 59.02 to 52.42 (11.2%), and MAPE from 2.62% to 2.42% (7.6%). These results indicate that GA-based hyperparameter optimization leads to a more accurate forecasting model.

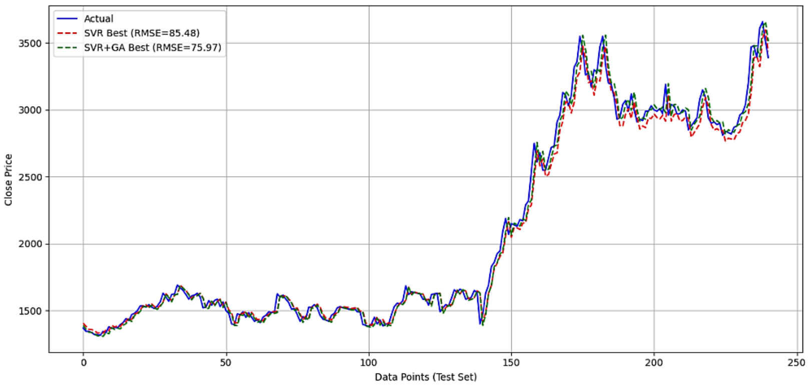


**Figure 2** Prediction Comparison: SVR Best vs SVR-GA Best

Figure 2 compares actual closing prices with the SVR and hybrid SVR–GA model predictions on the test set. Both models follow the general upward and downward trend of the series; however, the SVR–GA trajectory is visibly closer to the actual curve, especially during highly volatile periods with sharp price jumps and drops. Around several local peaks and troughs, the hybrid model more accurately reproduces both the magnitude and timing of price movements, whereas the baseline SVR tends to under-estimate or over-estimate these extremes. The visual evidence in Figure 2 is consistent with the quantitative improvements reported in Table 6.

### 3.5 Residual Analysis

The summary statistics of these residuals are presented in Table 7.

**Table 7.** Residual Statistics on the Test Set

| Model | Mean Residual | Std Residual | Min Residual | Max Residual |
|---|---|---|---|---|
| SVR (Best Parameters) | 26.78 | 81.31 | -232 | 341 |
| Hybrid SVR-GA | 9.44 | 75.50 | -235 | 263 |

Residuals are defined as the difference between actual and predicted prices at each time step. Table 7 summarizes the residual statistics for both models on the test set. The mean residuals of both models are close to zero, indicating no strong systematic over- or under-estimation. However, the hybrid SVR–GA exhibits a smaller mean residual (9.44 vs. 26.78) and a lower standard deviation (75.50 vs. 81.31) than the baseline SVR, which suggests that its prediction errors are less biased and less dispersed. Moreover, the maximum residual is reduced from 341 for SVR to 263 for SVR–GA, implying that large positive deviations from the actual prices occur less frequently in the hybrid model, even though the minimum residuals are of similar magnitude. Overall, the residual statistics confirm that the GA-optimized model produces smaller and more stable errors, in line with the global metrics reported in Table 6 and the visual behavior in Figure 2.

From a methodological perspective, the reduction in error can be attributed to the Genetic Algorithm's role in tuning the SVR hyperparameters. Instead of relying on manually selected or coarsely tuned values, GA performs a global search over a wider and continuous space of $C$, $\epsilon$, and $\gamma$, and selects combinations that provide a better trade-off between model complexity and generalization. Properly chosen $C$ and $\gamma$ allow the RBF kernel to capture nonlinear dynamics in the stock price series, while an appropriate $\epsilon$ controls the model's sensitivity to small fluctuations and noise. As a result, the hybrid SVR–GA model is less prone to underfitting in volatile regions and overfitting in relatively stable segments. This leads to consistently lower RMSE, MAE, and MAPE values, as well as reduced residual dispersion, thereby explaining why the error decreases after applying GA-based optimization.

## 4. CONCLUSION

This study proposes a hybrid SVR-GA model for forecasting the stock price of PT Aneka Tambang Tbk (ANTM.JK), where a Genetic Algorithm is used to optimize SVR hyperparameters and GA control parameters. The experimental results show that the SVR-GA model achieves lower prediction errors than the standard SVR model, with RMSE decreasing from 85.48 to 75.97, MAE from 59.02 to 52.42, and MAPE from 2.62% to 2.42%. These results indicate that GA-based optimization is more effective in adapting SVR parameters to the nonlinear and volatile characteristics of financial time series data, leading to improved generalization performance. Although both models are able to capture the overall trend of stock price movements, the hybrid SVR-GA model demonstrates more consistent predictive accuracy across all evaluation metrics. However, this study is limited to a fixed train-test split and OHLC features; therefore, future research is recommended to apply time series cross-validation, incorporate additional technical indicators (e.g., RSI, MACD, and moving averages), and compare the proposed model with other forecasting methods such as ARIMA or LSTM. Furthermore, future studies could explore comparing GA with other metaheuristic optimizers such as PSO or GWO to further enhance forecasting performance.

## REFERENCES

[1] S. Kumar S, D. Chitradevi, and S. Rajan, "Stock price prediction using deep learning LSTM (long short-term memory)," *2022 2nd Int. Conf. Adv. Comput. Innov. Technol. Eng. ICACITE 2022*, pp. 1787–1791, 2022, doi: 10.1109/ICACITE53722.2022.9823639.

[2] Y. P. Tsang, Y. Fan, Z. P. Feng, and Y. Li, "Examining supply chain vulnerability via an analysis of ESG-Prioritized firms amid the Russian-Ukrainian conflict," *J. Clean. Prod.*, vol. 434, no. May 2023, p. 139754, 2024, doi: 10.1016/j.jclepro.2023.139754.

[3] J. Mahardhika, "Saham Antam (ANTM) Terus Diserok, Target Harga Masih Jauh," Investor.id. Accessed: Sep. 16, 2025. [Online]. Available: https://investor.id/market/408852/saham-antam-antm-terus-diserok-target-harga-masih-jauh

[4] Kustodian Sentral Efek Indonesia, "Statistik Pasar Modal Indonesia – Januari 2025." Accessed: Sep. 18, 2025. [Online]. Available: https://web.ksei.co.id/files/Statistik_Publik_Januari_2025_final.pdf

[5] CNN Indonesia, "Perbandingan Anjlok Parah IHSG: Krisis 1998, Covid-19, dan Hari Ini," CNN Indonesia. Accessed: Sep. 18, 2025. [Online]. Available: https://www.cnnindonesia.com/ekonomi/20250318132738-92-1210179/perbandingan-anjlok-parah-ihsg-krisis-1998-covid-19-dan-hari-ini

[6] Bank Indonesia, "Data Inflasi Indonesia," Bank Indonesia. Accessed: Sep. 17, 2025. [Online]. Available: https://www.bi.go.id/id/statistik/indikator/data-inflasi.aspx?

[7] Bank Indonesia, "BI-Rate Turun 25 BPS Menjadi 5,75%: Mempertahankan Stabilitas, Mendorong Pertumbuhan Ekonomi," Bank Indonesia. Accessed: Sep. 17, 2025. [Online]. Available: https://www.bi.go.id/id/publikasi/ruang-media/news-release/Pages/sp_270825.aspx

[8] Melina, Sukono, H. Napitupulu, and N. Mohamed, "A Conceptual Model of Investment-Risk Prediction in the Stock Market Using Extreme Value Theory with Machine Learning: A Semisystematic Literature Review,"

*Risks*, vol. 11, no. 3, 2023, doi: 10.3390/risks11030060.

[9] L. N. Mintarya, J. N. M. Halim, C. Angie, S. Achmad, and A. Kurniawan, "Machine learning approaches in stock market prediction: A systematic literature review," *Procedia Comput. Sci.*, vol. 216, pp. 96–102, 2022, doi: 10.1016/j.procs.2022.12.115.

[10] P. Soni, Y. Tewari, and D. Krishnan, "Machine Learning Approaches in Stock Price Prediction: A Systematic Review," *J. Phys. Conf. Ser.*, vol. 2161, no. 1, 2022, doi: 10.1088/1742-6596/2161/1/012065.

[11] S. Rifky, W. S. Lestari, I. M. D. M. Sanjaya, and F. M. Sinaga, *Teori dan Penerapan AI di Berbagai Bidang*, no. June. 2024.

[12] A. H. Khan *et al.*, "A performance comparison of machine learning models for stock market prediction with novel investment strategy," *PLoS One*, vol. 18, no. 9 September, pp. 1–19, 2023, doi: 10.1371/journal.pone.0286362.

[13] A. Virigineni, M. Tanuj, A. Mani, and R. Subramani, "Stock Forecasting using HMM and SVR," *Proc. IC3IoT 2022*, 2022, doi: 10.1109/IC3IOT53935.2022.9767969.

[14] H. Oukhouya and K. El Himdi, "Comparing Machine Learning Methods—SVR, XGBoost, LSTM, and MLP—For Forecasting the Moroccan Stock Market," p. 39, 2023, doi: 10.3390/iocma2023-14409.

[15] S. A. Amellia Kharis, A. H. A. Zili, M. Malik, W. Nuryaningrum, and A. Putri, "Comparing machine learning models for Indonesia stock market prediction," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 38, no. 1, pp. 508–516, 2025, doi: 10.11591/ijeecs.v38.i1.pp508-516.

[16] O. A. Montesinos-López, A. M. López, and J. Crossa, *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Springer, 2022. doi: 10.1007/978-3-030-89010-0.

[17] Y. Zhao, G. Zheng, J. Wang, J. Liu, S. Dong, and J. Xin, "Efficient Support Vector Regression for Wideband DOA Estimation Using a Genetic Algorithm," *Sensors*, vol. 25, no. 9, pp. 1–19, 2025, doi: 10.3390/s25092915.

[18] P. Nayak, V. S. Kushwah, J. Parashar, and A. Shukla, "Stock Price Prediction Using LSTM and SVM," *Lect. Notes Networks Syst.*, vol. 971, pp. 291–302, 2024, doi: 10.1007/978-981-97-2089-7_26.

[19] S. R. Thumu and G. Nellore, "Improving Cryptocurrency Price Prediction Accuracy with Multi-Kernel Support Vector Regression Approach," *Int. Res. J. Multidiscip. Technovation*, vol. 6, no. 4, pp. 20–31, 2024, doi: 10.54392/irjmt2443.

[20] S. Ding, Z. Zhang, L. Guo, and Y. Sun, "An optimized twin support vector regression algorithm enhanced by ensemble empirical mode decomposition and gated recurrent unit," *Inf. Sci. (Ny).*, vol. 598, pp. 101–125, 2022, doi: 10.1016/j.ins.2022.03.060.

[21] N. M. Farhan and B. Setiaji, "Optimization by Nature: A Review of Genetic Algorithm Techniques," *Indones. J. Comput. Sci.*, vol. 12, no. 2, pp. 284–301, 2023, [Online]. Available: http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135

[22] U. Singh, K. Saurabh, N. Trehan, R. Vyas, and O. P. Vyas, "GA-LSTM: Performance Optimization of LSTM driven Time Series Forecasting: GA-LSTM: Performance Optimization..: U. Singh et al.," *Comput. Econ.*, vol. 66, no. 0123456789, pp. 1–36, 2024, doi: 10.1007/s10614-024-10769-0.

[23] M. S. Hasan, M. Tarequzzaman, M. Moznuzzaman, and M. A. Ahad Juel, "Prediction of energy consumption in four sectors using support vector regression optimized with genetic algorithm," *Heliyon*, vol. 11, no. 2, p. e41765, 2025, doi: 10.1016/j.heliyon.2025.e41765.

[24] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Glob. Transitions Proc.*, vol. 3, no. 1, pp. 91–99, 2022, doi: 10.1016/j.gltp.2022.04.020.

[25] M. Munawaroh, T. Azhima, Y. Siswa, and W. J. Pranoto, "Optimasi Metode BMR dan Simulated Annealing untuk Algoritma SVM dalam Mengatasi Imbalanced dan High Dimensional Data Stunting," *JPIP*, vol. 5, no. 3, pp. 1–19, 2024, [Online]. Available: https:// ijurnal.com/1/index.php/jpip/article/view/43

[26] N. Huyghues-beaufond, S. Tindemans, P. Falugi, M. Sun, and G. Strbac, "Robust and automatic data cleansing method for short-term load forecasting of distribution feeders," *Appl. Energy*, vol. 261, no. November 2019, p. 114405, 2020, doi: 10.1016/j.apenergy.2019.114405.

[27] A. Kurniadi Hermawan, A. Nugroho, and Edora, "Analisa Data Mining Untuk Prediksi Penyakit Ginjal Kronik Dengan Algoritma Regresi Linier," *Bull. Inf. Technol.*, vol. 4, no. 1, pp. 37–48, 2023, doi: 10.47065/bit.v4i1.475.

[28] M. Buyukkececi and M. C. Okur, "A Comprehensive Review of Feature Selection and Feature Selection Stability in Machine Learning," *Gazi Univ. J. Sci.*, vol. 36, no. 4, pp. 1506–1520, 2023, doi: 10.35378/gujs.993763.

[29] A. Gupta, Akansha, K. Joshi, M. Patel, and V. Pratap, "Visualization and forecasting of stock's closing price using machine learning," *Multimed. Tools Appl.*, vol. 83, no. 29, pp. 72471–72489, 2024, doi: 10.1007/s11042-024-18376-9.

[30] X. Wang, S. Gao, S. Zhou, Y. Guo, Y. Duan, and D. Wu, "Prediction of House Price Index Based on Bagging Integrated WOA-SVR Model," *Math. Probl. Eng.*, vol. 2021, pp. 1–15, 2021, doi: 10.1155/2021/3744320.

[31] E. Katya, "Exploring Feature Engineering Strategies for Improving Predictive Models in Data Science," *Res. J. Comput. Syst. Eng.*, vol. 4, no. 2, pp. 201–215, 2023, doi: 10.52710/rjcse.88.

[32] A. Pradasyah and A. Baita, "Comparative Study of Support Vector Regression and Long Short-Term Memory for Stock Price Prediction," *J. Appl. Informatics Comput.*, vol. 9, no. 4, pp. 1301–1311, 2025.

[33] P. J. M. Ali, "Investigating the Impact of Min-Max Data Normalization on the Regression Performance of K-Nearest Neighbor with Different Similarity Measurements," *ARO-The Sci. J. Koya Univ.*, vol. X, no. 1, pp. 85–91, 2022, doi: 10.14500/aro.10955.

[34] V. R. Joseph, "Optimal ratio for data splitting," *Stat. Anal. Data Min.*, vol. 15, no. 4, pp. 531–538, 2022, doi: 10.1002/sam.11583.

[35] B. Supri, Rudianto, Abdurohim, Badriatul Mawadah, and Helmi Ali, "Asian Stock Index Price Prediction

Analysis Using Comparison of Split Data Training and Data Testing," *JEMSI (Jurnal Ekon. Manajemen, dan Akuntansi)*, vol. 9, no. 4, pp. 1403–1408, 2023, doi: 10.35870/jemsi.v9i4.1339.

[36] T. B. Sianturi, I. Cholissodin, and N. Yudistira, "Penerapan Algoritma Long Short-Term Memory (LSTM) berbasis Multi Fungsi Aktivasi Terbobot dalam Prediksi Harga Ethereum," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 3, pp. 1101–1107, 2023, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12384

[37] Y. Syukriyah and A. Purnama, "Modelling Time Series Data for Stock Prices Prediction Using Bidirectional Long Short-Term Memory," *KNOWBASE Int. J. Knowl. Database*, vol. 04, no. 02, pp. 115–129, 2024, doi: 10.30983/knowbase.v4i2.8759.

[38] W. Shao, X. Lin, Y. Hunag, L. Qu, W. Zhuo, and H. Liu, "Effect of Kernel Functions on the Performance of Support Vector Regression Algorithm in Predicting Patient-Specific Organ Doses from CT Scans," *Radiat. Sci. Technol.*, vol. 11, no. 1, pp. 1–11, 2025, doi: 10.11648/j.rst.20251101.11.

[39] R. Haitham, A. Hussein, and A. Hussein, "Improving the prediction accuracy of a model using SVR with genetic algorithm," *ournal Econ. Adm. Sci.*, vol. 31, no. 145, 2025, doi: 10.33095/tzpewm39.

[40] K. L. Du, B. Jiang, J. Lu, J. Hua, and M. N. S. Swamy, "Exploring Kernel Machines and Support Vector Machines: Principles, Techniques, and Future Directions," *Mathematics*, vol. 12, no. 24, 2024, doi: 10.3390/math12243935.

[41] R. E. Caraka, R. C. Chen, T. Toharudin, M. Tahmid, B. Pardamean, and R. M. Putra, "EVALUATION PERFORMANCE OF SVR GENETIC ALGORITHM AND HYBRID PSO IN RAINFALL FORECASTING," *ICIC Int.*, vol. 11, no. 7, pp. 631–639, 2020, doi: 10.24507/icicelb.11.07.631.

[42] R. R. Abo-Alsabeh and A. Salhi, "The Genetic Algorithm: A study survey," *Iraqi J. Sci.*, vol. 63, no. 3, pp. 1215–1231, 2022, doi: 10.24996/ijs.2022.63.3.27.

[43] W. Zhou, Z. Yan, and L. Zhang, "A comparative study of 11 non-linear regression models highlighting autoencoder, DBN, and SVR, enhanced by SHAP importance analysis in soybean branching prediction," *Sci. Rep.*, vol. 14, no. 1, pp. 1–16, 2024, doi: 10.1038/s41598-024-55243-x.

[44] H. Zhang, Z. Nie, and Y. Lian, "A Semi-Trailer Path Planning Method Considering the Surrounding Traffic Conditions and Vehicle Roll Stability," *Appl. Sci.*, vol. 15, no. 5, 2025, doi: 10.3390/app15052353.

[45] A. Jierula, S. Wang, T. M. Oh, and P. Wang, "Study on accuracy metrics for evaluating the predictions of damage locations in deep piles using artificial neural networks with acoustic emission data," *Appl. Sci.*, vol. 11, no. 5, pp. 1–21, 2021, doi: 10.3390/app11052314.

**BIBLIOGRAPHY OF AUTHORS**

Muhammad Ulil Albab applied to the University of Muhammadiyah East Kalimantan, Department of Computer Engineering, and was accepted.

Taghfirul Azhima Yoga Siswa, S.Kom., M.Kom., is a Lecturer at Universitas Muhammadiyah Kalimantan Timur (UMKT). He holds a Bachelor's degree in Information Systems (S1) and a Master's degree in Computer Science (S2). His areas of expertise include Data Mining, Data Science, Internet of Things (IoT), Big Data, and System Evaluation..

Rofilde Hasudungan, S.Kom., M.Sc, is the Secretary of the International Study Program at Universitas Muhammadiyah Kalimantan Timur (UMKT). He completed his undergraduate studies in Informatics at Universitas Mulawarman in 2009 (S.Kom.) and went on to earn his Master's degree (M.Sc.) from Universiti Malaysia Pahang in 2018. Currently, he is actively involved in academic administration and research, contributing to the development of international study programs at UMKT.