CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 11 (1) (2026), Pages 593-612
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Comparison of ARIMA, Random Forest, and Hybrid ARIMA-Random Forest Models in Forecasting Indonesian Crude Oil Prices

Yeni Rahkmawati[1*], Selvi Annisa[1], Hardianti Hafid[2], Nuramaliyah[3], and Emeylia Safitri[3]

[1] *Department of Statistics, Universitas Lambung Mangkurat, Banjarbaru, Indonesia*
[2] *Department of Statistics, Universitas Negeri Makassar, Makassar, Indonesia*
[3] *Department of Statistics, Universitas Terbuka, Jakarta, Indonesia*

Abstract

The price of Indonesian crude oil (ICP) is highly volatile due to fluctuations in global demand, energy policies, and geopolitical tensions, making accurate forecasting challenging. This study compares three forecasting models: ARIMA, Random Forest, and Hybrid ARIMA–Random Forest. The models are evaluated using Time-Series Cross-Validation (TSCV) with MAPE, sMAPE, and RMSE as performance metrics. The results indicate that the Hybrid ARIMA–Random Forest model achieves the lowest MAPE and sMAPE, while Random Forest attains the lowest RMSE, and ARIMA exhibits the highest forecast errors. Diebold–Mariano (DM) tests confirm that ARIMA's predictive accuracy is significantly lower than both machine-learning-based models, whereas no significant difference is found between Random Forest and the hybrid model. Out-of-sample forecasts for January–June 2026 show relatively stable price movements within 59–63 USD per barrel, with short-term fluctuations reflected in wide prediction intervals. These findings suggest that Indonesian crude oil prices contain both linear and non-linear components, which are effectively captured by the hybrid approach. Overall, the Hybrid ARIMA–Random Forest model provides the most accurate forecasts in percentage-based metrics, offering a robust and reliable tool for policymakers, investors, and market participants navigating volatile oil markets.

Keywords: ARIMA; Forecasting; Hybrid ARIMA-Random Forest; Indonesian Crude Oil Price (ICP); Random Forest.

1. Introduction

The availability and utilization of energy are crucial in achieving sustainable development [1]. Energy resources are fundamental components that support life and serve as important factors of production in economic activities. Among the various types of energy resources, oil, natural gas, and coal are the most commonly utilized sources of energy [2]. The growing global energy demand, driven by industrial development in the 20th century, shifted energy use from biomass to fossil fuels such as coal, oil, and natural gas. The global energy transition that began with the shift from biomass to fossil fuels during the Industrial Revolution is moving towards renewable energy (RE) and emission reductions through increased RE utilization, reduced fossil fuel consumption, and

*Corresponding author. E-mail: yeni.rahkmawati@ulm.ac.id

Submitted: September 18, 2025 Reviewed: October 04, 2025 Accepted: March 01, 2026
DOI: https://doi.org/10.18860/cauchy.v11i1.36540

the adoption of electric vehicles. Although future energy policies prioritize enhancing renewable energy use, projections indicate that fossil fuels, particularly oil and natural gas, will continue to dominate the energy mix for a specific period. Indonesia's primary energy mix projections indicate that oil and natural gas will still play a significant role in meeting national energy demands. The Business as Usual scenario forecasts that the share of oil and natural gas in the total energy demand will reach 49% in 2025 and 39% in 2050, respectively. Meanwhile, the Current Policy scenario forecasts similar figures, at 45% and 44% [3].

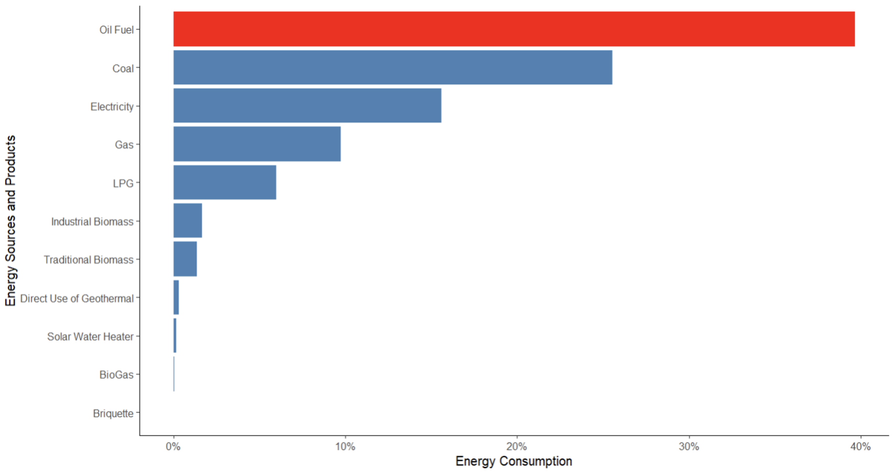

Fig. 1: Final Energy Consumption by Type (Million BOE).

On the energy demand side, final energy consumption increased by 6.29%, equivalent to 1,220 million barrels of oil equivalent (BOE). This increase indicates that energy consumption in 2023 reached its highest point in the last six years, signalling rapid growth in energy demand (HEESI, 2023). Fig. 1 shows that oil fuel dominates energy consumption (39.7%), followed by coal (25.5%) and electricity (15.6%). Other energy sources such as gas (9.7%), LPG (5.9%), industrial biomass (1.6%), and traditional biomass (1.3%) contribute smaller shares. The contribution of other energy sources remains minimal compared to primary sources. The 39.7% share of oil fuel in total energy consumption reflects Indonesia's high dependency on oil-based energy sources [4]. This dependence makes Indonesia vulnerable to fluctuations in global crude oil prices, as crude oil prices are one of the macroeconomic variables that significantly influence the economy.

The international crude oil price applied in Indonesia is the Indonesian Crude Oil Price (ICP), which serves as the base price for crude oil used in calculating and formulating the State Budget (APBN). The ICP represents the average price of Indonesian crude oil in the international market [5]. Global oil prices, which tend to be volatile, highly influence the selling price of Indonesian oil in the international market, known as the ICP. ICP directly correlates with Indonesia's energy export revenues, particularly crude oil. When the ICP increases, Indonesia can generate higher revenue from oil exports. However, at the same time, domestic energy prices, which depend on oil, also rise, affecting production costs and domestic energy consumption [6]. Changes in crude oil prices can significantly impact economic and social development, stability, and even national security in a country [7]. Fluctuations in crude oil prices indicate global economic prospects, currency exchange rate dynamics, and inflation rates. Despite technological advancements, such as electric vehicles and renewable energy sources, reducing dependence on fossil fuels, oil's role as a strategic commodity remains highly relevant in both the global and Indonesian economies [8]. Therefore, it is crucial to design scientific methods to accurately forecast crude oil price movements, helping mitigate the risks of extreme fluctuations in the crude oil market [9].

Previous studies have applied various methods to forecast crude oil prices. Some studies used traditional time series models, such as ARIMA [10] and ARIMAX [11]. Although these models effectively handle linear patterns, difficulties arise when accommodating non-linear data patterns.

[12] developed an ensemble model based on ARIMA to improve forecasting accuracy. Another widely used method to capture non-linear data patterns is the Artificial Neural Network (ANN), as seen in the research by [13] and [14], which used Radial Basis Function Neural Networks (RBF NN). Additionally, studies have applied machine learning-based methods, such as Random Forest (RF) [8] and Fuzzy Time Series (FTS) [15]. However, these studies predominantly focused on using ANN and machine learning techniques separately, without leveraging the strengths of traditional time series models like ARIMA. While ARIMA effectively captures linear components, it is less optimal for non-linear relationships. Various studies have begun combining ARIMA with machine learning algorithms to address this limitation. Researchers have developed ARIMA models combined with Random Forest (RF) and Support Vector Machine (SVM) to handle residuals from ARIMA containing non-linear patterns [16].

A study by [17] proposed a hybrid approach that combines the linear components of ARIMA with non-linear models such as ANN or SVM to improve forecasting accuracy. This hybrid model has proven to be more accurate in forecasting price movements that ARIMA alone cannot explain. Recent research also shows that combining ARIMA with Random Forest and Gradient Boosting can produce a more accurate model by leveraging the strengths of both approaches better to capture both linear and non-linear patterns [18].

Although hybrid approaches have been increasingly adopted, limitations remain in the evaluation design, particularly when dealing with small sample sizes. Splitting the data into training and testing sets may substantially reduce the number of available observations, potentially leading to less reliable parameter estimation and evaluation results. Therefore, a more appropriate evaluation framework is required for time series data, such as time-series cross-validation with a rolling window approach, which preserves the temporal dependency structure and provides more stable and representative performance estimates [19].

Based on this background, the present study not only develops a hybrid ARIMA–Random Forest model to capture both linear and non-linear patterns in crude oil price data, but also implements a time-series cross-validation procedure during the evaluation process and conducts systematic hyperparameter tuning of the Random Forest model to obtain an optimal configuration. Furthermore, the performance of the hybrid model is compared with the individual ARIMA and Random Forest models to determine the most accurate and reliable approach for forecasting future crude oil prices.

2. Methods

This section describes the data, modelling framework, validation strategy, and evaluation criteria used in this study. It begins with the data source and study period, followed by the ARIMA, Random Forest, and Hybrid ARIMA–Random Forest models, as well as the time-series cross-validation procedure and performance metrics employed to assess forecasting accuracy.

2.1. Data

This study is quantitative research that uses secondary time series data. The time series data consists of the monthly average Indonesian crude oil (ICP) prices from August 2017 to December 2025 (USD/Barrel), with 101 observations. We obtained this monthly data from the official website of the Ministry of Energy and Mineral Resources[1].

2.2. Autoregressive Integrated Moving Average (ARIMA) Model

The Autoregressive Integrated Moving Average (ARIMA(p, d, q)) model, also known as the Box–Jenkins model, is one of the classical approaches in time series analysis, first introduced by Box and Jenkins in 1976. This model is an extension of the Autoregressive Moving Average (ARMA(p, q)) model for non-stationary data, where differencing up to order d is applied to make

[1] https://migas.esdm.go.id/

the data stationary in the mean. Thus, ARIMA(p, d, q) combines autoregressive (AR), moving average (MA), and differencing components to handle time series data that exhibit trends or non-stationarity patterns [20].

The differencing process for the first order is defined as:

$$\nabla Y_t = Y_t - Y_{t-1}.$$

The Autoregressive Moving Average (ARMA(p, q)) becomes the ARIMA(p, d, q) when:

$$\nabla^d Y_t = (1 - B)^d Y_t,$$

where B is the backshift operator. In general, the ARIMA model can be expressed as:

$$\phi_p(B)(1 - B)^d Y_t = \theta_q(B) e_t, \tag{1}$$

where p is the order of autoregression, d is the order of differencing, and q is the order of the moving average [21].

We consider a time series data Y_t to be stationary if:

$$E(Y_t) = E(Y_{t-k}) = \mu,$$

which means the expected value of Y remains constant over time;

$$\text{Var}(Y_t) = \text{Var}(Y_{t-k}) = \sigma^2,$$

indicating that the variance of Y stays constant over time; and

$$\text{Cov}(Y_t, Y_s) = \text{Cov}(Y_{t-k}, Y_{s-k}),$$

which shows the covariance of Y for all time points $t_1, t_2, \ldots, t_n$ and for every lag k. Data stationarity is tested by analyzing the Augmented Dickey–Fuller (ADF) test results and considering the lambda (λ) value from the Box–Cox transformation to assess the need for variance stabilization. If the assumption of stationarity in variance is not satisfied, a variable transformation using the Box–Cox transformation is applied. If the data violates the assumption of stationarity in the mean, differencing is applied to correct this [22].

The steps to determine the ARIMA model for time series data are as follows:

1. **Model Identification**. Model identification involves procedures applied to the data to identify an appropriate model for further investigation. At this stage, the goal is to obtain initial values for p, d, and q required in the ARIMA model and to make preliminary estimates for the parameters. We refer to the resulting model as a tentative model. This model identification process can be performed by examining time series data plots, checking ACF plots, checking PACF plots, and conducting formal tests to assess the stationarity of the data.
2. **Parameter Estimation**. In the parameter estimation stage, we estimated the parameters of the identified model using the maximum likelihood method. We apply this step to all candidate models. Additionally, we compute the model selection criteria values. The model with the smallest model selection criterion value determines the best model among all candidate models.
3. **Model Diagnostics**. A model qualifies as suitable when its estimated values closely align with the actual values, ensuring the residuals exhibit white noise properties, meaning the residuals are independent, normally distributed with a mean of zero, and a constant variance. To assess whether the residuals are normally distributed, we can employ the Jarque–Bera test. Additionally, we can examine the residuals using residual plots, ACF residual plots, the Q^* statistic, or the Ljung–Box test [23]. The Ljung–Box test assesses

the independence between lags of the residuals. The null hypothesis tested is that the residuals are independent, while the alternative hypothesis is that the residuals are not independent, with the Q^* statistic defined as:

$$Q^* = n(n+2)\sum_{k=1}^{K}\frac{1}{n-k}r_k^2, \tag{2}$$

where n is the number of residuals; r_k is the ACF value at lag k; and K is the lag length. The statistic Q^* is then compared with the χ^2 value at the degrees of freedom $(K-p-q)$, where p and q are the orders of the ARIMA model [23].

2.3. Random Forest

Random Forest is an ensemble learning method that combines multiple decision trees, each built randomly. Unlike the single regression tree approach, which focuses on optimizing a single predictive model, Random Forest aggregates multiple predictor models, each of which might be suboptimal. Including randomness in constructing each tree allows for a broader exploration of the model space, resulting in empirically superior prediction performance [24].

The general definition of Random Forest according to [25] is as follows: $(\hat{h}(\cdot,\Theta_1),\ldots,\hat{h}(\cdot,\Theta_q))$ represent a collection of tree predictors, with $\Theta_1,\ldots,\Theta_q$ being i.i.d. random variables independent of L_n. The aggregation process in a Random Forest combines the set of random trees to obtain the Random Forest predictor $\hat{h}$. Fig. 2 illustrates this aggregation process.

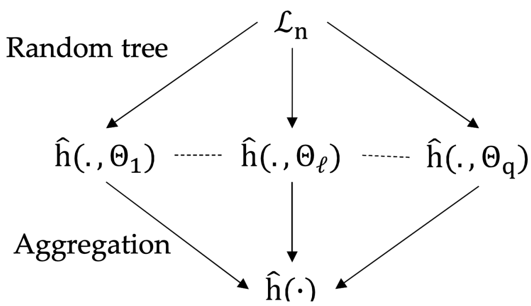

Fig. 2: Random Forest Aggregation Process.

In Random Forest for regression, we calculate the prediction by finding the average prediction from each regression tree, as shown in Eq. (3).

$$\hat{h}(x) = \frac{1}{q}\sum_{l=1}^{q}\hat{h}(x,\Theta_l), \tag{3}$$

where q is the number of regression trees formed and $\hat{h}(x,\Theta_l)$ represents the prediction from the l-th regression tree [8].

2.4. Hybrid ARIMA–Random Forest Model

According to [17], time series data consists of two main components: the linear autocorrelation structure and the non-linear component. A hybrid model that combines linear and non-linear components is used to capture both characteristics more accurately, as shown below:

$$Y_t = L_t + N_t,$$

where L_t represents the linear component and N_t represents the non-linear component. We estimate both components from the data. The data is then input into the first stage of the

ARIMA model to capture the linear component, so the residuals from the linear model will contain non-linear relationships.

If e_t represents the residual component at time t from the linear model, then:

$$e_t = Y_t - \hat{L}_t,$$

where $\hat{L}_t$ represents the forecast value for time t. Residuals play a crucial role in diagnosing the adequacy of the linear model. We consider a linear model inadequate if it still contains a linear correlation structure in the residuals. However, residual analysis cannot detect non-linear patterns in the data. Currently, no general diagnostic statistic is available to identify non-linear autocorrelation relationships. Therefore, even if the diagnostic tests pass, the model may still not be fully adequate because it may not properly model non-linear relationships. Significant non-linear patterns in the residuals indicate limitations in the ARIMA model [17].

The Random Forest (RF) model is used to model the non-linear component present in the residuals of the ARIMA model. Using n input nodes, the Random Forest model for these residuals can be expressed as:

$$e_{tRF} = f_{RF}(e_{t-1}, e_{t-2}, \ldots, e_{t-n}) + \mu_{RF},$$

where f_{RF} is the non-linear function determined by the Random Forest model, and μ_t^{RF} represents the random error component. If we denote the forecast results from Random Forest as $\hat{N}_t^{RF}$, we can express the combined forecast as follows [16]:

$$\hat{F}_{tRF} = \hat{L}_t + \hat{N}_{tRF}.$$

2.5. Time-Series Cross-Validation

Model evaluation in time series data requires special consideration due to the temporal dependence among observations. Unlike cross-sectional data, time series observations follow a chronological order and cannot be randomly permuted without violating the underlying dependency structure. Therefore, conventional k-fold cross-validation with random data partitioning is not appropriate for time series forecasting, as it may lead to data leakage, where future information is unintentionally used during model training, resulting in overly optimistic performance estimates [26], [27].

Time-Series Cross-Validation (TSCV) has been proposed as an alternative evaluation framework that preserves the temporal ordering of the data. The fundamental principle of TSCV is that the model is trained using past observations and validated on subsequent future observations. One commonly used scheme is the rolling-origin evaluation (also known as walk-forward validation), where the training set is progressively expanded (expanding window) or shifted forward with a fixed size (sliding window). This approach allows model performance to be assessed at multiple forecasting origins, providing more stable and representative error estimates compared to random validation procedures [26].

From a theoretical perspective, TSCV aligns with the primary objective of forecasting, namely predicting future values based solely on historical information. Empirical studies have demonstrated that time-series cross-validation produces more reliable estimates of out-of-sample forecasting accuracy and is particularly suitable for model selection and hyperparameter tuning in both statistical and machine learning forecasting models [19], [27].

2.6. Model Evaluation Metrics

Mean Absolute Percentage Error (MAPE) represents the average percentage of the error ratio between the observation at time t (Y_t) and the predicted value at time t ($\hat{Y}_t$), expressed as the following formula:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t} \times 100\right|,$$

The best model is selected based on the smallest MAPE value.

In addition to using MAPE as an accuracy indicator, model performance can also be evaluated using the Root Mean Squared Error (RMSE) to provide an understanding of the absolute deviation in its original units. RMSE quantifies the deviation of predicted values from actual values. The formula to calculate RMSE is:

$$\text{RMSE} = \left(\frac{1}{n} \sum_{t=1}^{n} (Y_t - \hat{Y}_t)^2 \right)^{1/2},$$

where Y_t is the observed value at time t and $\hat{Y}_t$ is the predicted value at time t. The smaller the RMSE value, the better the resulting model [21].

To obtain a comprehensive evaluation of forecasting performance, this study employs several widely used accuracy measures. Each metric captures different aspects of forecast errors, allowing for a more robust comparison among competing models. Specifically, the evaluation criteria include Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and Symmetric Mean Absolute Percentage Error (sMAPE). While MAPE expresses errors in percentage terms and facilitates relative comparison, RMSE emphasizes larger deviations due to the squaring process. Complementarily, sMAPE provides a symmetric percentage-based measure of forecast accuracy, reducing scale dependency and mitigating bias toward large observations by normalizing errors relative to both actual and predicted values.

Symmetric Mean Absolute Percentage Error (sMAPE) is a widely used metric for evaluating forecast accuracy. It measures the relative magnitude of forecast errors by scaling the absolute difference between observed and predicted values by their average. The sMAPE is defined as follows:

$$\text{sMAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)/2}$$

where y_i denotes the observed value, $\hat{y}_i$ represents the forecasted value, and n is the total number of observations [28].

2.7. Crude Oil Price

Crude oil is a naturally occurring hydrocarbon product that, under atmospheric pressure and temperature conditions, exists in liquid or solid form, including asphalt, mineral wax, ozokerite, and bitumen obtained through mining processes. However, it does not include coal or solid hydrocarbon deposits obtained from activities unrelated to the oil and gas industry [29]. Crude oil prices are at the core of the crude oil market, and forecasting these prices becomes a critical factor in managing various industrial sectors worldwide [30]. As a result, this topic has become one of the main focuses of financial research in the oil industry [9].

2.8. Data Analysis Procedures

The data in this study were processed using R programming language version 4.4.2. The analysis steps performed were as follows:

1. Indonesian crude oil monthly average price data from August 2017 to December 2025 were collected from the official source.
2. Data preprocessing was conducted, including checking for missing values, detecting outliers, and transforming the data when necessary to ensure stationarity and suitability for time series modelling.
3. Descriptive statistical analysis and graphical exploration were performed to identify trends, variability, and general price patterns of Indonesian crude oil.
4. The dataset was partitioned into training and testing sets using an 80%–20% splitting scheme. The first 80% of the observations (earliest period) were used to develop and

estimate the forecasting models, while the remaining 20% (most recent period) were reserved for out-of-sample evaluation.

5. The ARIMA model was identified using the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots to determine the appropriate orders of the autoregressive (AR) and moving average (MA) components. Several candidate ARIMA models were then estimated and compared using a time-series cross-validation (TSCV) approach with an expanding window scheme to select the best-performing specification. The best ARIMA model was selected based on the smallest Root Mean Squared Error (RMSE) obtained from the cross-validation procedure. Finally, diagnostic checking was conducted to examine the adequacy of the selected model, including residual independence and white-noise assumptions.
6. The Random Forest model was developed by first determining the input variables based on the Partial Autocorrelation Function (PACF) analysis of the time series data. Significant lags identified from the PACF were used as predictor variables. Subsequently, a time-series cross-validation (TSCV) approach with an expanding window scheme was implemented to evaluate model performance. Hyperparameter tuning was conducted using a grid search procedure to obtain the optimal combination of parameters. The tuning process considered different values of the number of variables randomly sampled at each split ($mtry = 1, 2, 3$) and the number of trees ($ntree = 200, 400, 600$).
   The best Random Forest model was selected based on the smallest Root Mean Squared Error (RMSE) obtained from the cross-validation procedure.
7. A Hybrid ARIMA–Random Forest model was subsequently developed to capture both linear and nonlinear patterns in the data. In the first stage, the selected ARIMA model was fitted to model the linear structure of the series, and the residuals obtained from the ARIMA model were extracted.
   In the second stage, the Random Forest algorithm was applied to model the nonlinear component contained in the ARIMA residuals using the selected lagged variables. Hyperparameter tuning was performed using grid search within a time-series cross-validation (TSCV) framework with an expanding window scheme.
   Finally, the hybrid forecasts were generated by combining the ARIMA forecasts (linear component) with the Random Forest predictions of the residuals (nonlinear component). This modelling framework follows the decomposition structure presented in Eq. (2.4), where the observed series is assumed to consist of additive linear (L_t) and nonlinear (N_t) components. The final hybrid forecast at time t is therefore obtained by summing the estimated linear and nonlinear components.
8. To compare the forecasting performance of the ARIMA, Random Forest, and Hybrid ARIMA–Random Forest models, forecasting accuracy was first evaluated using three performance metrics: Root Mean Squared Error (RMSE), Symmetric Mean Absolute Percentage Error (sMAPE), and Mean Absolute Percentage Error (MAPE). These measures were computed to provide a comprehensive comparison of model performance, and the model with the smallest error values was identified as the best-performing model. Furthermore, the Diebold–Mariano (DM) test was applied to examine whether the differences in forecasting accuracy between competing models were statistically significant.
9. The best-performing model, determined based on the smallest forecasting error values, was selected to generate future forecasts of Indonesian crude oil prices.

The overall analytical procedures and methodological steps of this research are illustrated in the flowchart presented in Fig. 3.

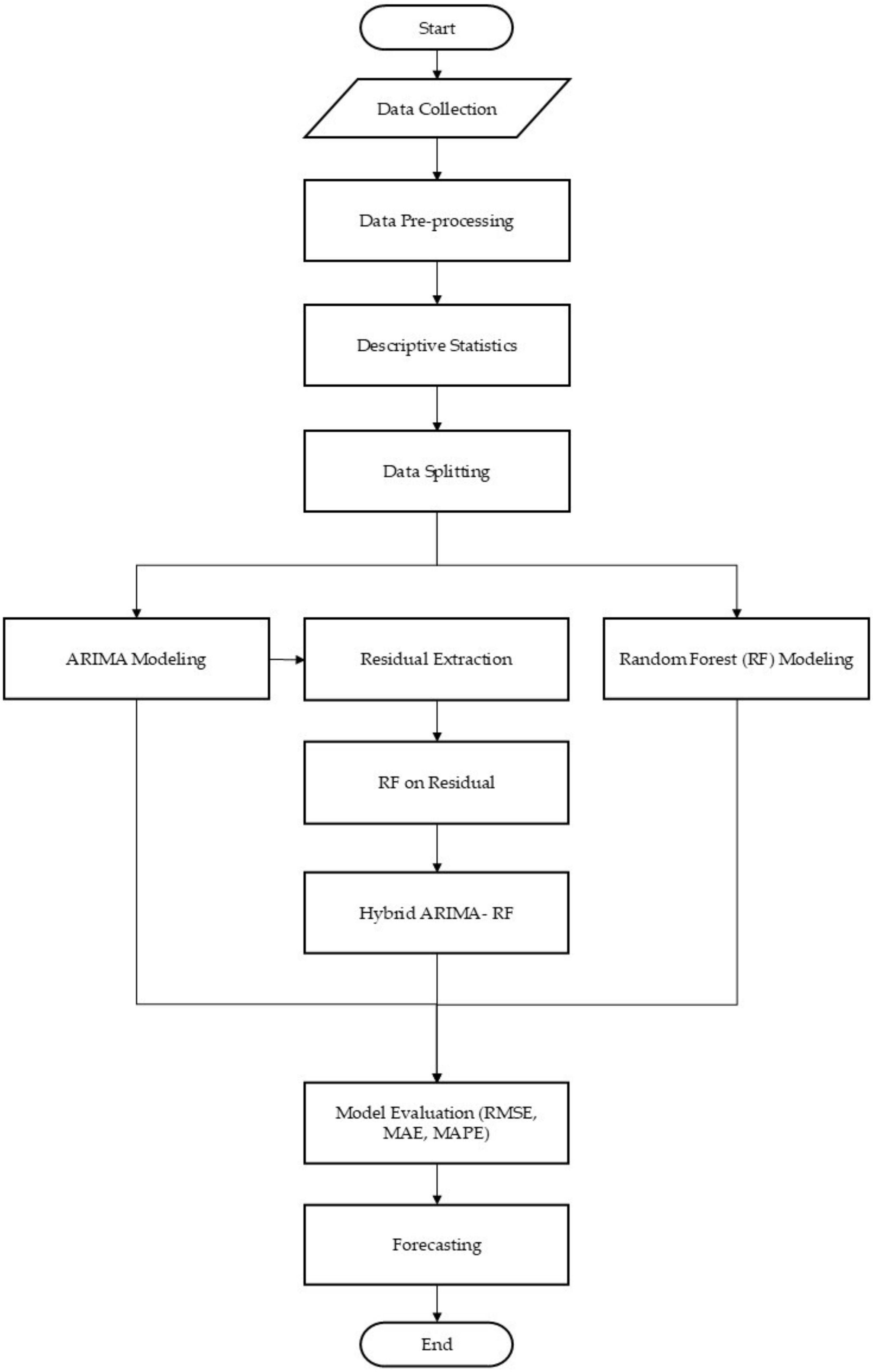

Fig. 3: Research Flowchart of Data Analysis Procedures.

3. Results and Discussion

This section presents the empirical findings of the study and discusses the performance of the proposed forecasting models. It begins with a descriptive overview of the Indonesian crude oil price data, followed by the modelling results, comparative evaluation, and forecasting outcomes.

3.1. Descriptive Statistics

Before conducting further analysis, an important first step is to understand the general characteristics of the data through descriptive statistics analysis. Based on the descriptive statistics results in Table 1, the Indonesian crude oil price (ICP) during the observation period had a minimum value of USD 20.66 and a maximum value of USD 117.62, with an average of USD 69.18 and a standard deviation of 16.93. The very low minimum value occurred in early 2020, when the COVID-19 pandemic caused a sharp decline in global energy demand due to restrictions on industrial and transportation activities [31]. The global geopolitical conflicts, particularly Russia's invasion of Ukraine in 2022, most likely triggered the price spike, disrupting global energy supplies and causing a drastic rise in oil prices [32].

Table 1: Statistic Summary of Indonesian Crude Oil Price

	Min	Max	Mean	Standard Deviation
Crude Oil Price (USD/Barrel)	20.66	117.62	69.18	16.93

This is supported by the ICP pattern presented in Fig. 4. The oil price data from 2017 to 2024 shows significant fluctuations. From 2017 to 2020, the price dropped drastically from around USD 70 per barrel to USD 20 per barrel due to the COVID-19 pandemic. However, the economic recovery post-pandemic and Russia's invasion of Ukraine in 2022 pushed the price up to USD 117.62 per barrel. By the end of 2023, the Indonesian crude oil price began to decrease, continuing until it reached about USD 61.10 per barrel in December 2025.

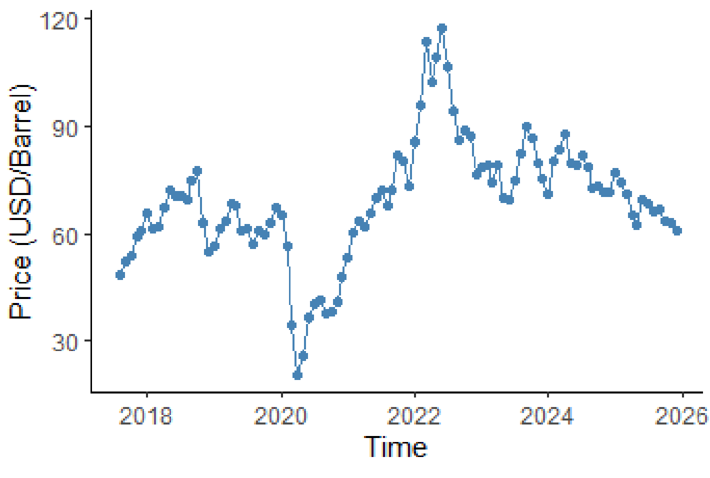

Fig. 4: Plot Time Series of Crude Oil Price (USD/Barrel)

This highly fluctuating data pattern reflects the volatile nature of oil prices, which are sensitive to global economic and political changes. The significant standard deviation indicates that the price variation is relatively high over time. The volatile characteristics of the data make studies related to forecasting crude oil prices highly relevant. This study further implemented forecasting models adapted to non-linear dynamics to provide accurate predictions.

3.2. ARIMA Model

The initial step in ARIMA modelling was to examine the data's stationarity in terms of its mean and variance. We presented the results of the variance stationarity test in Table 2 below.

Table 2: Stationarity Test in Variance

	λ
Box–Cox Lambda	1.117374

Based on the estimation results, the value of $\lambda = 1.117374$ is close to 1, suggesting that no transformation is needed for the crude oil price data because the variance is relatively stable. Next, we performed the mean stationarity test using the Augmented Dickey–Fuller (ADF) Test and presented the results in Table 3.

Table 3: Stationarity Test in Means

Data	Dickey–Fuller Statistic	P-value
Initial data	-2.0063	0.5731
First-order differenced data	-3.8117	0.02267

The results of the ADF test indicated that the original data were not stationary in terms of their mean, as the p-value was greater than 0.05. Therefore, the first differencing ($d = 1$) was performed, and the results showed that the data became stationary after differencing, as indicated by a p-value less than 0.05. The first differencing ($d = 1$) successfully removed the trend in the data, making it stationary and ready for ARIMA identification.

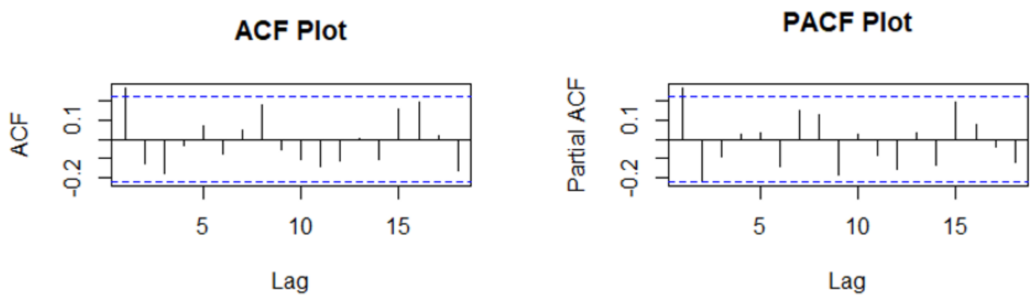

Fig. 5: ACF and PACF Plots

Fig. 5 displays the ACF and PACF plots after first differencing. Both the ACF and PACF show a clear cut-off at lag 1, indicating that the series may follow an ARIMA model with autoregressive and moving average components of order 1. This pattern suggests an initial specification of ARIMA$(1, 1, 1)$, where the differencing order is set to $d = 1$ to achieve stationarity.

To ensure that the selected model provides the best fit and to account for possible overfitting, several alternative specifications were examined by allowing the maximum autoregressive and moving average orders up to $p = 2$ and $q = 2$, respectively, while keeping $d = 1$. Consequently, five candidate models were considered: ARIMA$(1, 1, 1)$, ARIMA$(1, 1, 0)$, ARIMA$(0, 1, 1)$, ARIMA$(2, 1, 1)$, and ARIMA$(1, 1, 2)$.

All candidate models were estimated and evaluated using a time-series cross-validation (TSCV) procedure with an expanding window scheme. In this approach, the training window was progressively expanded while generating one-step-ahead forecasts, and the forecasting errors were computed at each iteration. The model with the smallest Root Mean Squared Error (RMSE) obtained from the TSCV procedure was selected as the best-performing ARIMA specification. The estimation results of the candidate models are presented in Table 4.

Table 4: Parameter Estimates, Significance Tests, and TSCV RMSE of Candidate ARIMA Models

Model	Parameter	Estimate	Std. Error	z value	P-value	RMSE
ARIMA(1,1,1)	ϕ_1	-0.083542	0.307246	-0.2719	0.7857	6.934861
	θ_1	0.404279	0.279327	1.4473	0.1478	
ARIMA(1,1,0)	ϕ_1	0.26670	0.10799	2.4698	0.01352*	6.831552
ARIMA(0,1,1)	θ_1	0.33225	0.10710	3.1022	0.001921*	6.850827
ARIMA(1,1,2)	ϕ_1	0.890261	0.092087	9.6676	< 2.2e-16*	7.127976
	θ_1	-0.605798	0.130772	-4.6325	3.613e-06*	
	θ_2	-0.354066	0.105084	-3.3694	0.0007534*	
ARIMA(2,1,1)	ϕ_1	0.58564	0.39200	1.4940	0.13518	7.044308
	ϕ_2	-0.28211	0.13202	-2.1368	0.03262*	
	θ_2	-0.27698	0.40396	-0.6857	0.49292	

* Significant at 5% level.

Based on the estimation results and the parameter significance tests presented in Table 4, it can be observed that several candidate models have statistically significant parameters at the 5% significance level. In particular, the ARIMA$(1, 1, 0)$, ARIMA$(0, 1, 1)$ and ARIMA$(1, 1, 2)$ models show significant coefficients, while the ARIMA$(1, 1, 1)$ and ARIMA$(2, 1, 1)$ models contain one or more insignificant parameters (p-value > 0.05).

However, model selection was not determined solely based on parameter significance. All candidate models were further evaluated using a time-series cross-validation (TSCV) procedure with an expanding window scheme. Based on the smallest Root Mean Squared Error (RMSE) obtained from the TSCV process, the ARIMA$(1, 1, 0)$ model was selected as the best-performing specification. This indicates that ARIMA$(1, 1, 0)$ provides the most accurate out-of-sample forecasting performance among the candidate models.

Therefore, ARIMA$(1, 1, 0)$ was chosen for subsequent diagnostic checking and further analysis. We conducted two main diagnostic tests on the residuals of the selected model, as presented in Table 5: the Ljung–Box test to evaluate the presence of residual autocorrelation and the Jarque–Bera test to assess the normality of the residuals.

Table 5: Model Diagnostic Test

Model	P-value	
	Ljung–Box Test	Jarque–Bera Test
ARIMA(1,1,0)	0.1298	0.05211

Based on Table 5, the ARIMA$(1, 1, 0)$ model shows Ljung–Box and Jarque–Bera p-values of 0.1298 and 0.05211, respectively. Since both p-values are greater than the 5% significance level, the residuals do not exhibit significant autocorrelation and can be considered approximately normally distributed. These results indicate that the ARIMA$(1, 1, 0)$ model satisfies the classical diagnostic assumptions.

Therefore, based on the residual diagnostic tests, the ARIMA$(1, 1, 0)$ model is considered adequate and appropriate for forecasting purposes.

The ARIMA(1,1,0) model, derived from Eq. (1), can be written as follows:

$$Y_t = (1 + \phi_1)Y_{t-1} - \phi_1 Y_{t-2} + e_t,$$

$$Y_t = 1.2667Y_{t-1} - 0.2667Y_{t-2} + e_t.$$

We evaluated the ARIMA(1,1,0) model using MAPE, sMAPE and RMSE to measure the prediction accuracy, allowing for comparison with other models.

3.3. Random Forest Model

We completed the ARIMA model identification in the previous stage, and the next step is to build an alternative model using the Random Forest algorithm. Random Forest is an ensemble learning method based on decision trees, capable of capturing non-linear patterns in time series data. The main advantage of this algorithm lies in its ability to model complex relationships without satisfying classical statistical assumptions such as linearity, normality, or homoscedasticity, which are typically prerequisites in traditional parametric models.

Research by [8] proves that the Random Forest algorithm provides accurate forecasting results for Indonesian crude oil prices. The modelling in this study began with an analysis of the Partial Autocorrelation Function (PACF) to identify significant lags as input variables.

Fig. 6 presents the PACF plot of the Indonesian crude oil price data. Based on the significance bounds shown in the figure, lag 1, lag 2, and lag 16 exceed the confidence limits, indicating that these lags have a statistically significant partial autocorrelation with the current value. Therefore, these three lags were selected as candidate input variables for the Random Forest model.

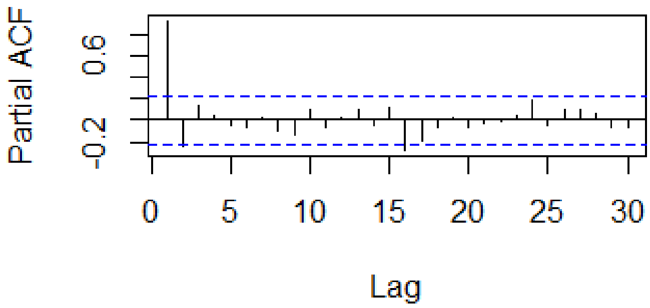

Fig. 6: PACF Plot of Indonesian Crude Oil Prices

Hyperparameter tuning of the Random Forest model was conducted using a grid search approach within a time-series cross-validation (TSCV) framework with an expanding window scheme. The tuning process evaluated different combinations of the number of variables randomly selected at each split (*mtry*) and the number of trees (*ntree*). Specifically, *mtry* values of 1, 2, and 3 were combined with *ntree* values of 200, 400, and 600, resulting in nine candidate model specifications. Each combination was assessed using the Root Mean Squared Error (RMSE) obtained from the TSCV procedure.

The complete tuning results are presented in Table 6.

Table 6: Random Forest Hyperparameter Tuning Results Based on TSCV

mtry	*ntree*	RMSE
1	200	7.166064
2	**200**	**6.798245**
3	200	7.140064
1	400	7.149043
2	400	6.878749
3	400	7.194258
1	600	7.111057
2	600	6.890961
3	600	7.164297

The results indicate that the smallest RMSE value was achieved when $mtry = 2$ and $ntree = 200$, suggesting that selecting two predictor variables at each split provides the most accurate forecasting performance. Although three lag variables were identified as significant from the PACF analysis, the tuning results show that considering two predictors at each node

($mtry = 2$) yields better predictive accuracy than using only one predictor or all available predictors. This finding implies that a moderate level of model complexity is sufficient to capture the underlying non-linear dynamics of the series without introducing unnecessary variability.

The best-performing Random Forest model was subsequently evaluated on the test dataset to assess its out-of-sample forecasting ability. The evaluation was conducted using MAPE, sMAPE, and RMSE as performance metrics. These measures provide complementary information regarding the model's predictive accuracy, allowing a comprehensive comparison with the ARIMA model developed in the previous stage.

3.4. Hybrid ARIMA–Random Forest Model

Based on the previous analysis, the best ARIMA model selected for the data was ARIMA(1,1,0). After obtaining the optimal ARIMA specification, the residuals from this model were extracted to capture the remaining non-linear patterns that could not be explained by the linear ARIMA structure. These residuals were then modelled using the Random Forest algorithm to construct the hybrid ARIMA–RF model.

Hyperparameter tuning for the Random Forest model on the ARIMA residuals was conducted using a grid search approach within a time-series cross-validation (TSCV) framework employing an expanding window scheme. This approach ensures that the temporal order of the data is preserved during validation, preventing information leakage from future observations. The tuning process evaluated combinations of the number of variables randomly selected at each split ($mtry$) and the number of trees ($ntree$). Specifically, $mtry$ values of 1, 2, and 3 were combined with $ntree$ values of 200, 400, and 600, resulting in nine candidate models. Each specification was evaluated using the Root Mean Squared Error (RMSE) derived from the TSCV procedure.

The complete tuning results are presented in Table 7.

Table 7: Random Forest Hyperparameter Tuning on ARIMA Residuals Based on TSCV

mtry	*ntree*	RMSE
1	200	6.696967
2	**200**	**5.909922**
3	200	6.717310
1	400	6.968369
2	400	5.972090
3	400	6.965884
1	600	6.741416
2	600	6.193564
3	600	6.832245

The results indicate that the smallest RMSE value was achieved when $mtry = 2$ and $ntree = 200$. This finding suggests that considering two predictors at each split provides better predictive accuracy in modeling the non-linear structure present in the ARIMA residuals. The result implies that a moderate model complexity is more effective than a very simple configuration in capturing the remaining non-linear patterns after the linear ARIMA component has been removed.

After selecting the optimal Random Forest model, the hybrid forecasting process was performed by combining the linear forecasts obtained from the ARIMA(1,1,0) model with the predicted residuals generated by the Random Forest model. Mathematically, the hybrid forecast can be expressed as:

$$\hat{Y}_t^{Hybrid} = \hat{Y}_t^{ARIMA} + \hat{e}_t^{RF},$$

where $\hat{Y}_t^{ARIMA}$ represents the linear forecast from the ARIMA model and $\hat{e}_t^{RF}$ denotes the predicted residual from the Random Forest model.

Finally, the performance of the hybrid ARIMA–RF model was evaluated on the test dataset using MAPE, sMAPE, and RMSE. These evaluation metrics provide a comprehensive assessment of forecasting accuracy and enable comparison with the single ARIMA and Random Forest models.

3.5. Evaluation Model

We evaluated the model performance by comparing the prediction results from the ARIMA, Random Forest, and Hybrid ARIMA–Random Forest models using three commonly used evaluation metrics in forecasting: Mean Absolute Percentage Error (MAPE), Symmetric Mean Absolute Percentage Error (sMAPE), and Root Mean Squared Error (RMSE). MAPE measures the average percentage error between the observed and predicted values, sMAPE measures the average symmetric percentage error between the observed and predicted values, providing a scale-independent evaluation of forecasting accuracy while reducing the bias toward large actual values, and RMSE provides information on the dispersion of prediction errors with greater penalty on large deviations. Table 8 presents the comparison results on the test dataset.

Table 8: Model Comparison on Test Data

Model	sMAPE (%)	MAPE (%)	RMSE
ARIMA(1,1,0)	17.8535	20.1313	15.1343
Random Forest	5.3286	5.3948	**4.4779*** **
Hybrid ARIMA–Random Forest	**5.2200***	**5.2621****	4.5476

* Smallest sMAPE, ** Smallest MAPE, *** Smallest RMSE.

Based on Table 8, the ARIMA(1,1,0) model produced substantially larger forecasting errors (sMAPE = 17.85%, MAPE = 20.13%, RMSE = 15.13), indicating that the linear ARIMA specification was insufficient to capture the underlying data dynamics.

In contrast, both machine learning-based models significantly improved predictive accuracy. The Hybrid ARIMA–Random Forest model achieved the smallest percentage-based errors (sMAPE = 5.22%, MAPE = 5.26%), while the Random Forest model yielded the lowest RMSE (4.48), suggesting slightly better performance in minimizing large deviations.

Although the Hybrid model marginally outperformed in percentage-based measures, the overall difference between the Hybrid and Random Forest models is small, indicating comparable predictive capability. These findings suggest that the dominant structure of the series is primarily non-linear, allowing the Random Forest model alone to effectively capture the underlying dependencies without requiring additional linear decomposition.

To formally assess whether the observed differences in forecasting performance between the competing models are statistically significant, the Diebold–Mariano (DM) test was employed. This test evaluates the null hypothesis of equal predictive accuracy between two forecasting models. The results of the pairwise comparisons are presented in Table 9.

Table 9: Diebold–Mariano Test Results for Forecast Accuracy Comparison

Model Comparison	DM Statistic	p-value
Random Forest vs Hybrid ARIMA–RF	-0.0894	0.9297
ARIMA vs Hybrid ARIMA–RF	5.0276	**0.0000645***
ARIMA vs Random Forest	5.2392	**0.0000398***

* Significant at the 5% level.

The comparison between the Random Forest and Hybrid models yielded a p-value of 0.9297, indicating no statistically significant difference in predictive accuracy between the two approaches. In contrast, both the Random Forest and Hybrid models significantly outperformed the ARIMA model (p-value < 0.001). These results suggest that incorporating non-linear modelling techniques provides a statistically significant improvement over the linear ARIMA framework.

Given that the Hybrid ARIMA–Random Forest model achieved the lowest percentage-based error measures (sMAPE and MAPE) and performed comparably to the Random Forest model in terms of RMSE, the Hybrid model was selected as the final forecasting model. Although the Diebold–Mariano test indicates that the difference between the Hybrid and Random Forest models is not statistically significant, the Hybrid model demonstrates slightly more consistent performance across multiple evaluation metrics.

Subsequently, forecasts of Indonesian crude oil prices for the period January 2026 to June 2026 were generated using the Hybrid ARIMA–Random Forest approach. The comparable performance between the Hybrid and Random Forest models suggests that non-linear modelling plays a dominant role in capturing the data dynamics. However, the marginal improvement observed in the Hybrid model indicates that incorporating the linear structure through ARIMA decomposition may still provide complementary information. Hybrid models are particularly advantageous when a time series contains both linear and non-linear components that interact in a complex manner. In this case, the results suggest that while non-linear patterns are prominent, the linear structure captured by ARIMA contributes additional refinement, leading to slightly improved percentage-based forecasting accuracy. Therefore, the Hybrid ARIMA–Random Forest model represents a balanced and robust forecasting approach for Indonesian crude oil prices.

3.6. Forecasting

The graph in Fig. 7 presents the forecasting results of Indonesian crude oil prices for the period January to June 2026. The blue line represents the historical observations from August 2017 to December 2025, while the red line shows the forecasts generated by the Hybrid ARIMA–Random Forest model. The shaded area illustrates the 95% prediction interval, reflecting the uncertainty associated with the forecasts.

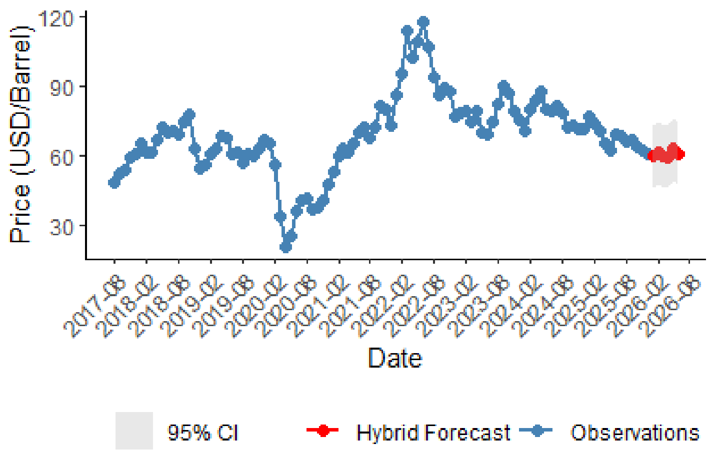

Fig. 7: Forecasting Indonesian Crude Oil Prices (January–June 2026).

The last observed crude oil price in December 2025 was 61.10 USD per barrel. The forecast indicates a moderate increase to 60.04 USD in January 2026, followed by a further rise to 61.65 USD in February. Prices are then projected to slightly decline to 60.42 USD in March and 59.50 USD in April. A notable increase is expected in May, reaching 63.17 USD, before stabilizing at 61.11 USD in June 2026. Overall, the projected pattern suggests short-term fluctuations rather than a strong upward or downward trend. The forecasts indicate relatively stable price movements within the range of approximately 59 to 63 USD per barrel during the first half of 2026.

The 95% prediction intervals provide additional insight into forecast uncertainty. For January

2026, the interval ranges from 46.63 to 73.45 USD per barrel, while for June 2026 it ranges from 47.70 to 74.52 USD per barrel. The relatively wide intervals reflect the inherent volatility of crude oil prices and the uncertainty in global supply-demand conditions.

Although the model is purely data-driven and does not explicitly incorporate exogenous economic variables, the projected short-term fluctuations may be associated with broader global market conditions. For instance, concerns over potential increases in U.S. trade tariffs could disrupt global economic activity and reduce oil demand. Additionally, reduced refinery operations in the U.S. and Europe during routine maintenance periods may exert downward pressure on prices. Seasonal consumption patterns, particularly in preparation for the summer holiday season, may also influence short-term demand dynamics [33].

The projected trajectory suggests relatively stable price dynamics with short-term adjustments rather than a pronounced upward or downward trend. These movements likely capture ongoing adjustments in global supply–demand balances, seasonal consumption patterns, and broader international economic conditions. However, the wide confidence intervals highlight substantial forecast uncertainty, indicating that actual prices remain vulnerable to external shocks, geopolitical tensions, and shifts in global economic activity.

Therefore, although the Hybrid ARIMA–Random Forest model anticipates short-term stabilization, policymakers and market participants should interpret the forecasts cautiously and account for the considerable uncertainty surrounding future price movements when formulating strategic decisions.

4. Conclusion

This study compared the forecasting performance of ARIMA, Random Forest, and Hybrid ARIMA–Random Forest models in predicting Indonesian crude oil prices (ICP) using Time-Series Cross-Validation (TSCV) with MAPE, sMAPE, and RMSE as evaluation metrics. The results indicate that the Hybrid ARIMA–Random Forest model outperforms the competing models in terms of MAPE and sMAPE, while the Random Forest model achieves the lowest RMSE. In contrast, the ARIMA(1,1,0) model exhibits the highest forecast errors across all metrics, indicating its limited ability to capture the complexity of the underlying data patterns. The Diebold–Mariano (DM) test further reveals that the predictive accuracy of the ARIMA model is significantly lower than that of both machine-learning-based models, whereas no statistically significant difference is found between the Random Forest and Hybrid ARIMA–Random Forest models.

Out-of-sample forecasts for the January–June 2026 period indicate relatively stable price movements within the range of 59–63 USD per barrel, characterized by short-term fluctuations and relatively wide prediction intervals. Overall, these findings confirm that machine-learning-based approaches, namely Random Forest and Hybrid ARIMA–Random Forest, are more effective than conventional ARIMA in modeling the dynamics of Indonesian crude oil prices, although their relative superiority depends on the evaluation metric considered.

CRediT Authorship Contribution Statement

Yeni Rahkmawati: Conceptualization, Methodology, Writing–Original Draft. **Selvi Annisa**: Data Curation, Formal Analysis, Writing–Review & Editing. **Hardianti Hafid**: Software, Validation, Visualization. **Nuramaliyah**: Supervision, Project Administration, Funding Acquisition. **Emeylia Safitri**: Data Collection, Investigation, Writing–Review & Editing.

Declaration of Generative AI and AI-assisted technologies

Generative AI and AI-assisted technologies were used during the preparation of this manuscript. Specifically, ChatGPT (OpenAI, GPT-5, 2025) was employed to assist in language polishing, improving clarity of sentences, and checking consistency in formatting. The final responsibility for the content rests fully with the authors, and all interpretations, analyses, and conclusions were made by the authors without reliance on AI tools.

Declaration of Competing Interest

The authors declare no competing interests.

Funding and Acknowledgments

This research received no external funding.

Data and Code Availability

The Indonesian Crude Oil Price (ICP) data were obtained from the official website of the Directorate General of Oil and Gas, Ministry of Energy and Mineral Resources (ESDM), Republic of Indonesia: https://migas.esdm.go.id/post/harga-minyak-mentah.

The R scripts for ARIMA, Random Forest, Hybrid ARIMA–Random Forest, Time-Series Cross-Validation (TSCV), and the Diebold–Mariano test are available from the corresponding author upon reasonable request. Code are provided for academic purposes subject to institutional data policies.

References

[1] H. Khan, I. Khan, and T. T. Binh, "The heterogeneity of renewable energy consumption, carbon emission and financial development in the globe: A panel quantile regression approach," *Energy Reports*, vol. 6, pp. 859–867, 2020. DOI: 10.1016/j.egyr.2020.04.002.

[2] L. F. Arifah, M. Basorudin, M. A. Majid, and M. Choirunnisa, "Studi empiris pengaruh harga minyak mentah dunia dan variabel moneter terhadap perekonomian indonesia periode 1996-2018," *Jurnal Ekonomi-Qu*, vol. 10, no. 1, pp. 23–44, 2020. DOI: 10.35448/jequ.v10i1.8577. http://jurnal.untirta.ac.id/index.php/Ekonomi-Qu.

[3] A. E. Setyono and B. F. T. Kiono, "Dari energi fosil menuju energi terbarukan: Potret kondisi minyak dan gas bumi indonesia tahun 2020–2050," *Jurnal Energi Baru dan Terbarukan*, vol. 2, no. 3, pp. 154–162, 2021. DOI: 10.14710/jebt.2021.11157.

[4] K. ESDM, A. W. Kencono, M. Dwinugroho, E. S. Baruna, and N. Ajiwihanto, *Handbook Of Energy & Economic Statistics Of Indonesia 2015*. 2015, p. 73.

[5] C. Tjahjaprijadi, *Dampak penurunan indonesian crude oil price terhadap pertumbuhan ekonomi indonesia*, 2015. https://fiskal.kemenkeu.go.id/kajian/2015/12/31/145740273503251-dampak-penurunan-indonesian-crude-oil-price-terhadap-pertumbuhan-ekonomi-indonesia.

[6] D. Suryani, M. Fadhilla, and A. Labellapansa, "Indonesian crude oil price (icp) prediction using multiple linear regression algorithm," *J. RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 6, pp. 1057–1063, 2022. DOI: 10.29207/resti.v6i6.4590.

[7] G. Wu and Y. J. Zhang, "Does china factor matter? an econometric analysis of international crude oil prices," *Energy Policy*, vol. 72, pp. 78–86, 2014. DOI: 10.1016/j.enpol.2014.04.026.

[8] S. Annisa et al., "Peramalan harga minyak mentah indonesia menggunakan algoritma random forest," *Jurnal Gaussian*, vol. 13, pp. 472–479, 2025. DOI: 10.14710/j.gauss.13.2.472-479.

[9] J. L. Zhang, Y. J. Zhang, and L. Zhang, "A novel hybrid method for crude oil price forecasting," *Energy Economics*, vol. 49, pp. 649–659, 2014. DOI: 10.1016/j.eneco.2015.02.018.

[10] D. E. Krislianti, E. Zukhronah, and Y. Susanti, "Peramalan harga minyak menggunakan autoregressive integrated moving average dan support vector regression," in *Prosiding Seminar Pendidikan Matematika dan Matematika*, vol. 7, 2023. DOI: 10.21831/pspmm.v7i1.302.

[11] I. F. Amri, A. Wulandari, K. N. Abidah, A. C. Irawan, and M. Al Haris, "Pemodelan arimax untuk meramalkan harga minyak mentah dunia," *Square Journal of Mathematics and Mathematics Education*, vol. 5, no. 1, pp. 47–58, 2023. DOI: 10.21580/square.2023.5.1.17074.

[12] E. Setiyowati, A. Rusgiyono, and T. Tarno, "Model kombinasi arima dalam peramalan harga minyak mentah dunia," *Jurnal Gaussian*, vol. 7, no. 1, pp. 54–63, 2018. DOI: 10.14710/j.gauss.v7i1.26635.

[13] J. Veri, S. Surmayanti, and G. Guslendra, "Prediksi harga minyak mentah menggunakan jaringan syaraf tiruan," *MATRIK Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 3, pp. 503–512, 2022. DOI: 10.30812/matrik.v21i3.1382.

[14] R. A. Fauzannissa, H. Yasin, and D. Ispriyanti, "Peramalan harga minyak mentah dunia menggunakan metode radial basis function neural network," *Jurnal Gaussian*, vol. 5, no. 1, pp. 193–202, 2015. DOI: 10.14710/j.gauss.5.1.193-202. https://ejournal3.undip.ac.id/index.php/gaussian/article/view/11049.

[15] M. A. Ramadhani, B. H. Mustawinar, D. R. Arifanti, and Y. Yuliani, "Prediksi harga minyak dunia dengan fuzzy time series," *Proximal Jurnal Penelitian Matematika dan Pendidikan Matematika*, vol. 7, no. 1, pp. 305–309, 2024. DOI: 10.30605/proximal.v7i1.3471.

[16] M. Kumar and M. Thenmozhi, "Forecasting stock index returns using arima-svm, arima-ann, and arima-random forest hybrid models," *International Journal of Banking, Accounting and Finance*, vol. 5, no. 3, pp. 284–308, 2014. DOI: 10.1504/IJBAAF.2014.064307.

[17] P. G. Zhang, "Time series forecasting using a hybrid arima and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003. DOI: 10.1016/S0925-2312(01)00702-0.

[18] S. Bhuvaneshwari and S. N. S. Rajini, "A novel hybrid model for stock price forecasting: Combining arima, random forests, and gradient boosting techniques," *Library of Progress-Library Science, Information Technology & Computer*, vol. 44, no. 2, pp. 2027–2034, 2024. DOI: 10.48165/bapas.2024.44.2.1.

[19] W. Sulandari, Y. Yudhanto, S. Subanti, E. Zukhronah, and M. Z. Subarkah, "Implementing time series cross validation to evaluate the forecasting model performance," *KnE Life Sciences*, vol. 8, no. 1, pp. 229–238, 2024. DOI: 10.18502/kls.v8i1.15584.

[20] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th. New Jersey: John Wiley & Sons, 2016.

[21] D. C. Montgomery, C. L. Jennings, and M. Kulahci, *Introduction to Time Series Analysis and Forecasting*, 3rd. New Jersey: John Wiley & Sons, 2008.

[22] W. W. S. Wei, *Time Series Analysis: Univariate and Multivariate Methods*, 2nd. United States of America: Pearson Education, Inc., 2006. http://linkinghub.elsevier.com/retrieve/pii/016920709190015N.

[23] J. D. Cryer and K.-S. Chan, *Time Series Analysis: With Applications in R*. Springer, 2008. DOI: 10.1007/978-0-387-75959-3.

[24] R. Genuer and J.-M. Poggi, *Random Forests with R*. Springer, 2020. DOI: 10.1007/978-3-030-56485-8_3.

[25] L. Breiman, "Random forest," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[26] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd. Melbourne, Australia: OTexts, 2018.

[27] C. Bergmeir, R. J. Hyndman, and J. M. Benítez, "Bagging exponential smoothing methods using stl decomposition and box–cox transformation," *International Journal of Forecasting*, vol. 34, no. 2, pp. 303–312, 2018. DOI: 10.1016/j.ijforecast.2017.11.002.

[28] V. Cerqueira, L. Roque, and C. Soares, "Modelradar: Aspect-based forecast evaluation," *Machine Learning*, pp. 114–229, 2025. DOI: 10.1007/s10994-025-06877-z.

[29] Republik Indonesia, *Undang undang republik indonesia nomor 11 tahun 2020 tentang cipta kerja*, https://peraturan.bpk.go.id/, 2020.

[30] H. Shin, T. Hou, K. Park, C. K. Park, and S. Choi, "Prediction of movement direction in crude oil prices based on semi-supervised learning," *Decision Support Systems*, vol. 55, no. 1, pp. 348–358, 2013. DOI: 10.1016/j.dss.2012.11.009.

[31] A. Sugiyono, J. Santosa, Adiarso, and E. Hilmawan, "Pemodelan dampak covid-19 terhadap kebutuhan energi di indonesia," *Jurnal Sistem Cerdas*, vol. 3, no. 2, pp. 65–73, 2020. DOI: 10.37396/jsc.v3i2.65.

[32] Ministry of Finance of the Republic of Indonesia, *Macroeconomic framework and principles of fiscal policy in 2023*, 2023. https://fiskal.kemenkeu.go.id/files/kemppkf/file/1684478331_kem_ppkf_2023.pdf.

[33] Kementerian Energi dan Sumber Daya Mineral Republik Indonesia, *Icp maret turun menjadi usd71,11/barel, dampak peningkatan tarif perdagangan as*, 2023. https://esdm.go.id/id/media-center/arsip-berita/icp-maret-turun-menjadi-usd7111-barel-dampak-peningkatan-tarif-perdagangan-as.