JURNAL ABSIS
Jurnal Pendidikan Matematika dan Matematika
Vol. 8 No. 2, August 2025

EVALUATION OF THE PERFORMANCE OF TBATS AND SARIMA METHODS IN FORECASTING AIR TEMPERATURE IN INDONESIA

Ananda Aprilia[1], Madona Yunita Wijaya[2], Dhea Urfina Zulkifli[3],
Irma Fauziah[4], Nur Inayah[5]
[1-5] Universitas Islam Negeri Syarif Hidayatullah Jakarta, Indonesia
anandaaprilia684@gmail.com

ABSTRACT Indonesia is a tropical archipelago located along the equator, where air temperature patterns exhibit seasonal trends and unstable fluctuations. This instability can impact several sectors, including agriculture—making it difficult for farmers to determine planting and harvesting times—electricity demand, which increases during hotter periods, and public health, as erratic weather may reduce productivity and elevate the risk of diseases such as dehydration, asthma, and respiratory infections. This study aims to evaluate the performance of the Trigonometric, Box-Cox Transformation, ARMA Errors, Trend, and Seasonal (TBATS) model and the Seasonal Autoregressive Integrated Moving Average (SARIMA) model in forecasting air temperature in Indonesia. The dataset used comprises 2-meter air temperature records in Indonesia from January 1940 to August 2024, obtained from ECMWF. The evaluation method applied is cross-validation with a rolling basis. The results show that the RMSE for the TBATS model is 21.3843%, while the SARIMA model has an RMSE of 21.2958%. These results indicate that SARIMA has a slightly better performance than TBATS. However, both methods perform well in forecasting air temperature in Indonesia, as their RMSE percentages are within an acceptable range. This research is expected to contribute to the scientific literature on air temperature forecasting in Indonesia and encourage further studies on hybrid models that integrate TBATS and SARIMA. Additionally, it may support efforts to mitigate the adverse impacts of air temperature changes in the country.

Keywords: air temperature forecasting, tbats model, sarima model, model performance evaluation

ABSTRAK Indonesia is a tropical archipelagic country located along the equator, where air temperature patterns are characterized by seasonal trends and unstable fluctuations. This instability affects several sectors, such as agriculture—where farmers struggle to determine planting and harvesting times—energy, with increased electricity demand during high temperatures, and public health, as fluctuating weather conditions may reduce productivity and increase the risk of illnesses such as dehydration, asthma, and respiratory infections. This study aims to evaluate the performance of two forecasting methods: Trigonometric,

https://doi.org/10.30606/absis.v8i2.3052

Box-Cox Transformation, ARMA Errors, Trend, and Seasonal components (TBATS) and the Seasonal Autoregressive Integrated Moving Average (SARIMA) model in predicting air temperature in Indonesia. The dataset consists of 2-meter air temperature records in Indonesia from January 1940 to August 2024, sourced from ECMWF. The evaluation method employed is cross-validation using a rolling basis approach. The results show that the RMSE for TBATS is 21.3843%, while the RMSE for SARIMA is 21.2958%. Although SARIMA yields a slightly lower RMSE, indicating marginally better performance, both models demonstrate adequate accuracy in forecasting Indonesia's air temperature. This study contributes to the existing literature on climate forecasting in Indonesia and encourages further research on hybrid models that integrate the strengths of both TBATS and SARIMA. Moreover, the findings may support efforts to mitigate the adverse effects of temperature fluctuations in the country.

Kata-kata kunci: peramalan suhu udara, model tbats, model sarima, evaluasi kinerja model

INTRODUCTION

Air temperature is one of the key elements contributing to climate change (Salsabila, Kusuma, & Ruchjana, 2022). Geographic and topographic factors influence the climate in each region (Kafara, Rumlawang, & Sinay, 2017). Indonesia's location along the equator and its archipelagic nature with a tropical climate result in high-intensity sunlight exposure, which in turn contributes to rising air temperatures (Aprianti, Faulina, & Usman, 2024). Air temperature in Indonesia exhibits unique characteristics, including complex seasonal patterns, discernible trends, and unstable fluctuations caused by natural phenomena such as global warming, La Niña, and El Niño.

An analysis of 31 BMKG observation stations indicates that Indonesia's average air temperature in August 2024 reached 26.9°C—marking the second-highest anomaly since 1981—exceeding the climatological average by 0.71°C. This deviation signals a potential climate disturbance that may affect various life sectors (Adityo Wicaksono, 2024). Increases in air temperature, sea surface temperature, and irregular rainfall patterns negatively impact agriculture, fisheries, public health, and ecological balance. For example, Wolff et al. (2021) found that deforestation in Berau, East Kalimantan over the past 16 years led to a 0.95°C increase in daily maximum temperatures, contributing to higher mortality rates among forest-dwelling wildlife. In order to mitigate the adverse effects of irregular temperature changes, particularly rising temperatures, reliable forecasts of future air temperatures in Indonesia are essential (Jaya & Chadidjah, 2020). Such information can be used to plan and manage natural resources and to develop strategies for adapting to climate variability.

Research on air temperature forecasting has been previously conducted. Susanti et al. (2020) utilized an ARIMA model that achieved a Mean Absolute Percent Error (MAPE) of less than 5%. However, the presence of complex seasonal patterns and trends in Indonesia's temperature data necessitates the use of models capable of handling such complexities—namely the Trigonometric, Box-Cox Transformation,

https://doi.org/10.30606/absis.v8i2.3052

ARMA Errors, Trend, and Seasonal (TBATS) model and the Seasonal Autoregressive Integrated Moving Average (SARIMA) model. The TBATS method is particularly suitable for capturing intricate seasonal structures and periodic seasonal behavior. Fajar and Nonalisa (2021) demonstrated that the TBATS model effectively handles trend and seasonality, yielding MAPE values under 10%. Similarly, Hutagalung and Sari (2024) confirmed that TBATS performs well in managing data with seasonal and trend variability.

Kusumawati and Kuswanto (2014) compared TBATS with Exponential Smoothing in modeling electricity consumption, noting that while TBATS had larger residuals, it could handle irregular seasonal data. In a related study, Guo et al. (2022) compared ARIMA and TBATS models in forecasting dengue fever incidence in Jiangsu, China, and found TBATS superior, with a MAPE of 27.44%. Despite its strengths, TBATS has limitations, particularly its slow computational speed and high resource demands when applied to large datasets (Karabiber & Xydis, 2019).

Given these computational challenges, SARIMA offers a viable alternative. It is known for its computational efficiency and suitability for regular seasonal patterns and modest trends. Agustin et al. (2022) showed that SARIMA effectively forecasted water quality data with MSE and RMSE values below 1%. Similarly, Maulana et al. (2017) applied SARIMA to rainfall forecasting and obtained a MAPE of 12.02%. In a study comparing ARIMA, SARIMA, and NNAR models for average temperature forecasting in Palembang, Solihat et al. (2022) found that while NNAR performed best with a MAPE of 1.35%, SARIMA followed closely with 1.78%, demonstrating its high forecasting capability.

However, few studies to date have comprehensively compared the performance of TBATS and SARIMA in forecasting air temperature in Indonesia using a rolling cross-validation approach. Therefore, this study aims to fill that gap by providing a robust performance evaluation of both models. The findings are expected to assist Indonesia's Meteorological, Climatological, and Geophysical Agency (BMKG) in improving daily and seasonal temperature forecasts. Accurate temperature forecasts can help farmers schedule planting and harvesting times more effectively, support the state electricity company (PLN) in anticipating electricity demand, and assist the transportation sector in planning safe and efficient operations.

METHODS

This study uses secondary data, specifically the Indonesian 2m Temperature Data for the period of January 1940 to August 2024 which amount to 1016 data, sourced from the ECMWF website. ECMWF is an international meteorological agency that is globally recognized for providing high-quality data with long-term coverage, ensuring that the data it offers is validated and reliable. This dataset represents Indonesian air temperature measurements taken at a height of 2 meters above ground level, sea, or deep water. This dataset was chosen because the 2m temperature represent atmospheric conditions directly experienced by humans, which can significantly affect daily activities. The data is divided into 900 training data points (January 1940 to December 2014) and 116 testing data points (January 2015 to August 2024). The data is divided into training and testing sets to objectively evaluate the model's performance and to avoid overfitting. The training data is used to build the model, while the testing data is used to assess the model's ability to make accurate forecast. After separating the data, the next step involves checking the stationarity of the data. Stationarity test is a fundamental requirement in forecasting using the SARIMA method. Moreover, if the data used is already stationary, it ensures the validity of statistical tests and helps avoid spurious correlations. The Box-Cox Test is applied to evaluate the stationarity of variance, while the ADF Test (Augmented Dickey-Fuller Test) and the PP Test (Phillips-Perron Test) will assess the stationarity of the mean. This study employs the Box-Cox test, ADF test and PP test to strengthen the evaluation of stationarity, encompassing both mean and variance stationarity. If the p-value obtained from the ADF Test and PP Test is lower than the significance level (0.05), the data can be considered stationary in the mean. The research is employed two forecasting methods: Trigonometric, Box-Cox Transformation, ARMA Error, Trend, and Seasonal (TBATS) and Seasonal Autoregressive Integrated Moving Average (SARIMA). This study employs a cross-validation approach to assess the model's performance.

Trigonometrics, Box Cox, Transformation, Arma Error, Trend, And Seasonal (TBATS)

TBATS is a branch of Exponential Smoothing and BATS (Box-Cox transformation, ARMA errors, Trend and Seasonal) methods. The TBATS model applies the Box-Cox transformation to stabilize the variance in time series data, handles ARMA errors, and uses trigonometry o manage intricate seasonal patterns. The process of analyzing data using TBATS is aided by R software and the model is identified using the tbats() function in the 'forecasting' package. TBATS consists of the following equations.

The equation for Box Cox transformation is:

$$y_t^{(\omega)} = \begin{cases} \frac{y_t^{\omega}-1}{\omega}, & \text{if } \omega \neq 0 \\ \ln(y_t), & \text{if } \omega = 0 \end{cases} \tag{1}$$

Box Cox transformation has several conditions, as shown in Table 1

Table 1. Box Cox Transformation Conditions

ω	Transformation
-1	$y_t^{(\omega)} = \frac{1}{y_t}$

https://doi.org/10.30606/absis.v8i2.3052

ω	Transformation
-0.5	$\mathrm{y_t}^{(\omega)} = \frac{1}{\sqrt{\mathrm{y_t}}}$
0	$\mathrm{y_t}^{(\omega)} = \ln(\mathrm{y_t})$
0.5	$\mathrm{y_t}^{(\omega)} = \sqrt{\mathrm{y_t}}$
1	$\mathrm{y_t}^{(\omega)} = \mathrm{y_t}$
2	$\mathrm{y_t}^{(\omega)} = \mathrm{y_t}^2$

The equation for seasonal pattern is :

$$y_t^{(\omega)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^{M} s_{t-m_i}^{(i)} + d_t \tag{2}$$

The equation for local and global trends is :

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t \tag{3}$$

$$b_t = (1-\phi)b + \phi b_{t-1} + \beta d_t \tag{4}$$

The equation for an error component modeled with ARMA is:

$$d_t = \sum_{i=1}^{p} \phi_i d_{t-1} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \tag{5}$$

The equation for seasonal patterns modeled with Fourier series is :

$$s_t^{(i)} = \sum_{j=1}^{M} s_{j,t}^{(i)} \tag{6}$$

$$s_{j,t}^{(i)} = s_{j,t-1}^{(i)} \cos \lambda_j^{(i)} + s^{*(i)}_{j,t-1} \cos \lambda_j^{(i)} + \gamma_1^{(i)} d_t \tag{7}$$

$$s^{*(i)}_{j,t} = -s_{j,t-1}^{(i)} \cos \lambda_j^{(i)} + s^{*(i)}_{j,t-1} \cos \lambda_j^{(i)} + \gamma_2^{(i)} d_t \tag{8}$$

where:

$\mathrm{y_t}^{(\omega)}$: Box Cox transformation results of actual data $\mathrm{y_t}$
$\mathrm{y_t}$: actual data with t = 1,2,3,…,n
ω : estimated value of Box Cox transformation
$\phi_p(B)$: $1 - \phi_1 B - \cdots - \phi_p B^p$, characteristic polynomial of the AR process

$\theta_q(B)$: $1-\theta_1 B-\cdots-\theta_q B^q$, characteristic polynomial of the MA process
l_t : level (local value) at time t
b_t : global trend at time t
d_t : error (deviation) at time t
ε_t : white noise (random error) at time t
s_t : seasonal component at time t
α : smoothing coefficient for level
β : smoothing coefficient for trend
γ_i : seasonal smoothing parameter
λ : fourier frequency

The TBATS method includes smoothing parameters, namely α, β, γ_i (Munim, 2022). Typically, γ_i consist of γ_1 and γ_2, each with different interpretations. γ_1 represents the amplitude of the first seasonal component in the Fourier series (low frequency), and its value indicates the degree of influence of the first seasonal pattern on the data. In contrast, γ_2 represents the amplitude of the second seasonal component in the Fourier series (higher frequency), which captures more detailed aspects of complex seasonal patterns (De Livera, Hyndman, & Snyder, 2011). The TBATS method produces a model in the form of TBATS$(\Omega, \{p,q\}, \phi, \{m_i, k_i\})$, where:

Ω : Box-Cox transformation component, which is useful for stabilizing variance
$\{p,q\}$: ARMA order
ϕ : damping parameter, which regulates the gradual slowing of the trend over time.
$\{m_i, k_i\}$: seasonal component, where m_i represents the duration of the i-th seasonal period and k_i denotes the number of trigonometric harmonics utilized to model the i-th seasonal pattern.

Seasonal Autoregressive Integrated Moving Average (SARIMA)

SARIMA is an enhanced version of the ARIMA (Autoregressive Integrated Moving Average) model, designed to account for and incorporate seasonal patterns. The SARIMA model can be expressed as $(p,d,q)(P,D,Q)^S$, where (p,d,q) denotes the non-seasonal orders, (P,D,Q) specifies the seasonal orders and S indicates the number of periods within a single season. The SARIMA equation can be formulated as follows :

$$\Phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D Y_t=\theta_q(B)\Theta_Q(B^S)\alpha_t \qquad (9)$$

where:

p : non-seasonal AR (autoregressive) order

d : non-seasonal differencing order
q : non-seasonal MA (moving average) order
P : seasonal AR (autoregressive) order
D : seasonal differencing order
Q : seasonal MA (moving average) order
$(1-B)^d$: non-seasonal differencing operator
$(1-B^S)^D$: seasonal differencing operator
Y_t : value of time series data at time t
S : number of periods in one season
α_t : error value at time t in the random residual (white noise)
$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_P B^P$, non-seasonal AR polynomial
$\Phi_p(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \cdots - \Phi_p B^{PS}$, seasonal AR polynomial
$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$, non-seasonal MA polynomial
$\Theta_Q(B^S) = 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \cdots - \Theta_Q B^{QS}$, seasonal MA polynomial

Selection of the best models

After identifying the model and estimating the parameters, the next is check the residuals and evaluate the forecasting results of both models by calculating the MAPE and RMSE. The selection of MAPE and RMSE as evaluation metrics is based on the seasonal and fluctuating characteristics of air temperature data in Indonesia, which require models to effectively capture both absolute and relative errors. The MAPE and RMSE values can be calculated using the following formulas:

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{10}$$

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{n}\frac{(\hat{y}_i - y_i)^2}{n}} \tag{11}$$

However, to assess the accuracy of forecasting, not only using the MAPE and RMSE values but must perform a cross-validation process to evaluate model performance to avoid overfitting. This study employs cross-validation with a rolling basis, which preserves the temporal structure of the data by using only past observations to forecast future values. This approach is well-suited for seasonal patterns and long-term trends, while also helping to prevent information leakage and reduce the risk of overfitting. The concept of this cross-validation is to divide the data into several folds to test the training and testing data in a rolling manner, with the condition that the training data increases with each fold (Shrivastava, 2020). Figure 1 shows an overview of cross-validation using a rolling base.

The first step in cross-validation using a rolling base is to divide the data into several folds evenly and sequentially. In rolling-based cross-validation, the training data consist of the data from the first fold up to the i-th fold, while the testing data are the data from the (i+1)-th fold. The training data for each fold will gradually increase, while the testing data will remain the same. The result of this cross-validation with a rolling basis is the RMSE value. The model with the smallest RMSE value is deemed superior in forecasting accuracy.

Figure 1. Overview of cross-validation using a rolling base

FINDING AND DISCUSSION

The initial step in analyzing data using TBATS and SARIMA is to understand the data and identify trends and seasonal patterns within it. Indonesia's air temperature data is shown in Figure 2.

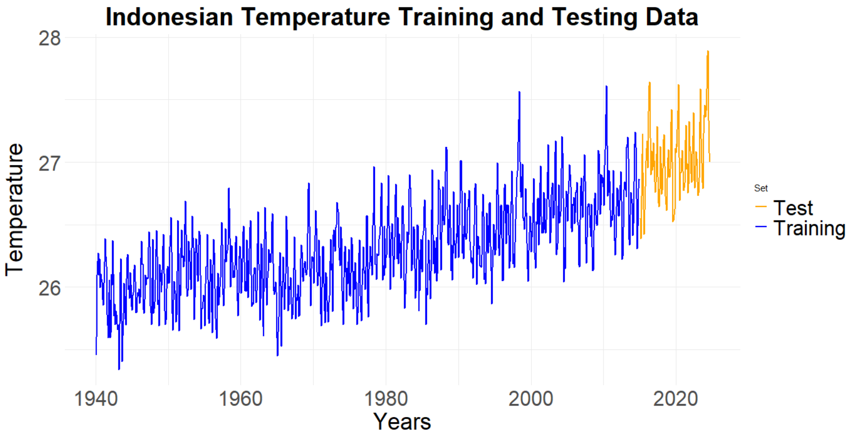

Figure 2. Indonesia's Air Temperature Data

The blue graph in Figure 2 represents the training data, while the orange graph represents the testing data. Figure 2 shows an overall upward trend, indicated by a gradual increase in average values over time. The regular up-and-down pattern suggests the presence of a consistent seasonal component. This increase in fluctuation requires conducting a stationarity test using the Box-Cox transformation, the Augmented Dickey-Fuller (ADF) test, and the Phillips-Perron (PP) test. The stationarity test is crucial because the SARIMA model requires the data to be stationary. If the data are not stationary, the model may be unable to accurately capture the underlying structure of the series, thereby compromising the reliability and accuracy of the resulting forecasts. The Box Cox Test produces a ω value of 0.2, indicating that the stationarity of the data in variance is not met and must be transformed as per the criteria in Table 1. The transformation applied is $y_t^{(\omega)} = y_t^{0.2}$. After the Box Cox transformation, the data is retested using the ADF and PP tests. The results of both tests are presented in Table 2. Both tests yield p-values below the significance level of 0.05. This suggests that the data's stationarity in the mean is satisfied, and differencing is unnecessary.

Table 2. The result of ADF test and PP tests

	Augmented Dickey Fuller Test	Phillips-Peron Test
Test Statistic	-6.4639	-9.1876
p-value	0.01	< 0.0001

The data that satisfies stationarity in both mean and variance is shown in Figure 3.

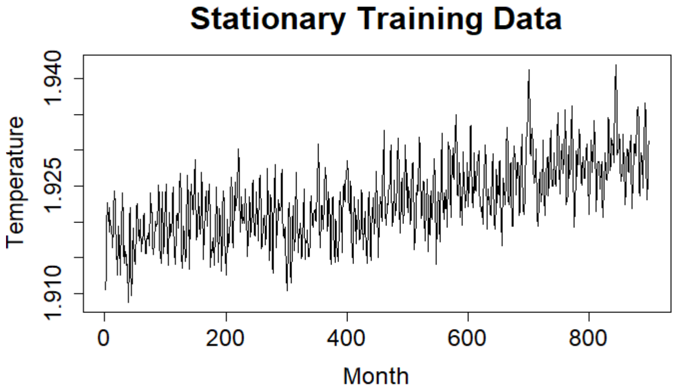

Figure 3. Stationary Training Data

Figure 3 shows fluctuations that tend to be stable, but have not yet aligned with the horizontal line, as seasonal patterns are still present in the data. Therefore, the stationary training data must be further analyzed using ACF and PACF plots. Figure 4 presents the results of the ACF and PACF plots.

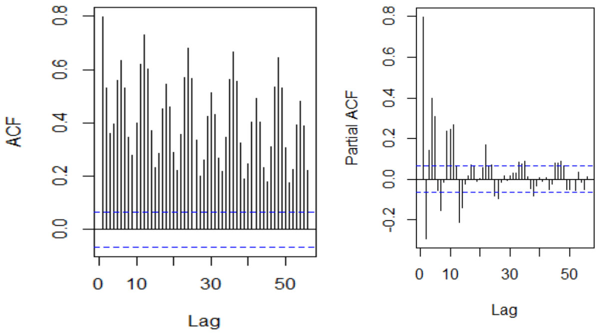

Figure 4. ACF and PACF Plot of the stationary data

Figure 4 shows the ACF plot, illustrating the presence of a seasonal pattern at lag multiples of 12, indicating that the data exhibits a seasonal pattern with a period of 12.

Trigonometrics, Box Cox, Transformation, Arma Error, Trend, And Seasonal (TBATS)

The temperature change forecast in Indonesia using the TBATS method resulted in the TBATS(0,{0,0},0.8,{<12,5>}) model. The value of 0 represents that no transformation was applied to the data. The value {0,0} indicates the p and q values in the ARMA polynomial, which do not address the trend as there are no significant ARMA components. The value 0.8 represents the damping parameter, indicating that there is a slowdown in the trend and it is approaching stability. The value {<12,5>} represents the seasonal component, where 12 is the first seasonal period, which is monthly, and 5 is the number of trigonometric harmonics. The parameter estimation results from the TBATS method are listed in Table 3.

Table 3. TBATS Parameters

TBATS parameters	
Omega (Ω)	0.000161
Alpha (α)	0.7254857
Beta (β)	-0.1578621
Damping Parameter (ϕ)	0.8
Gamma-1 Values (γ_1)	0.0002579581
Gamma-2 Values (γ_2)	0.0002910172

Table 3 indicates that the value of Ω is very small, indicating that very little data has been transformed and that the data has not changed significantly. The alpha value in the table indicates that the model is responsive in adjusting to recent changes in the data. The beta value generated by the TBATS model is negative, which interprets as a decreasing trend. However, because the beta value is small and the seasonal trend is more dominant than the linear trend, the resulting graph will still show an upward trend and seasonal patterns, without indicating a downward trend. The damping parameter value of 0.8, which is close to 1, this means that the trend component is exponentially damped and tends to decelerate over time. In other words, the model prevents the trend from increasing or decreasing indefinitely, which results in more stable long-term forecasts. The gamma value represents the seasonal component, which is expressed using Fourier series analysis. Based on Table 3, the value of γ_2 is greater than γ_1, indicating that γ_2 has a greater influence on the data. However, since both gamma values are very small, the influence of the seasonal amplitudes is relatively weak, indicating that the seasonal patterns are considered highly stable over time. After identifying the model and estimating the parameters, the next step is diagnostic testing using the Shapiro-Wilk test and the Ljung-Box test. Table 4 displays the results of both tests.

Table 4. The Result Shapiro-Wilk and Ljung-Box Tests

	Test Statistic	p-value	Result
Shapiro-Wilk	0.99698	0.08773	Normally distributed
Ljung-Box	9.4608	0.489	There is no significant autocorrelation

Based on Table 4, the p-value results from the Shapiro-Wilk test and the Ljung-Box test exceed the significance level of 0.05. This indicates that the residuals of the TBATS model follow a normal distribution and exhibit no autocorrelation, meaning the residuals are randomly generated. This suggests that the model fits the patterns in the data well.

Seasonal Autoregressive Integrated Moving Average (SARIMA)

After the transformation and stationarity test, the data satisfies stationarity in mean and variance. However, from the ACF plot in Figure 4, the lag multiples of 12 still exhibit an unstable pattern, and differencing is needed for that seasonal pattern of 12. The result of differencing for the seasonal pattern of 12 is shown in Figure 5.

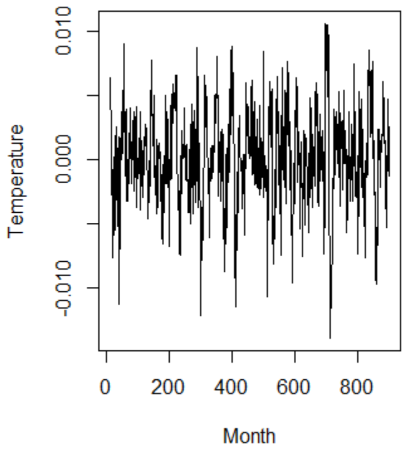

Figure 5. The result of the data that has been seasonally differenced with a period of 12

Figure 5 reveals that the data are now stationary, as it is within the horizontal field. After differencing, the ADF and PP tests must be conducted again to check for stationarity. Table 5 presents The outcomes of the ADF and PP tests.

Table 5. The Result of ADF and PP test after differencing the seasonal pattern-12

	Augmented Dickey Fuller Test	Phillips-Peron Test
Test Statistic	-10.743	-13.3156
p-value	0.01	< 0.0001

The ADF and PP test results in Table 5 demonstrate that the p-value is below the significance level of 0.05, showing that after differencing the seasonal pattern of 12, the data remains stationary in mean. Once the stationarity of the data has been tested, the next step is to determine the model by analyzing the ACF and PACF plots shown in Figure 6.

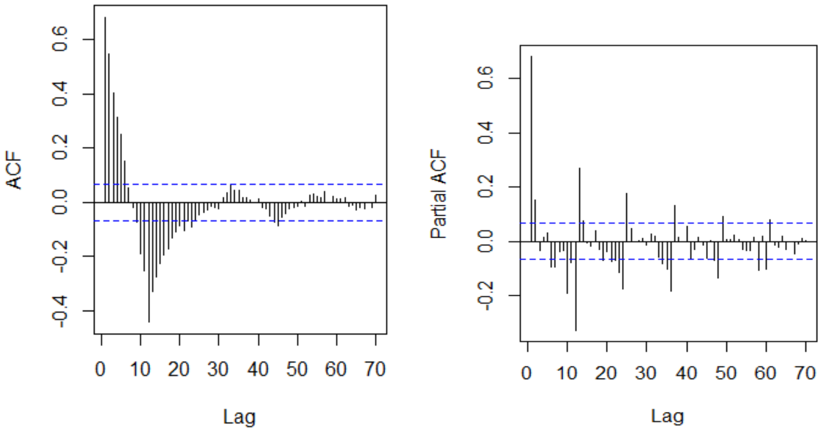

Figure 6. Plot the ACF and PACF of the data that has been seasonally differenced with a period of 12

Based on Figure 6, the resulting model is SARIMA(2,0,6)(5,1,1)12. The ACF plot shows that there are six significant lags, so the non-seasonal MA parameter (q) for the SARIMA model is 6. As for the seasonal MA parameter, it is determined by the lags that are multiples of 12 in the ACF plot. In this plot, only the 12th lag is significant, so the value of Q is 1. The PACF plot shows significant lags at the first and second positions, meaning the non-seasonal AR parameter (p) is 2. Additionally, significant lags at multiples of 12 indicate that the seasonal AR parameter is 5, as there are five significant lags at multiples of 12. The values of d and D are determined by how many times the data is differenced. In this SARIMA model, d is 0 because no differencing was applied to the non-seasonal data, while D is 1 because the seasonal data was differenced once. The SARIMA model can also be obtained using the autosarima function in RStudio. For the case of air temperature data in Indonesia, the model generated by the autosarima function is SARIMA(2,0,0)(4,1,0)12. The SARIMA(2,0,6)(4,1,0)12 is a fusion of the model derived from the ACF and PACF plots

https://doi.org/10.30606/absis.v8i2.3052

and the one from the autosarima function. The AIC values for the three models are presented in Table 6.

Table 6. Comparison of AIC Values from SARIMA Models

Models	AIC
SARIMA(2,0,6)(5,1,1)12	-8477.03
SARIMA(2,0,0)(4,1,0)12	-8401.95
SARIMA(2,0,6)(4,1,0)12	-8398.70

Table 6 shows that the SARIMA(2,0,6)(5,1,1)12 model has the lowest AIC value. However, after checking the significance of the parameters for both the (2,0,6)(5,1,1)12 and SARIMA(2,0,6)(4,1,0)12 models, the parameters in both models were found to be insignificant. On the other hand, the parameters in the SARIMA(2,0,0)(4,1,0)12 model are significant. Therefore, the model that should be chosen is SARIMA(2,0,0)(4,1,0)12 . After selecting the model, diagnostic tests should be conducted on the residuals of the SARIMA(2,0,0)(4,1,0)12 model using the Shapiro-Wilk test to check for normality and the Ljung-Box test to check for autocorrelation in the residuals. Table 7 displays the results of these diagnostic tests.

Table 7. The Result Shapiro-Wilk and Ljung-Box Tests

	Test Statistic	p-value	Results
Shapiro-Wilk	0.9954	0.008453	Not normally distributed
Ljung-Box	16.598	0.3435	There is no significant autocorrelation

Based on Table 7, the Shapiro-Wilk test produces a p-value lower than the significance level of 0.05. This p-value indicates that the residuals of the SARIMA model do not adhere to a normal distribution. However, the Ljung-Box test yields a p-value higher than the significance level of 0.05, meaning that there is no autocorrelation in the residuals of the SARIMA model, or the residuals are random. In addition to using the Shapiro-Wilk test, normality can be checked using a histogram and a Normal Quantile-Quantile (QQ) Plot. Figure 7 shows the histogram and Normal QQ Plot of the residuals from the SARIMA(2,0,0)(4,1,0)12 model.

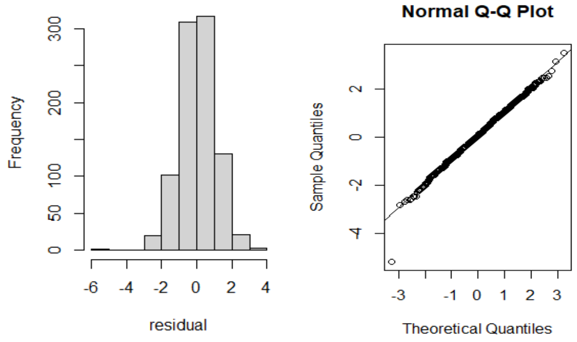

Figure 7. Histogram and Normal QQ of the residuals from the SARIMA model

Based on Figure 7, the histogram approximates a normal distribution, and the Normal QQ plot shows that the residuals lie along the straight line, with only one residual outside the line. Both of these observations indicate that the residuals of the SARIMA(2,0,0)(4,1,0)12 model are close to a normal distribution. Although the Shapiro-Wilk test indicated that the data are not normally distributed statistically, visual inspections using the Q-Q plot and histogram showed a distribution that is approximately normal. This is not considered a significant issue because the Shapiro-Wilk test is highly sensitive, especially with large sample sizes, and can detect minor deviations that are not practically impactful. Therefore, the data are considered to meet the normality assumption in a practical sense, and the analysis can proceed.

The Results of Forecasting Air Temperature in Indonesia and Cross Validation with a Rolling Basis

The following are the results of air temperature forecasting in Indonesia for the next 5 years using the TBATS and SARIMA methods.

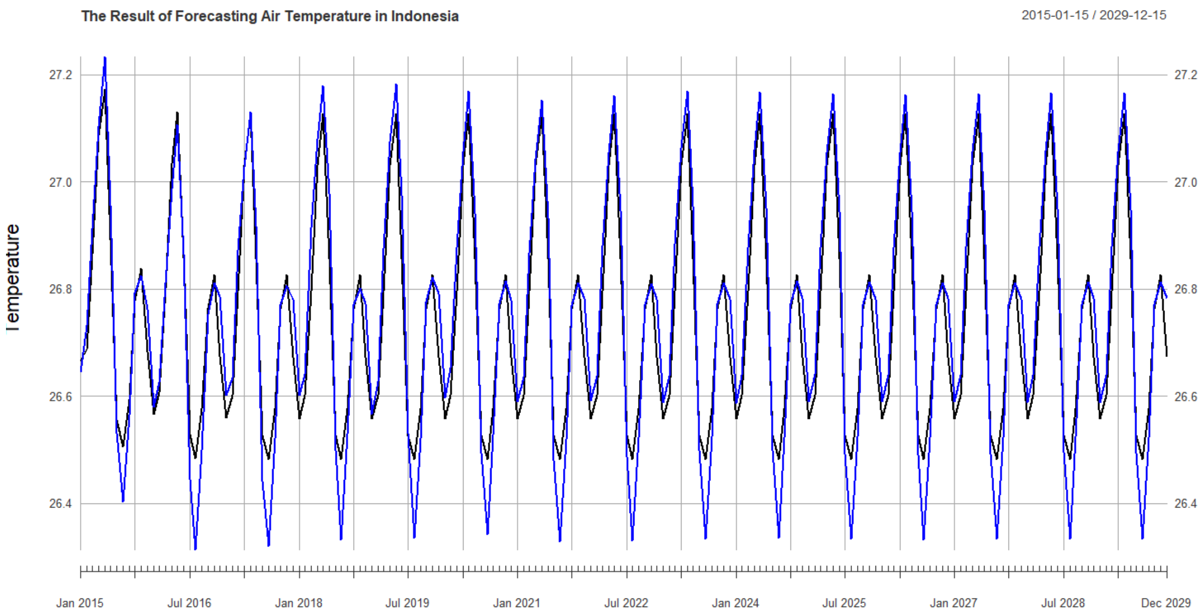

Figure 8. The Result of Forecasting using the TBATS and SARIMA Methods

Based on Figure 8, the black graph is the TBATS method of forecasting and the blue is the SARIMA method of forecasting. The forecasting result from the TBATS and SARIMA Methods are almost similar. Indonesia's future air temperature is projected to follow a seasonal pattern with regular fluctuations and stable variations, thereby allowing the forecasting results to contribute to efforts aimed at mitigating the adverse effects of unpredictable weather, particularly in the areas of agriculture, energy, and public health. Next, cross-validation process with rolling basis to evaluate the model with divides the data into 15 folds, with each fold containing 67 data points. Cross-validation using a rolling base results in the RMSE values for each fold, which are listed in Table 8.

https://doi.org/10.30606/absis.v8i2.3052

Table 8. The Results of Cross Validation using a Rolling Base

Fold Number	RMSE Values of TBATS Method	RMSE Values of SARIMA Method
2	0.2011772	0.2150456
3	0.2042598	0.1795092
4	0.2264087	0.2031667
5	0.1987422	0.206062
6	0.2500406	0.271091
7	0.1763572	0.1905056
8	0.2091329	0.1978624
9	0.2314101	0.2230308
10	0.1486877	0.173737
11	0.2729072	0.2826522
12	0.179365	0.1643387
13	0.2136591	0.2280157
14	0.303022	0.2926455
15	0.1786267	0.1537426
Average	0.2138426	0.2129575

Based on Table 8, the average cross-validation results with the rolling basis for the SARIMA method show a lower RMSE value compared to the TBATS method. A comparison of the MAPE values calculated on the testing data and the average cross-validation results with the rolling basis is presented in Table 9.

Table 9. Evaluation Results of the TBATS and SARIMA Models

Models	MAPE	Cross Validation (RMSE)
TBATS(0, {0,0}, 0.8, {<12,5>})	1.0667%	21.3843%
SARIMA(2,0,0)(4,1,0)12	1.0417%	21.2958%

The MAPE results and the average cross-validation using a rolling base are consistent, as both show that the SARIMA method produces slightly lower values. The comparison of the cross-validation using a rolling base for both methods is shown in the form of a boxplot in Figure 9.

https://doi.org/10.30606/absis.v8i2.3052

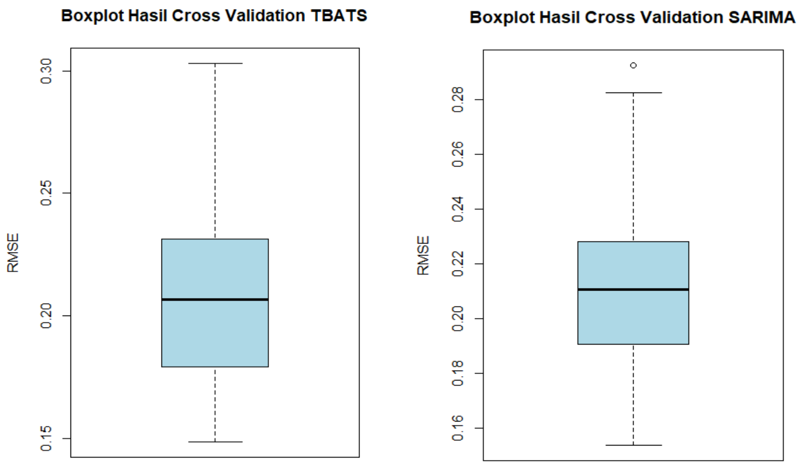

Figure 9. Boxplot Comparison of Cross Validation Results with a Rolling Basis for Both Methods

Based on Figure 9, the boxplot results for both methods appear similar. Therefore, based on Table and Figure 9, both the TBATS and SARIMA methods are equally good at predicting Indonesia's air temperature data. However, because the RMSE value for SARIMA is lower than that for TBATS, the SARIMA method is slightly superior to the TBATS method.

CONCLUSIONS AND RECOMMENDATIONS

Forecasting results show that Indonesia's air temperature in the future will have a seasonal pattern with regular fluctuations and stable variations. Based on the results of the cross-validation with a rolling basis, the TBATS model produces an RMSE value of 21.3843%, while the SARIMA model has a lower RMSE value of 21.2958%. This indicates that both SARIMA and TBATS methods perform effectively in terms of forecasting accuracy. However, although the RMSE value for SARIMA is lower than that for TBATS, the difference is not statistically significant. This research is expected to support decision-making in natural resource planning and management, particularly in the agriculture, energy, and public health sectors by providing more accurate temperature forecasts that can help mitigate the adverse impacts of temperature changes in Indonesia. This research produces small MAPE values and the RMSE cross validation results are in the quite good category because they are close to the lower limit of the sufficient category (20% - 50%). Future research is recommended to explore the development of a hybrid method that combines TBATS and SARIMA, leveraging the strengths of each model to improve forecasting performance. In addition, subsequent studies may also consider the application of machine learning approaches to enhance forecasting accuracy, particularly in capturing nonlinear patterns and complex seasonal structures.

REFERENCES

Adityo Wicaksono. (2024, November 18). *Anomali suhu udara bulan Agustus 2024*. Badan Meteorologi Klimatologi dan Geofisika (BMKG). https://www.bmkg.go.id/berita/?p=anomali-suhu-udara-bulan-agustus-2024&lang=ID&tag=anomali-suhu-udara-bulanan

Agustin, A., Rahani, F. F., & Indikawati, F. I. (2022). Prediksi kualitas air menggunakan metode Seasonal Autoregressive Integrated Moving Average (SARIMA). *Jurnal Manajemen Informatika (JAMIKA)*, *12*(2), 137–150. https://doi.org/10.34010/jamika.v12i2.8022

Aprianti, A., Faulina, N., & Usman, M. (2024). Generalized Space Time Autoregressive (GSTAR) model for air temperature forecasting in the South Sumatera, Riau, and Jambi Provinces. *InPrime: Indonesian Journal of Pure and Applied Mathematics*, *6*(1), 1–13. https://doi.org/10.15408/inprime.v6i1.36049

De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, *106*(496), 1513–1527. https://doi.org/10.1198/jasa.2011.tm09771

Fajar, M., & Nonalisa, S. (2021). Forecasting chili prices using TBATS. *International Journal of Scientific Research in Multidisciplinary Studies*, *7*(2), 1–5.

Guo, Z., Gong, H., Li, Y., Tao, C., & LuoJing, Z. (2022). Time series analysis and prediction of scarlet fever incidence trends in Jiangsu Province, China: Using ARIMA and TBATS models. *Research Square*, 1–18. https://doi.org/10.21203/rs.3.rs-2259096/v1

Hutagalung, A. A., & Sari, R. F. (2024). Forecasting Indihome users by using Trigonometrics, Box Cox, Transformation, Arma Error, Trend, and Seasonal (TBATS) methods. *Mathline: Jurnal Matematika dan Pendidikan Matematika*, *9*(1), 243–256. https://doi.org/10.31943/mathline.v9i1.579

Jaya, I. G. M., & Chadidjah, A. (2020). Modeling and forecasting of the temperature in Bandung, Indonesia: The next ten years. *International Journal of Innovative Science, Engineering and Technology (IJISET)*, *7*(7), 19–27.

Kafara, Z., Rumlawang, F., & Sinay, L. (2017). Peramalan curah hujan dengan pendekatan Seasonal Autoregressive Integrated Moving Average (SARIMA): Studi kasus curah hujan bulanan di Kota Ambon, Provinsi Maluku. *Barekeng: Jurnal Ilmu Matematika dan Terapan*, *11*(1), 63–74. https://doi.org/10.30598/barekengvol11iss1pp63-74

Karabiber, O. A., & Xydis, G. (2019). Electricity price forecasting in the Danish day-ahead market using the TBATS, ANN and ARIMA methods. *Energies*, *12*(5), Article 928. https://doi.org/10.3390/en12050928

Kusumawati, F. A., & Kuswanto, H. (2014). Pemodelan beban sistem listrik Jawa-Bali dengan menggunakan pendekatan flexible seasonality forecasting. *Prosiding Seminar Nasional Matematika Universitas Jember*, *1*(1), 121–131.

Maulana, H. A., Muliah, Sampe, M. Z., & Hanifah, F. (2017). Pemodelan dan peramalan deret waktu: Studi kasus suhu permukaan laut di selatan Jawa Timur. *Math Educa Journal*, *1*(2), 187–199.

Munim, Z. H. (2022). State-space TBATS model for container freight rate forecasting with improved accuracy. *Maritime Transport Research*, *3*, Article 100057. https://doi.org/10.1016/j.martra.2022.100057

Salsabila, S., Kusuma, D. A., & Ruchjana, B. N. (2022). Penerapan model Vector Autoregressive orde tiga pada data suhu udara rata-rata di Kabupaten Malang dan Kabupaten Sidoarjo. *Teorema: Teori dan Riset Matematika*, *7*(2), 247–254. https://doi.org/10.25157/teorema.v7i2.7200

Shrivastava, S. (2020, December 1). *Cross-validation in time series*. Medium. https://medium.com/@soumyachess1496/cross-validation-in-time-series-566ae4981ce4

Solihat, F. L. N., Putri, F. A., Lastiar, N., & Pontoh, R. S. (2022). Analisis komparatif peramalan suhu rata-rata Palembang menggunakan ARIMA, SARIMA, dan NNAR. *Seminar Nasional Statistika Aktuaria I*.

Susanti, L., Hasanah, P., & Winarni, W. (2020). Peramalan suhu udara dan dampaknya terhadap konsumsi energi listrik di Kalimantan Timur. *Barekeng: Jurnal Ilmu Matematika dan Terapan*, *14*(3), 399–412. https://doi.org/10.30598/barekengvol14iss3pp399-412

Wolff, N. H., Zeppetello, L. R. V., Parsons, L. A., Aggraeni, I., Battisti, D. S., Ebi, K. L., ... & Spector, J. T. (2021). The effect of deforestation and climate change on all-cause mortality and unsafe work conditions due to heat exposure in Berau, Indonesia: A modelling study. *The Lancet Planetary Health*, *5*(12), e882–e892. https://doi.org/10.1016/S2542-5196(21)00279-5