Jurnal Varian
Vol. 8, No. 3, October 2025, pp 281–292
e-ISSN: 2581-2017

Prediction of CO_2 Emissions Using ANN, ARIMAX, and Hybrid ARIMAX-ANN Models

Afifah Dayan Syaharani, Hervira Nur Shafira, Hikmal Mardian Irianto, Fitri Kartiasih
Politeknik Statistika STIS, East Jakarta, Indonesia

Article Info

Article history:

Received : 03-24-2025
Revised : 10-23-2025
Accepted : 10-30-2025

Keywords:

ANN;
ARIMAX;
CO_2 Emissions;
Energy Consumption;
GDP;
Hybrid ARIMAX-ANN.

ABSTRACT

The escalation of carbon dioxide (CO_2) emissions has emerged as a critical environmental concern, particularly in the context of Indonesia's pursuit of sustainable development. This study aims to forecast CO_2 emissions in Indonesia using annual time-series data spanning 1967–2023. Three methodological approaches are employed: an artificial neural network (ANN), an autoregressive model with exogenous variables (ARIMAX), and a hybrid ARIMAX-ANN model. The dataset comprises Gross Domestic Product obtained from the World Bank, along with per capita CO_2 emissions, per capita natural gas consumption, and per capita hydropower consumption sourced from Our World in Data. The findings of this research demonstrate that the hybrid ARIMAX-ANN model provides the best forecasting performance, as evidenced by the lowest RMSE, MAPE, and MAE values among the other two models. These results suggest that the hybrid model is currently the most reliable for predicting CO_2 emissions in the Indonesian context. The study enriches the expanding literature on emission forecasting by providing empirical evidence to support data-driven policymaking for climate change mitigation and sustainable energy development in Indonesia.

Accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: https://doi.org/10.30812/varian.v8i3.5045

Corresponding Author:

Fitri Kartiasih,
Statistics Study Program, Politeknik Statistika STIS,
Email: fkartiasih@stis.ac.id

How to Cite:
Syaharani, A. D., Shafira, H. N., Irianto, H. M., Kartiasih, F. (2025). Prediction of CO_2 Emissions Using ANN, ARIMAX, and Hybrid ARIMAX-ANN Models. *Jurnal Varian*, *8*(3), 281–292.

A. INTRODUCTION

The quality of the environment is a significant metric for evaluating a nation's developmental status and level of civilization. One of the main threats to the environment is greenhouse gas emissions, of which carbon dioxide (CO_2) accounts for the largest proportion. Economic production processes generate greenhouse gases that are released into the atmosphere, triggering the greenhouse effect, climate change, and rising average global temperatures (Yamacli & Tuncsiper, 2024). These phenomena affect global ecosystems, threaten food security, and increase the risk of natural disasters such as floods and droughts.

The relationship between economic growth and environmental pollution has long been a concern in environmental studies (Amalia et al., 2025; Kartiasih & Pribadi, 2020; Kartiasih & Setiawan, 2020). The relationship between economic development and environmental degradation has been a persistent subject of debate within environmental research (Kartiasih, Sandi, et al., 2025; Miswa & Kartiasih, 2025). According to the Environmental Kuznets Curve (EKC) hypothesis, environmental pollution tends to rise during the initial phases of economic growth as industrialization expands. However, once a certain level of income is attained, emissions start to fall as a result of the implementation of environmentally sound technologies and more stringent (Kartiasih, Azhari, & Rinangku, 2025; Siregar et al., 2025). Empirical research often points to an inverted U-shaped curve relating GDP and CO_2 emissions (Han & Lin, 2025). However, this relationship is not consistently observed across all countries, with developing countries exhibiting the most variation.

Global Carbon Project shows that Indonesia, as a rapidly growing economy, has a significant position in contributing to global CO_2 emissions. In 2023, Indonesia ranked eighth globally in fuel-based energy emissions and second in emissions from land-use change. In the ASEAN region, Indonesia ranks first with total fuel-based energy emissions of 773 $MtCO_2$, far surpassing Vietnam (335 $MtCO_2$) and Thailand (264 $MtCO_2$). Responding to this challenge, Indonesia has committed through the renewal of its Nationally Determined Contribution (NDC) as part of ratifying the Paris Agreement. This commitment includes an emissions-reduction target of 29% to 41% by 2030. Indonesia's target for reducing carbon emissions by 2030 has been updated from the initial goal of a 29% reduction (equivalent to 835 million tons of CO_2) to a 32% reduction, representing 912 million tons of CO_2. In this effort, forecasting CO_2 emissions is a strategic step toward understanding future trends and supporting effective, data-driven policymaking.

Forecasting CO_2 emissions in Indonesia requires a comprehensive understanding of the relationship between emissions and various external factors that influence them. In this context, several previous studies have identified dynamic relationships between exogenous variables such as economic growth, energy consumption, and urbanization. Research by Marwa et al. (2022) examines the two-way causality between economic growth and electricity consumption, while a study by Bashir et al. (2021) demonstrates the causal relationship between urbanization, energy consumption, and CO_2 emissions in Indonesia.

A study by Khan et al. (2023) examined the nexus between per capita income and greenhouse gas emissions across 108 developing nations from 2000 to 2016. The results validated the Environmental Kuznets Curve (EKC) hypothesis, indicating that rising income levels, particularly when paired with enhanced human capital, contribute to a decline in environmental degradation. Another study by Lin et al. (2018) re-examined the relationship between coal consumption, CO_2 emissions, and economic growth in China from 1969 to 2015 using the Bootstrap ARDL method. While the analysis did not identify a long-term equilibrium among the variables, the Granger causality test uncovered a short-run bidirectional relationship between economic growth and CO_2 emissions, suggesting that each variable influences the other in the short term. The causal relationship between CO_2 emissions and economic growth has also been examined in the context of Malaysia. For instance, Mehraaein et al. (2021) employed an ARDL approach using data from 1971 to 2014 and found evidence of an inverted-U-shaped relationship consistent with the Environmental Kuznets Curve (EKC) hypothesis, along with unidirectional causality running from economic growth to environmental degradation.

The environmental Kuznets curve (EKC) explains the relationship between income inequality and income levels, has become an essential framework in this field (Mahmood et al., 2023). The EKC posits that in the early stages of economic growth, environmental degradation worsens. However, after reaching a certain per capita income threshold, these adverse effects begin to diminish. While per capita income is considered an independent variable within the EKC hypothesis, existing literature includes studies that incorporate additional data to refine CO_2 emissions modeling. Some studies use variables beyond per capita income to enhance the accuracy of CO_2 emissions modeling. A recent study analyzed the relationships among CO_2 emissions, urbanization, energy consumption, and economic growth in Pakistan using the ARDL and VECM approaches, revealing both short- and long-run causal effects of urbanization and economic growth on CO_2 emissions (Sufyanullah et al., 2022).

Several studies have developed models to analyze the factors influencing CO_2 emissions in various countries using different approaches. For instance, Iftikhar et al. (2024) proposed a hybrid model combining regression and time series models to predict CO_2 emissions in Pakistan, demonstrating that this combination can improve prediction accuracy. Additionally, Yamacli & Tuncsiper (2024) conducted a comparative analysis of ARIMAX and deep learning methodologies for modeling CO_2 emissions in Turkey. Recent research has increasingly utilized the Autoregressive Distributed Lag (ARDL) method to test the Environmental Kuznets Curve (EKC) hypothesis across different national contexts. For example, Ali et al. (2021) applied the ARDL approach to analyze Pakistan's case and identified an inverted U-shaped relationship between economic growth and CO_2 emissions, aligning with the EKC hypothesis.

Although a variety of forecasting techniques are employed internationally, Indonesia has seen limited application of sophisticated methods such as ARIMAX and ANN. Research by Yamacli & Tuncsiper (2024) in Turkey shows that deep learning-based methods yield more accurate predictions than traditional models. Meanwhile, Iftikhar et al. (2024) validated the effectiveness of the proposed hybrid forecasting method, demonstrating its high accuracy and efficiency in estimating CO_2 emissions. The optimized hybrid model successfully predicted Pakistan's per capita CO_2 emissions.

This study aims to compare CO_2 emission forecasting methods using ANN, ARIMAX, and hybrid ARIMAX-ANN approaches in Indonesia. The exogenous variables include income per capita and energy consumption from natural gas and hydropower per capita. These variables were chosen because they have a significant influence on CO_2 emissions, as shown by various previous studies. The novelty of this research lies in the application of hybrid methods to CO_2 emissions in Indonesia, an area that has rarely been explored. In addition, this study evaluates the accuracy of various time-series modeling approaches by including exogenous variables, such as GDP per capita and energy consumption from natural gas and hydropower per capita, which are relevant to Indonesia's unique characteristics.

By integrating hybrid methods and analyzing the performance of ANN, ARIMAX, and hybrid ARIMAX-ANN models individually, this research provides deep insights into the underlying patterns and trends of CO_2 emissions in Indonesia. The results not only contribute to the literature on greenhouse gas emissions forecasting but also offer practical implications for policymakers in designing more effective emissions reduction strategies. The approach used is a comparison of the ANN, ARIMAX, and hybrid ARIMAX-ANN models, which are rarely applied to CO_2 emissions in Indonesia. This research also identifies the most suitable method based on model evaluation criteria, such as RMSE, MAE, and MAPE. It aims to determine whether Indonesia can achieve the new NDC (Nationally Determined Contributions) target for CO_2 emission reduction. This approach supports Indonesia's Net Zero Emissions 2060 target while providing a strong basis for data-driven decision-making amid global climate change challenges.

B. RESEARCH METHOD

The study begins by splitting the dataset into 80% for training and 20% for testing. The first stage applies an Artificial Neural Network (ANN) with backpropagation to capture nonlinear patterns within the data. The second stage constructs an Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) model using historical per capita CO_2 emissions, with Gross Domestic Product (GDP), hydropower energy consumption, and natural gas consumption as exogenous variables. The model is analyzed using autoregressive, differencing, and moving-average components. In the third stage, a hybrid ARIMAX-ANN model is developed, where ARIMAX predictions are used to generate residuals, which the ANN subsequently models to improve accuracy. Finally, model performance is evaluated using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), with accuracy interpreted according to established guidelines. The overall modeling process consists of four main approaches, each described in the following subsections.

1. Artificial Neural Network (ANN)

The second method applied in this research to model CO_2 emissions is an Artificial Neural Network (ANN). ANN is a machine learning technique that mimics the workings and structure of the human brain, consisting of a number of interconnected neurons as processing units. The training of artificial neural networks (ANN) involves the adjustment of connection weights through the backpropagation algorithm (Alam & AlArjani, 2021). ANNs possess a high degree of adaptability, allowing them to model a wide range of data patterns, including nonlinear relationships. Furthermore, their ability to handle complex data structures, including those that violate normality assumptions, makes them suitable for situations where traditional statistical methods struggle (Zhang et al., 2018). The forecasting output for the j_{th} sample is determined using a mathematical expression, as shown in Equation (1).

$$y_j = \sum_{i=1}^{n} \omega_{ki,\, j} f\left(h_k\right) + \beta_j \tag{1}$$

The weight connecting the hidden node to the output node at the i-th iteration is denoted by $\omega_{ki,\, j}$, β_j represents the bias of the j_{th} output node, The result of the hidden node after applying the activation function is. Finally, y_j refers to the output of the j_{th} sample.

2. ARIMAX

This study compares three different methods to predict the amount of CO_2 emissions per capita. The study's first approach, the ARIMAX model (autoregressive integrated moving average with exogenous variables), predicts future CO_2 emissions by incorporating both GDP and energy consumption as explanatory variables, and the historical behavior of CO_2 emissions through autoregressive components. The ARIMAX model was chosen because it combines the historical pattern of CO_2 emissions (represented by the ARIMA model) with the influence of external factors such as income and energy consumption. This combination makes the ARIMAX model a powerful tool for analyzing and predicting CO_2 emissions.

This study refers to Janhuaton et al. (2024) which examines the impact of gross domestic product and energy consumption from hydroelectric power plants and gas per capita on CO_2 emissions in Indonesia. The relationship can be modeled using the autoregressive model with exogenous variables, as shown in Equation (2). Furthermore, the structure of the time series models applied in this study includes the autoregressive process in Equation (3), the differencing operator in Equation (4), the moving average process in Equation (5), and the regression form of exogenous variables in Equation (6).

$$\Delta^d Y_t = \sum_{i=1}^{P} \emptyset_i \, \Delta^d Y_{t-i} + \sum_{i=1}^{P} \theta_j \varepsilon_{t-j} + \beta_k X_{tk} + \theta_j \varepsilon_t \, , \tag{2}$$

$$Y_t = \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \ldots + \emptyset_p Y_{t-p} + \varepsilon_t \, , \tag{3}$$

$$\Delta^d Y_t = (1 - B)^d \, Y_t, \tag{4}$$

$$Y_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{5}$$

$$Y_t = \beta_1 X_{t1} + \beta_2 X_{t2} + \ldots + \beta_k X_{tk} \tag{6}$$

The ARIMAX model is defined as follows: Y_t denotes the value of the time series at time t, ϕ represents the coefficient associated with the autoregressive component, Δ dsignifies the differencing operation performed d times, B is the backshift operator, θ is the coefficient of the moving average component, β_k represents the coefficient of the exogenous variable X_{tk} and ε represents the white noise error term. This study uses the ARIMAX model, which is a development of the ARIMA model. This model includes four main components: autoregressive (AR), which measures the dependence of current values on previous values, integration (I), which handles trends in the data, moving average (MA), which captures the random component, and exogenous variable (X), which represents the influence of outside factors (Sutthichaimethee & Ariyasajjakorn, 2017).

3. Hybrid ARIMAX-ANN

The Hybrid ARIMAX-ANN model combines two models: an ARIMAX linear model and a neural network with a back-propagation algorithm as a nonlinear model. In the Hybrid ARIMAX-ANN model, the ANN is used to predict the ARIMAX model's errors, thereby improving forecasting accuracy (Mahmudi et al., 2024). The mathematical formulation of the Hybrid ARIMAX-ANN model is presented in Equation (7).

$$y_t = L_t + N_t + \varepsilon_t \tag{7}$$

Where L_t is the linear component contained in the given data, N_t is the nonlinear component. After the ARIMAX linear model is applied to the data, the next step is to calculate the residuals—the differences between the original data (Yt) and the linear model's predictions (Lt). This residual contains only the nonlinear part of the data. Next, the residuals are analyzed to determine whether any linear patterns remain. If there is none, the next step is to check whether the residuals exhibit a nonlinear pattern (Supriya, 2020).

4. Performance Evaluation

The study employs RMSE, MAPE, and MAE to assess the model's predictive capabilities. These evaluation metrics assess the discrepancy between predicted and observed values; smaller values indicate better model performance. The predictive accuracy is gauged using three key indicators, Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE). RMSE, outlined in Equation (8), captures the average squared deviation between forecasts and actual outcomes. MAPE, described in Equation (9), conveys the average percentage difference between predicted and actual values. Meanwhile, MAE, presented in Equation (10), indicates the mean size of prediction errors.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \widehat{y_i})^2}{n}} \tag{8}$$

$$MAPE = \left(\frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - \widehat{y_i}|}{|y_i|} \right) \times 100 \tag{9}$$

$$MAE = \left(\frac{1}{n} \right) \times \sum_{i=1}^{n} |y_i - \widehat{y_i}| \tag{10}$$

In the above equation, n indicates the total number of observations, and is the actual value of the i_{th} observation, and y_i is the predicted value of the i-th observation. In addition, in the research by Ağbulut (2022), the interpretation guidelines for evaluating the MAPE metric are categorized into four levels of forecasting accuracy, as shown in Table 1. The overall modeling process in this study, which integrates ANN, ARIMAX, and a hybrid ARIMAX-ANN, is illustrated in Figure 1.

Table 1. Interpretation Guidelines for the MAPE in Forecasting Accuracy Assessment

Forecasting Accuracy	MAPE Range
High Accuracy	$\leq 10\%$
Good Accuracy	$> 10\%$ and $\leq 20\%$
Sufficient Accuracy	$> 20\%$ and $\leq 50\%$
Inaccurate	$> 50\%$

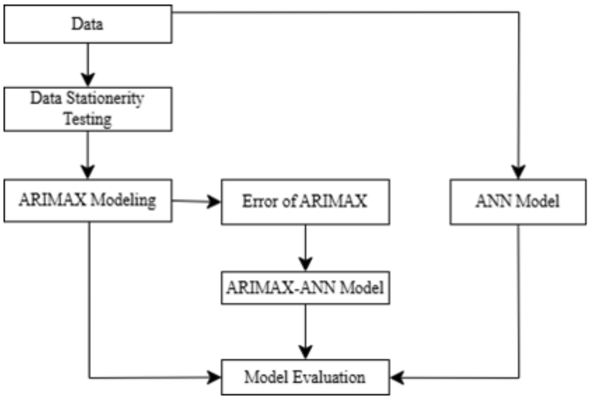

Figure 1. Flowchart ARIMAX, ANN, and Hybrid ARIMAX-ANN

C. RESULT AND DISCUSSION

1. Descriptive Data

Table 2 presents descriptive statistics for the four variables analyzed —CO_2 emissions, GDP, natural gas consumption, and hydropower generation energy consumption per capita. Each variable has a minimum value, first quartile (Q1), median, mean (average), third quartile (Q3), and maximum value. The CO_2 emissions variable has an average of -1.2050 and a median of -1.1159, indicating a relatively symmetrical distribution with a slight negative trend. The range of CO_2 emission values is quite wide, from -2.6434 to -0.2304, indicating variation in changes in CO_2 emissions over time. Meanwhile, the GDP variable has a mean of 1445.5 and a median of 732.3, indicating a right-skewed distribution (positive), with some very high values raising the mean above the median. The maximum GDP per capita of 4876.3 indicates significant economic growth during the period.

For energy consumption variables, such as natural gas and hydroelectricity, the variability is quite high. For example, the variable Hydro has a mean of 114.41 and a median of 121.10, indicating a more balanced distribution compared to the variable energy consumption from natural gas per capita. The energy consumption from natural gas variable shows a large variation, with a minimum value of 58.87 and a maximum of 1805.11. The mean of 1087.45, which is lower than the median of 1438.45, indicates that the Gas data distribution has a negative skew or leans to the left.

Table 2. Interpretation Guidelines for the MAPE in Forecasting Accuracy Assessment

Statistics	CO_2	GDP	Hydropower	Natural Gas
Minimum	0.2304	53.2	23.31	58.87
Q1	0.6588	454.9	49.34	481.55
Median	1.1159	732.3	121.10	1438.45
Mean	1.2050	1445.5	114.41	1087.45
Q3	1.6403	2218.5	145.47	1637.46
Maximum	2.6434	4876.3	258.30	1805.11

Indonesian CO_2 emission data from 1967 to 2023 were employed in this study. The model was trained on data from 1967 to 2007 and subsequently evaluated on data from 2008 to 2023. The ARIMAX model uses CO_2 per capita data as endogenous variables, and gross domestic product (GDP) and energy consumption from both hydropower and natural gas per capita as exogenous variables. The ANN model uses gross domestic income and per capita energy consumption from hydropower and natural gas as inputs. Meanwhile, the hybrid ARIMAX-ANN model uses the ARIMAX residuals as one of its input variables.

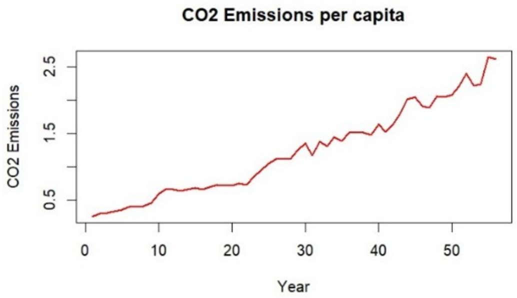

Figure 2. Trends in CO_2 emissions per capita

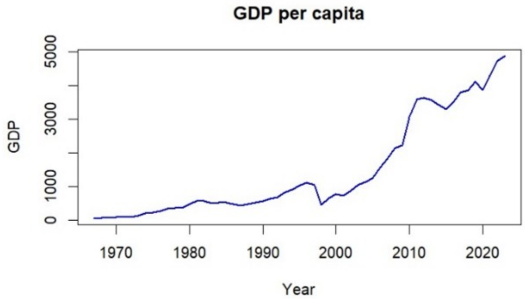

Figure 3. GDP per capita trends

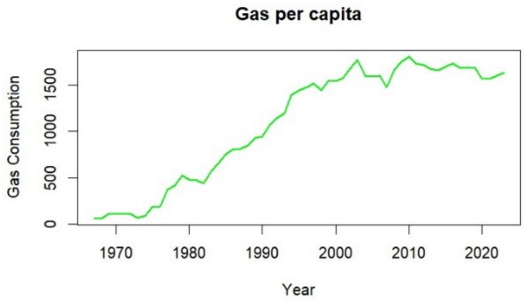

Figure 4. Gas Consumption trend per capita

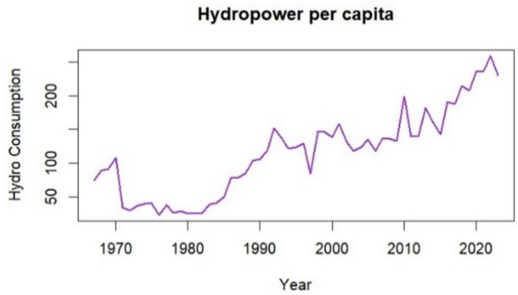

Figure 5. Consumption trends from hydropower per capita

The trend of Indonesia's yearly CO_2 emissions from 1967 to 2023 is depicted in Figure 2. The data originates from Our World in Data, which uses information from the Global Carbon Project. The graph shows an almost consistent increase in emissions, indicating a growth trend throughout the years. In the 1967–2023 period, the GDP per capita graph (Figure 3) shows a steady increase, punctuated by significant declines in the late 1990s and 2021, reflecting the impact of the economic crisis and the COVID-19 pandemic. Similarly, the graph of natural gas consumption per capita, shown in Figure 4, shows a consistent increase, with more stable growth than hydroelectric power generation per capita, indicating a gradual shift towards cleaner energy sources. The graph of energy consumption from hydropower generation per capita shows a consistent upward trend, reflecting the increasing adoption of renewable energy over time (Figure 5).

As a first step, the collected data were pooled, and the Augmented Dickey-Fuller (ADF) test was conducted to analyze data stationarity at the initial level, first differentiation, and second differentiation. The ADF test results show that, at this level, not all variables are stationary because the p-value is greater than 0.05. In the first differentiation, most variables show significant results. Still, total stationarity is achieved only in the second differentiation, with all probabilities equal to 0.000 and t-statistic values meeting the stationarity criteria. Details of the test results are shown in Table 3.

Table 3. Stationarity Test Results (ADF)

Test		CO_2	GDP	Hydro Power	Natural Gas
Level	Intercept	0.0370	35.8951	1.5449	54.9479
	Prob.	0.9990	0.9998	0.9865	0.4815
	t stat	1.4420	1,9688	0.5351	-1.5889
	Intercept	0.0357	-44.9829	4.9072	55.5710
	Trend	0.0142	5.6923	1.2948	-0.3191
	Prob.	0.1272	0.9771	0.0498	0.9868
	t stat	-3.0549	-0.5675	-3.4937	-0.3606
1st Difference	Intercept	0.0703	64.4409	4.7664	27.6615
	Prob.	0.0000	0.0000	0.0000	0.0000

DOI: https://doi.org/10.30812/varian.v8i3.5045

Test		CO_2	GDP	Hydro Power	Natural Gas
	t stat	-7.3387	-5.5167	-7.5393	-7.0232
	Intercept	0.0315	-23.6292	-6.0199	59.6000
	Trend	0.0014	3.2928	0.3768	-1.0535
	Prob.	0.0000	0.0000	0.0000	0.0000
	t stat	-7.6526	-5.9861	-7.9426	-7.3245
2nd Difference	Intercept	0.0038	9.4783	1.0547	-0.6212
	Prob.	0.0000	0.0000	0.0000	0.0000
	t stat	-9.8175	-8.2517	-9.0715	-6.6632
	Intercept	-0.0017	-4.3896	5.1783	24.5408
	Trend	0.0002	0.4633	-0.1353	0.7895
	Prob.	0.0000	0.0000	0.0000	0.0000
	t stat	-9.7181	-8.1694	-8.9393	-6.7358

Once the data are stationary, the next step is to perform time-series analysis and modeling. This research uses three methods to model CO_2 emissions, each based on per capita income and energy consumption from three main sources.

2. Artificial Neural Network (ANN)

The ANN method begins with data normalization, followed by splitting the normalized data into training (the first 40 data points) and test (the last 17 data points) sets in an 80:20 ratio. ANN modeling employs a neural network to predict CO_2 emissions based on per capita income (GDP), natural gas consumption, and hydroelectric power generation, with a single hidden layer containing 1 neuron per unit. Since the data is continuous, the linear output parameter is set to TRUE. The training results are visualized in the network plot shown in Figure 6. During the testing phase, the ANN model predicts CO_2 emissions using the test data, and the predictions are then denormalized back to the original scale using the denormalize function.

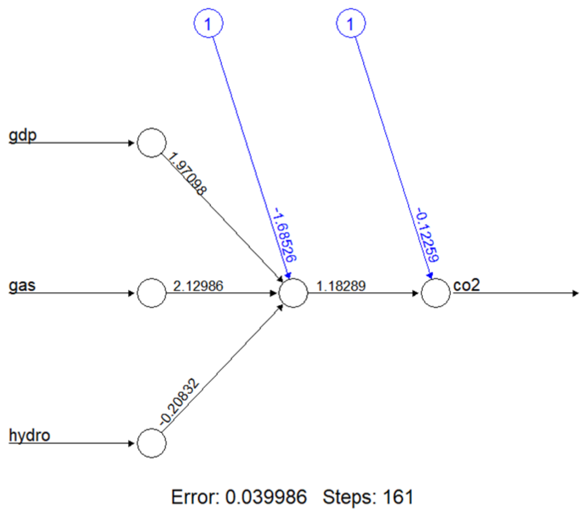

Figure 6. ANN plot

3. ARIMAX

All selected exogenous variables, GDP per capita, natural gas, and hydropower, exhibited Variance Inflation Factor (VIF) values below 10, suggesting minimal multicollinearity. To evaluate stationarity, the Augmented Dickey-Fuller (ADF) test was applied, confirming that the data became stationary after first differencing. Table 4 presents the ARIMAX model's parameter estimates, showing that gas consumption has a positive and statistically significant impact at the 5% level. In contrast, the coefficients for GDP and hydropower are not statistically significant. The constant term is also significant at the 5% level. In addition, the model includes an AR(1) component that is statistically significant at the 1% level.

DOI: https://doi.org/10.30812/varian.v8i3.5045

Table 4. ARIMAX Parameter Estimation Results

Variables	Coefficient	Standard Error (s.e)	Probability
C	0.263915	0.035375	0.0000
AR(1)	0.430130	0.206961	0.0000
GDP	0.000221	0.000137	0.1074
Gas	0.000559	0.000267	0.0361
Hydro	-0.000212	0.000497	0.6686

The model is obtained using the Auto ARIMA approach, which automatically selects the best parameters to model the regression with errors modeled as ARIMA(1,0,0). Auto ARIMA evaluates various combinations of autoregressive (AR), differencing (I), and moving average (MA) parameters to minimize information criteria such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). In this model, no MA or moving average component is selected, but there is one AR(1) autoregressive component.

4. ARIMAX-ANN

The hybrid ARIMAX-ANN combines the strengths of linear and nonlinear methodologies by using ARIMAX to identify and represent linear trends, while leveraging artificial neural networks (ANN) to uncover and learn complex nonlinear patterns in the dataset. This integrated framework provides a more comprehensive and flexible forecasting strategy. The modeling process begins with the development of an ARIMAX model to generate initial predictions of CO_2 emissions. The residuals, defined as the discrepancies between the predicted and actual values, are subsequently modeled using an ANN to account for nonlinear components not captured by the linear ARIMAX structure.

The residuals from the ARIMAX model, after normalization, were used as target variables in the artificial neural network training. The chosen ANN architecture consists of a single hidden layer with 1 neuron, following a configuration similar to the pure ANN model. Once the ANN training is complete, the model is used to predict residuals on the test data. The residual predictions obtained are then normalized back to the original scale. The final prediction of the hybrid model is obtained by combining the ARIMAX prediction with the ANN residual prediction on the test data. Thus, the hybrid model can integrate the strengths of ARIMAX in capturing long-term linear patterns and ANN in handling more complex non-linear components, as illustrated in Figure 7.

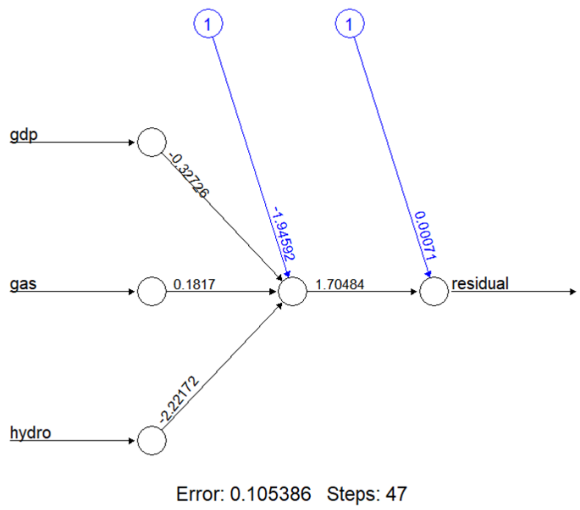

Figure 7. Hybrid ARIMAX-ANN plot

5. Performance Evaluation

Based on the model performance evaluation results in Table 5, the hybrid ARIMAX-ANN model achieves the best forecasting performance, as indicated by the lowest MAE, MAPE, and RMSE values. These findings show that the hybrid approach effectively integrates ARIMAX's strength in capturing linear patterns with ANN's capability to model nonlinear complexities, resulting in forecasts that are both accurate and robust. On the other hand, although ARIMAX and ANN perform adequately on

DOI: https://doi.org/10.30812/varian.v8i3.5045

their own, they are less capable than the hybrid model in detecting and modeling the nuanced structures within the data. Overall, the hybrid ARIMAX-ANN model is recommended as the most reliable method for predicting CO_2 emissions in this study. In contrast, further refinement of the individual models may enhance their future performance.

Table 5. Metric Evaluation Performance

Model	MAE	MAPE	RMSE
ANN	1.0547	47.71%	1.0988
ARIMAX	0.2045	8.71 %	0.2496
Hybrid ARIMAX-ANN	0.2040	8.69%	0.2491

The hybrid ARIMAX-ANN model demonstrated the highest forecasting accuracy among the three models and was therefore selected to predict CO_2 emissions per capita. Table 6 presents the forecast results for the period 2024–2030. The results show that CO_2 emissions are projected to slightly decrease from 2.6076 tons per capita in 2023 to 2.2447 tons per capita in 2024, before gradually increasing each year to reach 2.4455 tons per capita in 2030. This pattern indicates that although emissions initially decline, they are expected to rise steadily in the following years, reflecting a gradual upward trend toward 2030.

Table 6. Results of forecasting CO_2 Emissions per capita using ARIMAX-ANN

Year	CO_2 emissions
2024	2.2447
2025	2.2709
2026	2.3028
2027	2.3374
2028	2.3731
2029	2.4092
2030	2.4455

To complement the numerical forecast results shown in Table 6, a graphical representation of the predicted trajectory of CO_2 emissions per capita is provided. Visualizing the forecast alongside historical data enables clearer interpretation of the trend pattern and facilitates comparison between past actual values and future projections. This visualization highlights the persistent upward trend in emissions, as projected by the hybrid ARIMAX-ANN model. The forecast plot is shown in Figure 8.

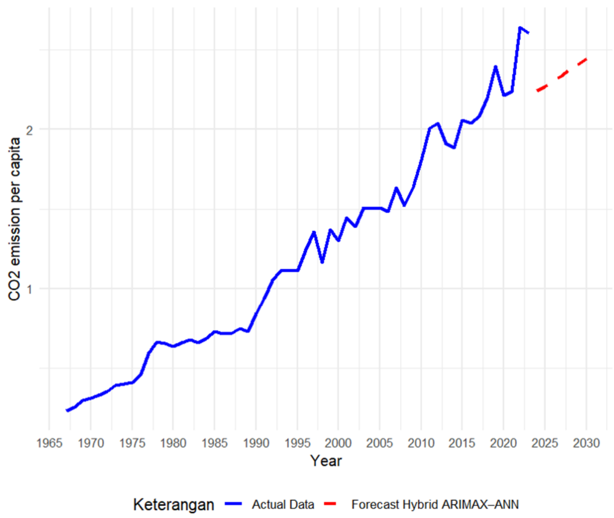

Figure 8. Forecasting CO_2 emissions per capita using Hybrid ARIMAX-ANN

The rising trend in CO_2 emissions per capita observed in the Hybrid ARIMAX-ANN forecasting results indicates that external factors, such as economic growth and energy consumption, remain the primary contributors to carbon emissions. This projection underscores the need for more effective mitigation policies to slow the rate of emissions growth, particularly through energy efficiency strategies and the expedited transition to greener energy sources. Ensuring sustainable economic growth without compromising environmental quality is crucial for the future.

The results of this study are supported by Wongsathan & Chankham (2016), who revealed that a hybrid model combining ARIMAX and Neural Networks outperformed standalone ARIMA, ARIMAX, and ANN models in forecasting PM10 levels during the summer season in Chiang Mai, Thailand. Similarly, Iftikhar et al. (2024) demonstrated that a hybrid forecasting approach offered high accuracy and efficiency in estimating CO_2 emissions, with the refined model effectively projecting Pakistan's per capita CO_2 output. In line with these findings, Wahidah et al. (2024) applied a hybrid ARIMAX–NN model to forecast Indonesia's export values, demonstrating that the hybrid approach provided more accurate and stable predictions than single models, thereby reinforcing the reliability of hybrid forecasting in capturing both linear and nonlinear data patterns. Based on these results, the hybrid ARIMAX-ANN is the most recommended model for predicting CO_2 emissions in Indonesia, given its ability to capture nonlinear relationships among variables. Although the hybrid ARIMAX-ANN approach can produce accurate predictions, the forecasting results indicate that the emission-reduction target set in Indonesia's latest NDC (Nationally Determined Contributions) has not yet been achieved. This finding makes an important contribution to supporting Indonesia's greenhouse gas mitigation policy and demonstrates the potential of statistical methods for forecasting environmental data. The development of hybrid or machine learning-based models remains necessary to address the challenges of more complex data patterns in the future.

D. CONCLUSION AND SUGGESTION

Based on the research results, the Hybrid ARIMAX–ANN model proved to be the most effective approach in predicting CO_2 emissions in Indonesia. This is evidenced by its superior performance metrics compared to the other two models, with a MAE value of 0.2040, MAPE of 8.69%, and RMSE of 0.2491, as shown in Table 5. These metrics highlight the hybrid model's robust performance, demonstrating its ability to combine the strengths of statistical methods and machine learning: effectively modeling linear trends via ARIMAX and capturing nonlinear dynamics via an ANN. In contrast, the standalone ARIMAX model performed slightly lower (MAPE = 8.71%), while the ANN model produced the weakest results (MAPE = 47.71%), indicating that it struggled to capture the dominant linear trends in the data. These findings imply that hybridizing ARIMAX and ANN improves forecasting accuracy for data with mixed linear and nonlinear dynamics.

Forecasting results using the hybrid ARIMAX-ANN model indicate that CO_2 emissions in Indonesia are expected to continue increasing, contrary to the 30% reduction target in the Nationally Determined Contribution (NDC) commitment. To achieve this target, evaluation and improvement of energy policies and mitigation strategies are urgently needed to ensure realistic, sustainable emission reductions that are aligned with Indonesia's NDC target. Accelerating the transition to renewable energy, improving energy efficiency in key sectors, and reducing dependence on fossil fuels are important steps that must be prioritized. In addition, it is necessary to develop a more comprehensive forecasting model by considering external factors, such as global energy price dynamics, international policy changes, and technological developments.

ACKNOWLEDGEMENT

The authors would like to express their deepest gratitude to the Department of Statistics, Politeknik Statistika STIS, for the continuous support and valuable academic environment that made this research possible. The authors sincerely thank all colleagues and mentors who provided constructive feedback and guidance throughout the preparation of this paper. Their encouragement and expertise have been invaluable in completing this research successfully.

DECLARATIONS

AUTHOR CONTIBUTION

First Author: Conceptualization, software data processing and analysis. Second Author: Writing original draft, layouting, editing and data visualization. Third Author: Conceptualization, writing original draft and methodology. Fourth Author: Editing and reviewing. All authors discussed the results and contributed to the final manuscript.

FUNDING STATEMENT

This study received no external funding.

COMPETING INTEREST

The authors declare that they have no competing interests in this article.

REFERENCES

Ağbulut, Ü. (2022). Forecasting of transportation-related energy demand and CO2 emissions in Turkey with different machine learning algorithms. *Sustainable Production and Consumption*, *29*, 141–157. https://doi.org/10.1016/j.spc.2021.10.001

Alam, T., & AlArjani, A. (2021). A Comparative Study of CO_2 Emission Forecasting in the Gulf Countries Using Autoregressive Integrated Moving Average, Artificial Neural Network, and Holt-Winters Exponential Smoothing Models (A. Donateo, Ed.). *Advances in Meteorology*, *2021*, 1–9. https://doi.org/10.1155/2021/8322590

Ali, M. U., Gong, Z., Ali, M. U., Wu, X., & Yao, C. (2021). Fossil energy consumption, economic development, inward FDI impact on CO_2 emissions in Pakistan: Testing EKC hypothesis through ARDL model. *International Journal of Finance & Economics*, *26*(3), 3210–3221. https://doi.org/10.1002/ijfe.1958

Amalia, M. F., Junianto, R. R., Kartiasih, F., & Rahmadani, R. (2025). Examining the Impact of Energy Use, Economic Growth, and Forest Area on CO_2 Emissions: Consequences for Achieving the SDGs. *Jurnal Ilmiah Pendidikan Lingkungan dan Pembangunan*, *25*(02), 1–26. https://doi.org/10.21009/plpb.v25i02.42230

Bashir, A., Susetyo, D., SUHEL, S., & Azwardi, A. (2021). Relationships between Urbanization, Economic Growth, Energy Consumption, and CO_2 Emissions: Empirical Evidence from Indonesia. *The Journal of Asian Finance, Economics and Business*, *8*(3), 79–90. https://doi.org/10.13106/JAFEB.2021.VOL8.NO3.0079

Han, T. T. T., & Lin, C.-Y. (2025). Exploring long-run CO_2 emission patterns and the environmental kuznets curve with machine learning methods. *Innovation and Green Development*, *4*(1), 100195. https://doi.org/10.1016/j.igd.2024.100195

Iftikhar, H., Khan, M., Żywiołek, J., Khan, M., & Linkolk López-Gonzales, J. (2024). Modeling and forecasting carbon dioxide emission in Pakistan using a hybrid combination of regression and time series models. *Heliyon*, *10*(13), e33148. https://doi.org/10.1016/j.heliyon.2024.e33148

Janhuaton, T., Ratanavaraha, V., & Jomnonkwao, S. (2024). Forecasting Thailand's Transportation CO_2 Emissions: A Comparison among Artificial Intelligent Models. *Forecasting*, *6*(2), 462–484. https://doi.org/10.3390/forecast6020026

Kartiasih, F., Azhari, A., & Rinangku, R. E. B. I. (2025). Carbon Trading Potential in the Forest Sector: Analysis at the Province Level in Indonesia. *EKO-REGIONAL: Jurnal Pembangunan Ekonomi Wilayah*, *20*(1), 14–29. https://doi.org/10.32424/er.v20i1.14668

Kartiasih, F., & Pribadi, W. (2020). Environmental Quality and Poverty Assessment in Indonesia. *Jurnal Pengelolaan Sumberdaya Alam dan Lingkungan (Journal of Natural Resources and Environmental Management)*, *10*(1), 89–97. https://doi.org/10.29244/jpsl.10.1.89-97

Kartiasih, F., Sandi, I. M., Abioga, N. R., Aditya, R. D., & Sabrina, R. (2025). Analysis of Carbon Emissions in ASEAN Manufacturing: Input-Output and Panel Data Approach. *Efficient: Indonesian Journal of Development Economics*, *8*(1), 323–329. https://doi.org/10.15294/jsnnrt53

Kartiasih, F., & Setiawan, A. (2020). *Aplikasi Error Correction Mechanism* dalam Analisis Dampak Pertumbuhan Ekonomi, Konsumsi Energi, dan Perdagangan Internasional terhadap Emisi CO_2 di Indonesia. *MEDIA STATISTIKA*, *13*(1), 104–115. https://doi.org/10.14710/medstat.13.1.104-115

Khan, M., Rana, A. T., & Ghardallou, W. (2023). FDI and CO_2 emissions in developing countries: The role of human capital. *Natural Hazards*, *117*(1), 1125–1155. https://doi.org/10.1007/s11069-023-05949-4

Lin, F.-L., Inglesi-Lotz, R., & Chang, T. (2018). Revisit coal consumption, CO_2 emissions and economic growth nexus in China and India using a newly developed bootstrap ARDL bound test. *Energy Exploration & Exploitation*, *36*(3), 450–463. https://doi.org/10.1177/0144598717741031

Mahmood, H., Hassan, M. S., Rej, S., & Furqan, M. (2023). The Environmental Kuznets Curve and Renewable Energy Consumption: A Review. *International Journal of Energy Economics and Policy*, *13*(3), 279–291. https://doi.org/10.32479/ijeep.14270

Mahmudi, M., Safitri, I. B., & Zulkifli, D. U. (2024). Modeling Rainfall in Jakarta with Hybrid ARIMAX-ANN Model. *Kontinu: Jurnal Penelitian Didaktik Matematika*, *8*(1), 1. https://doi.org/10.30659/kontinu.8.1.1-20

Marwa, T., Bashir, A., Atiyatna, D. P., Hamidi, I., Mukhlis, M., & Sukanto, S. (2022). The Link between Economic Growth, Electricity Consumption, and CO_2 Emissions: Evidence from Indonesia. *Signifikan: Jurnal Ilmu Ekonomi*, *11*(2), 253–272. https://doi.org/10.15408/sjie.v11i2.26286

Mehraaein, M., Afroz, R., Rahman, M. Z., & Muhibbullah, M. (2021). Dynamic Impact of Macroeconomic Variables on the Ecological Footprint in Malaysia: Testing EKC and PHH. *The Journal of Asian Finance, Economics and Business*, *8*(5), 583–593. https://doi.org/10.13106/JAFEB.2021.VOL8.NO5.0583

Miswa, S. D., & Kartiasih, F. (2025). Nexus between rural poverty and environmental quality: Empirical evidence from Indonesia. *Asia-Pacific Journal of Regional Science*, *9*(1), 57–81. https://doi.org/10.1007/s41685-024-00370-6

Siregar, M. R., Halma, K., Alifiyanti, R., & Kartiasih, F. (2025). Nexus between Extreme Poverty and CO2 Emissions in Indonesia: An Empirical Investigation with VECM Approach. *Journal of Economics Research and Social Sciences*, *9*(2), 264–281. https://doi.org/10.18196/jerss.v9i2.26339

Sufyanullah, K., Ahmad, K. A., & Sufyan Ali, M. A. (2022). Does emission of carbon dioxide is impacted by urbanization? An empirical study of urbanization, energy consumption, economic growth and carbon emissions - Using ARDL bound testing approach. *Energy Policy*, *164*, 112908. https://doi.org/10.1016/j.enpol.2022.112908

Supriya, K. (2020). A Study on the Performance of ARIMAX-ANN Hybrid Model over the Other Forecasting Models while Forecasting the Damage Caused by Brown Plant Hopper (Nilaparvata lugens) in Telangana State. *International Journal of Current Microbiology and Applied Sciences*, *9*(9), 2609–2616. https://doi.org/10.20546/ijcmas.2020.909.326

Sutthichaimethee, P., & Ariyasajjakorn, D. (2017). Forecasting Energy Consumption in Short-Term and Long-Term Period by Using Arimax Model in the Construction and Materials Sector in Thailand. *Journal of Ecological Engineering*, *18*(4), 52–59. https://doi.org/10.12911/22998993/74396

Wahidah, F. P., Wahyuningsih, S., & Siringoringo, M. (2024). Forecasting the Value of Indonesia's Exports using Model Hybrid Arimax-NN. *Science & Technology Asia*, *29*, 6274. https://doi.org/10.14456/SCITECHASIA.2024.71

Wongsathan, R., & Chankham, S. (2016). Improvement on PM-10 Forecast by Using Hybrid ARIMAX and Neural Networks Model for the Summer Season in Chiang Mai. *Procedia Computer Science*, *86*, 277–280. https://doi.org/10.1016/j.procs.2016.05.062

Yamacli, D. S., & Tuncsiper, C. (2024). Modeling the CO2 Emissions of Turkey Dependent on Various Parameters Employing ARIMAX and Deep Learning Methods. *Sustainability*, *16*(20), 8753. https://doi.org/10.3390/su16208753

Zhang, Z., Beck, M. W., Winkler, D. A., Huang, B., Sibanda, W., & Goyal, H. (2018). Opening the black box of neural networks: Methods for interpreting neural network models in clinical applications. *Annals of Translational Medicine*, *6*(11), 216–216. https://doi.org/10.21037/atm.2018.05.32