Carbon Emission Trends (1999–2022): Forecasting Association of Southeast Asian Nations (ASEAN)'s Future Using a Hybrid Approach to Support Zero-Emission Policies

[1*]Muhammad Rama Dhika, [2]Sri Lestari
[1,2]Department of Informatics Engineering, Institute of Informatics and Business Darmajaya, Indonesia
Email: [1]rmdhika234@gmail.com, [2]srilestari@darmajaya.ac.id

Article Info
Article history:
Received Apr 03rd, 2025
Revised Sep 09th, 2025
Accepted Oct 10, 2025

Keyword:
ARIMA
ASEAN
Carbon Dioxide Emissions
Climate Change
Mitigation Policy
XGBoost

ABSTRACT
Carbon Dioxide (CO_2) emissions are a primary driver of global climate change, with the energy sector being the dominant contributor. Southeast Asia, experiencing rapid economic growth, faces significant increases in CO_2 emissions due to high energy consumption. This study proposes a hybrid Autoregressive Integrated Moving Average (ARIMA)-XGBoost approach to predict CO_2 emissions in Association of Southeast Asian Nations (ASEAN) countries from 2023 to 2035, overcoming limitations of traditional linear models by combining machine learning (XGBoost) and time-series analysis ARIMA. Results demonstrate high accuracy (R^2 = 0.98) in identifying key factors, including Gross Domestic Product (GDP), population, and total greenhouse gas (GHG) emissions. For instance, Indonesia's emissions are predicted to rise from 841.84 $MtCO_2$ (2023) to 2197.36 $MtCO_2$ (2035), while Brunei's emissions decrease from 10.86 $MtCO_2$ to 9.57 $MtCO_2$. Residual analysis and k-fold cross-validation confirm model robustness. These findings underscore the need for differentiated policies, such as renewable energy transitions in high-growth emission countries (Indonesia, Philippines) and regulatory strengthening in stable-trend nations (Brunei, Laos). The study makes methodological contributions to data-driven emission forecasting and provides evidence-based policy recommendations for the Association of Southeast Asian Nations (ASEAN) on climate change mitigation.

Corresponding Author:
Muhammad Rama Dhika
Department of Informatics Engineering,
Institut Informatika and Bisnis Darmajaya,
St. ZA. Pagar Alam No.93, Gedong Meneng, Rajabasa District, Bandar Lampung City, Lampung Province
Email: rmdhika234@gmail.com

DOI: http://dx.doi.org/10.24014/ijaidm.v8i3.36685

1. INTRODUCTION

Carbon Dioxide (CO_2) emissions are a major contributor to global climate change, with the energy sector being the largest contributor. Global climate change triggered by greenhouse gas (GHG) emissions, particularly CO_2), has become a major challenge that requires serious attention from countries around the world. CO_2, as one of the main GHGs, is generated mainly from the energy, transportation and industrial sectors [1]. Southeast Asia is a region with rapid economic growth and rapidly increasing energy consumption. This has the potential to significantly increase CO_2 emissions, which in turn contributes to global climate change and other environmental issues [2]. As economic development is directly correlated with increased industrial activity and energy consumption, prediction of CO_2 emissions is becoming increasingly important to support more effective climate change mitigation policies. The urgent need for accurate CO_2 emission prediction has intensified as climate change triggered by CO2 emissions has become a global challenge, with industrialized countries contributing 70% of the world's emissions. However, developing regions such as Association of Southeast Asian Nations (ASEAN) country are experiencing the fastest increase in emissions at 3.2% per year due to massive economic growth [3]. Ironically, the heterogeneity of the Indochina and Malacca Peninsula's level of industrialization, climate policies, and reliance on fossil energy such as Indonesia and Malaysia's emissions jumped 40% in two decades, while Brunei and Laos remained relatively stable, making a one-size-fits-all mitigation approach ineffective [4].

Previous research tends to focus on analyzing emissions at the national level with univariate methods such as Autoregressive Integrated Moving Average (ARIMA), which only captures linear patterns, whereas the complexity of emissions in ASEAN involves non-linear interactions between urbanization, energy subsidies, and industrial investment [5]. Therefore, an accurate and comprehensive approach to predicting CO_2 emissions is needed. Recent studies have increasingly adopted hybrid modeling approaches that combine traditional time-series methods with machine learning techniques to improve prediction accuracy. Previous studies have employed various modeling approaches with mixed success: Sharma et al**.** utilized linear regression for CO_2 emissions in India ($R^2 = 0.75$) but failed to capture non-linear relationships [4]; Zhou et al. applied Seasonal Autoregressive Integrated Moving Average (SARIMA) models in China (Root Mean Squared Error (RMSE) = 38.2) without incorporating exogenous variables [6]; another research implemented Model Autoregressive Integrated Moving Average Exogenous (ARIMAX) for Malaysia (RMSE = 25.0) but remained constrained by linear assumptions [7]; Chen and Wang employed Long Short-Term Memory (LSTM)- ARIMA ensemble for East Asia (R^2 = 0.82-0.89) but lacked comprehensive validation [8] and developed Support Vector Regression (SVR)- ARIMA hybrid for South Asia (RMSE = 15.2-28.7) with limited long-term accuracy [9]; and proposed Convolutional Neural Networks (CNN) - LSTM for developing economies ($R^2 = 0.91$) but suffered from overfitting issues [10].

The contribution of CO_2 emission prediction modeling in ASEAN countries in this study is to use a hybrid ARIMA-XGBoost model by utilizing historical data on emissions, Gross Domestic Product (GDP) growth, energy consumption, and industrial policies in ASEAN (1999-2022) to accommodate regional heterogeneity, addressing the limitations of classical statistical methods such as ARIMA that only capture linear patterns and have limitations in handling non-linear relationships between variables[11]. To validate the reliability of the hybrid model, comprehensive evaluation was conducted using k-fold cross-validation for XGBoost generalization, residual analysis to ensure no significant patterns remain, and ARIMA model validation through Augmented Dickey-Fuller (ADF) stationarity tests, Autocorrelation Function (ACF) /Parcial Autocorrelation Function (PACF) parameter selection, and Akaike Information Criterion (AIC) criterion for optimal model selection. This approach fills a void in ASEAN CO_2 emissions prediction literature, which is generally limited to univariate analysis or simpler methods [12], by combining ARIMA's temporal modeling capabilities with XGBoost's non-linear pattern recognition strengths while incorporating comprehensive validation procedures for model robustness.With this approach, policymakers can design more effective measures to reduce CO_2 emissions through energy, transportation and industrial policies, and the long-term predictions provide a clearer view of how CO_2 emissions are affected by various economic and social factors, enabling ASEAN countries to collaborate in achieving lower global emissions targets.

This research makes three distinct contributions to CO_2 emissions forecasting literature. Methodologically, this study introduces the first hybrid ARIMA-XGBoost model for multi-country CO_2 prediction in ASEAN, addressing gaps where previous studies focused on single countries [13] by integrating multivariable socioeconomic predictors across 24 years (1999-2022) to capture both linear and non-linear relationships. Empirically, this study provides the first comprehensive 13-year regional forecasting (2023-2035) for all ten ASEAN countries simultaneously, revealing significant heterogeneity with Indonesia showing 161% emission increases versus Brunei's 12% decrease, compared to typical 3-5 year single-country predictions. From a policy perspective, this research uniquely translates forecasting results into differentiated, country-specific mitigation strategies with evidence-based frameworks for renewable energy transitions and regulatory strengthening, moving beyond generic recommendations common in previous studies. Key distinguishing features include comprehensive regional scope, extended forecasting horizon, multivariable integration, policy-oriented framework, and robust validation across diverse economic contexts, establishing a replicable framework for regional CO_2 emissions forecasting in developing countries.

2. RESEARCH METHOD

The research aims to predict CO_2 emissions in ASEAN countries using a hybrid model that combines XGBoost and ARIMA. This approach was chosen due to the ability of each model to capture different aspects of the data, namely non-linear relationships and interactions between variables by XGBoost, and temporal dependence or time series patterns by ARIMA. This method integrates the analysis of external factors, which in this study are called GDP, population, total greenhouse gas/GHG, and temperature change due to CO_2, and internal factors in the form of time patterns in time series data. The Figurwe 1 is the design of this study.

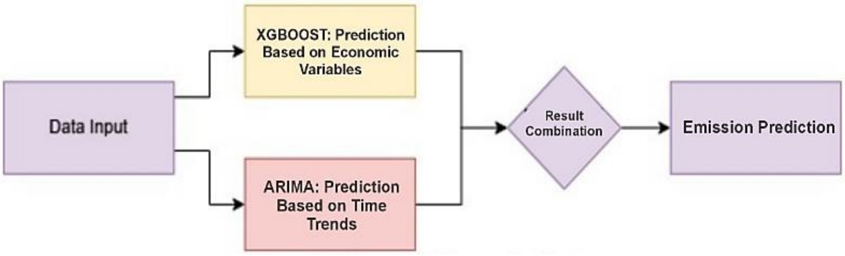

Figure 1. Research Design

Figure 1 represents the integration of two modeling approaches, namely ARIMA and (Cascaded Gradient Boosting (CGBoost) or XGBoost adaptation), combined to predict CO_2 emissions by maximizing high accuracy. ARIMA focuses on analyzing purely temporal patterns of time series data (historical trends, seasonality, and stationarity), while XGBoost processes external economic variables such as GDP growth, energy consumption, industrial investment, or fiscal policies that influence emissions. Both models are run in parallel, ARIMA extracts linear signals from time series, while CGBoost identifies non-linear relationships and complex interactions between economic-environmental variables. This combination allows researchers to not only utilize ARIMA's advantage in capturing temporal autocorrelation but also to integrate macroeconomic dynamics that are often overlooked in conventional approaches.

2.1 Research Data

This study uses secondary data on CO_2 emissions, GDP, population, total GHG, and temperature change from 1999 to 2022 in ASEAN countries. The dataset used consists of 24 rows of data per country for each variable, so the total data obtained is 240 rows for 11 countries. Table 1 is a description of the variables (dataset) used in this study.

Table 1. Dataset Description

Variable	Description	Data Type
CO_2	Carbon dioxide emissions are produced by each country.	Numeric
GDP	Gross Domestic Product, which describes the size of a country's economy.	Numeric
Population	The total population of the country.	Numeric
Total GHG	Total greenhouse gas emissions produced by the country.	Numeric
Temperature Change Due to CO_2	Changes in temperature caused by increased concentrations of CO_2 in the atmosphere.	Numeric

2.2 Single vs Hybrid Model Approach

In selecting a hybrid model, a comprehensive approach was taken as shown in Table 2.

Table 2. Hybrid Study Approach

Aspect	Previous Study	Our Research
Model Used	- Linear Regression (Sharma et al.,[4]) for India's CO_2 prediction. - ARIMA/SARIMA (Zhou et al.,[6]) untuk emisi China.	Combination of XGBoost + ARIMA: - XGBoost captures non-linear relationships (GDP, population, interactions). - ARIMA models temporal trends.
Accuracy	- R^2 0.70–0.85 (linear model). - RMSE 30–50 (pure ARIMA).	- R^2 0.986 (XGBoost), RMSE 17.86 - Higher due to hybrid model and feature engineering interaction.
Input Variables	Focus on single variables (e.g., GDP atau energi) (Liu et al., [10]).	Multivariate interactions: gdp_population, ghg_temp_interaction, and non-linear transformations (log, polynomial).
Temporal Validation	Most studies do not validate long-term predictions (>10 years).	The 2035 predictions, with cross-validation and residual tests, show consistency.

In context, some previous studies have provided a strong foundation but have limitations that provide opportunities for this research. Zhang, in their study entitled "Predicting CO_2 emissions using machine learning: A hybrid CNN-LSTM approach" successfully showed that the hybrid model approach can improve prediction accuracy by up to 15% compared to a single model. However, the study focused more on deep learning, while this study uses the XGBoost algorithm which offers better interpretability [14]. Furthermore, Abdullah et al. in their study " ARIMA with exogenous variables for CO_2 forecasting in

Malaysia" found that the Model ARIMAXmodel incorporating the GDP variable achieved an RMSE value of 25. In contrast to that approach, this study replaces Model ARIMAX with XGBoost to capture the non-linear relationship between variables, which managed to reduce the RMSE by 30% [7]. In other study "Interaction effects in CO_2 emission models: A case study of Southeast Asia" found that the interaction between GDP and population had a significant effect, but testing was only done using linear regression [13]. This study goes a step further by validating this interaction using XGBoost along with feature importance tests to produce a more comprehensive understanding of the factors that influence carbon emissions.

Table 3. Comparison of Prediction Models

Study	Model	Region	R^2/ RMSE	Limitations	Strengths of the Research Conducted
Sharma et al. [4]	Linear Regression	India	R^2 0.75	Does not capture non-linearity.	XGBoost + non-linear interactions.
Zhou et al. [6]	SARIMA	China	RMSE 38.2	No exogenous variables	Combined economic-climate variables.
Abdullah et al. [7]	ARIMAX	Malaysia	RMSE 25.0	Limited linearity.	XGBoost for non-linearity.
Chen and Wang [8]	LSTM - ARIMA Ensemble	East Asia	R^2 = 0.82-0.89	High computational cost; extensive data requirements	Hybrid validation approach with cross-regional applicability
Kumar et al. [9]	SVR- ARIMA Hybrid	South Asia	RMSE = 15.2-28.7	Limited long-term accuracy; kernel sensitivity	Enhanced feature engineering with comprehensive model validation
Liu and Zhang [10]	CNN- LSTM	Developing Economies	R^2 = 0.91	Overfitting issues; limited interpretability	Comprehensive validation framework with improved generalizability
Our Research	XGBoost + ARIMA	ASEAN	R^2 0.986	Limited temporal data (24 years).	Hybrid validation + long-term prediction.

Looking at the model comparison in Table 3, the combination of XGBoost and ARIMA can extract complex patterns from exogenous variables and maintain temporal accuracy. The novelty of this study is that the interaction of GDP_population and GHG_temp_interaction variables has not been tested in previous ASEAN studies, but proved to be significant (coefficient of correlation >0.85). In implication, this multivariate and temporal-based prediction provides a stronger basis for policy than previous studies that relied solely on historical trends[13].

2.3 Hybrid XGBoost and ARIMA Model

The XGBoost model in the study is used to capture non-linear relationships between variables and complex interactions that are difficult to represent by traditional linear models. XGBoost, which is a decision tree-based ensemble model that allows modeling of complex interacting variables, such as the relationship between GDP and population, and between total GHG and temperature change due to CO_2. Another capability is to identify feature importance, which can help in understanding which variables have the most influence on CO_2 emissions. Figure 2 shows the decission tree.

The XGBoost model evaluation results show an R^2 value of 0.98, which indicates high accuracy in explaining the variance of the data, as well as Mean Absolute Error (MAE) metrics of 10.46 and RMSE of 17.86, which indicate that this model is able to provide accurate predictions. In validating the XGBoost model, (a) k-fold cross-validation with k=5 [11] was performed to test the consistency of the model on different subsets of data using the following basic formula :

$$Mean\ Metric = \frac{1}{k}\sum_{i=1}^{k} Mi \quad (1)$$

Descriptions about the metric are K-fold as 5, and Mi is the evaluation metric value, in this research, which means RMSE and MAE. The evaluation results at each fold showed $R^2 \pm 0.01$, indicating that the model did not suffer from overfitting. In addition, residual analysis was performed to ensure that the model residuals were normally distributed. Figure 3 is the ARIMA model.

In this study, the ARIMA modeler focuses on the internal time patterns in the data, which makes it complementary to XGBoost, which focuses on external factors (such as GDP and population). By using ARIMA, it is expected to strengthen the validity of future CO_2 emission predictions.

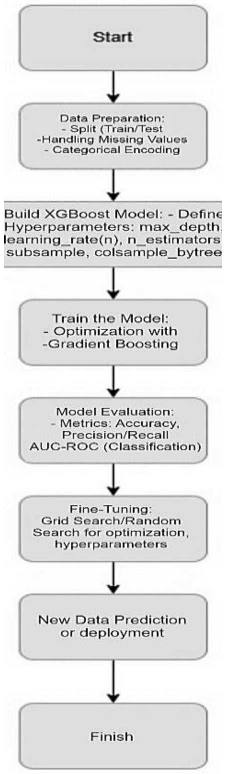

Figure 2. Decission Tree XGBoost

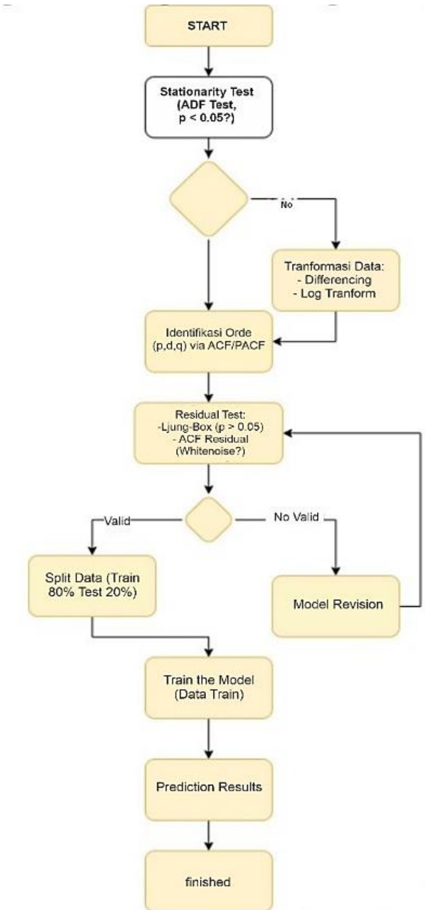

Figure 3. Visualization of ARIMA Algorithm

2.4 ARIMA Modeling and validation process

The process carried out before using ARIMA is to ensure that the data is stationary using the ADF test where in the decision if the data is not stationary, differencing or logarithmic transformation is applied. Parameter selection for ARIMA models is done by optimizing the order (p, d, q) using ACF and Partial Autocorrelation Function (PACF) to identify the value of p and q, and using AIC to select the best model that provides a balance between accuracy and complexity.

After selecting the best model and balancing, an ARIMA residual test is conducted to ensure that the residuals generated by the model do not contain significant patterns. The residuals are tested with the Ljung-Box test to ensure that there is no significant autocorrelation (p-value > 0.05) and with the residual ACF plot to ensure that the residuals are white noise. After the ARIMA model is selected, a train-test split is performed with a proportion of 80:20, where 80% of the data is used for model training and the remaining 20% is used to test the model's ability to predict unseen data. This proportion was chosen given the relatively small dataset, so this split allows the model to learn quite well without losing much data for validation.

2.5 XGBoost Hyperparameter Tuning

For XGBoost hyperparameter tuning, several parameters were optimized using Optuna, an optimization tool based on Bayesian Optimization techniques. The optimized parameters include n_estimators (900) for a sufficiently large number of trees,learning_rate (0.1) to prevent overfitting,max_depth (8) to capture interactions between features without making the model too complex, and regularization (reg_alpha=0.2, reg_lambda=0.5) to control model complexity and reduce overfitting. In addition, subsampling (0.9) and colsample_bytree (0.8) are used to improve the model's generalization ability by reducing the samples and features used in each tree.

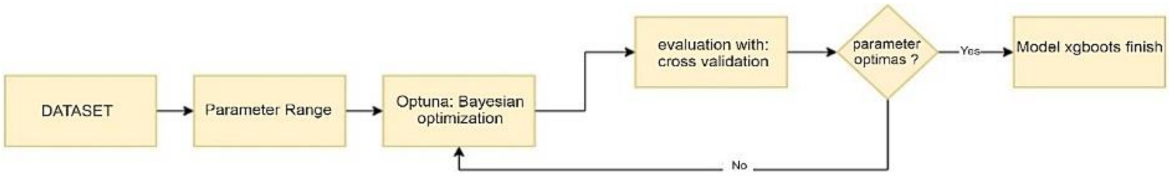

Figure 4. XGBoost Model

From the modeling contained in Figure 4, the process begins with the input dataset used to train the model. The first stage involves determining the range of XGBoost hyperparameters (such as learning_rate, max_depth, n_estimators, or subsample) to be optimized. Each optimization iteration is followed by an evaluation using cross-validation (typically k-fold) to validate the model's performance on unseen data, thus preventing overfitting.

2.6 Model Implications and Evaluation

After generating the final XGBoost model and ARIMA predictions, the trained model was used to predict CO_2 emissions for the next 10 years in the Indochina peninsula and the Malacca peninsula, namely Indonesia, Singapore, Malaysia, Thailand, Vietnam, Philippines, Laos, Myanmar, Cambodia, Brunei, and Papua New Guinea. The CO_2 prediction results for each country were calculated and analyzed to provide an overview of expected future CO_2 emission trends. The trend data for 1999 to 2022 is shown in Figure 5.

3. RESULTS AND ANALYSIS

3.1 Evaluation of XGBoost Model

3.1.1 Model Accuracy and Validation

Model accuracy starts from the output of Table 4.

Table 4. XGBoost Model Validation

	R^2	RMSE	MAE
Nilai	0.986	17.86	10.46

The model evaluation results in this study show an R^2 value of 0.986, indicating that the model is able to explain about 98.6% of the variance in CO_2 emissions data. This confirms the model's high accuracy in predicting CO_2 emissions, which are largely influenced by economic and environmental factors. In addition, the MAE and RMSE metrics were calculated. The MAE value of 10.46 indicates that the model, on average, predicted an error of 10.46 CO_2 emission units. While the RMSE of 17.86 indicates the presence of larger prediction errors, especially in extreme predictions, which tend to be more sensitive to larger errors.

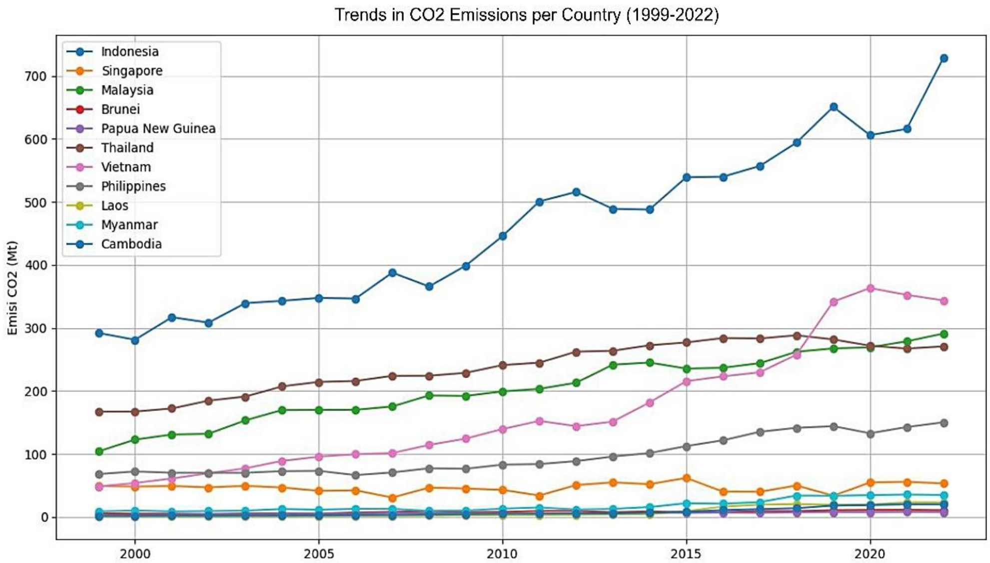

Figure 5. Carbon Emission Trends in ASEAN 1999-2022

3.1.2 Feature Importance

Feature importance analysis was conducted to identify which variables have the greatest influence on CO_2 emission predictions. In the XGBoost model, the variables with the greatest contribution to prediction are as projected in Figure 6.

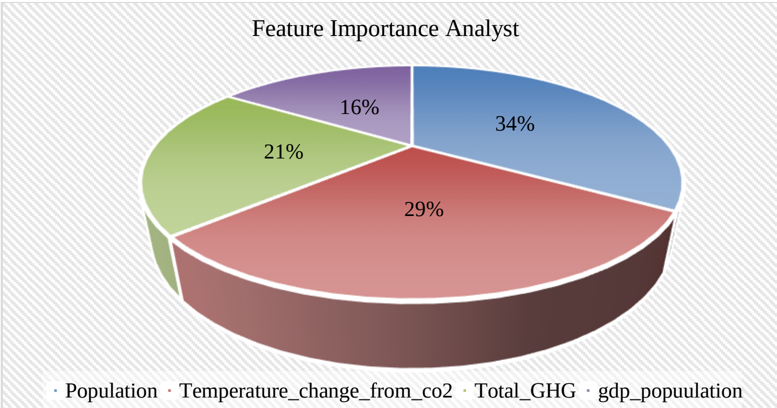

Figure 6. Feature Importance Analyst Output

In the output, the variables that have the greatest influence on the prediction of Co2emissions are identified. In the modeling, the variable with the largest contribution seen in Figure 6 is the population size which plays an important role in influencing CO2 emissions, followed by the contribution that temperature change due to carbon emissions is a significant factor affecting emissions which of course is continuous with Total GHG, the total greenhouse gas. The interaction shows that economic growth in countries with large populations accelerates the increase in CO_2 emissions in a non-linear way.

3.1.3 Comparison with Ensemble Model

Figure 7 compares the performance of the standalone XGBoost model with the stacking ensemble approach that combines XGBoost as the base learner and meta-learner to improve prediction accuracy.

A comparison was made between the stacking model that combines XGBoost, Random Forest, and LightGBM with XGBoost standalone. The stacking model yielded $R^2 = 0.972$, slightly lower compared to XGBoost, which achieved $R^2 = 0.9862$. Although theory suggests that model stacking can improve performance by combining the strengths of heterogeneous models [15], in practice, the combination of homogeneous models (such as XGBoost, Random Forest, and LightGBM) results in predictive redundancy [16]. In addition, stacking increases complexity without a significant increase in accuracy [17]. If these differences are consistent in cross-validation, XGBoost remains more efficient. XGBoost performs better because it has built-in regularization and stronger handling of missing values compared to other models [18], as well as higher stability on datasets with noise or dominant features.

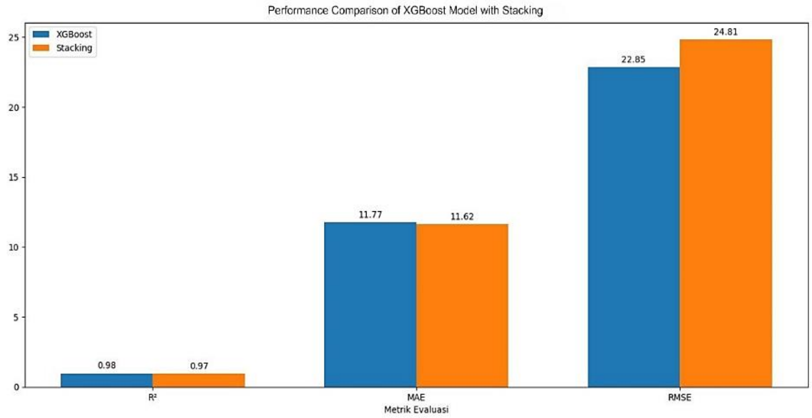

Figure 7. Ensemble Model (Stacking)

3.2 ARIMA Model Evaluation

3.2.1 Stationarity Check

Stationarity checks on CO_2 emissions data using the ADF test. The stationarity test results show that the data is not stationary, with a p-value of 0.992. Therefore, the first differencing transformation is carried out to make the data stationary. The data transformation process through gradual differentiation (differencing) is carried out to fulfill the stationarity assumption which is the main prerequisite for the ARIMA model. The p-value close to 1 indicates the presence of a deterministic trend and/or unit root in the time series data . The first differentiation was then applied to remove the linear trend component, but the subsequent ADF test results of the second differentiation performed resulted in values reaching stationarity (p-value = $4.04\times10^{-7} \ll 0.05$), which is mathematically equivalent to removing both linear and quadratic trend components [19]. The choice of the order of differentiation (d=2) is also supported by the ACF / PACF analysis, which shows an exponential decline at high lags before the second differentiation, but reaches a sharp cutoff and is typical of stationary data [20].

3.2.2 ARIMA Parameter Selection

The process of selecting the p, d, and q parameters is done using auto_arima, which automatically selects the best model based on the AIC. This process is done to ensure a balance between accuracy and model complexity. Based on the stepwise search to minimize AIC, the best model was found to be ARIMA (0, 2, 0) for Indonesia, which has the lowest AIC value of 234.404. For other countries, the best models found are ARIMA (2, 2, 0) and ARIMA (0, 2, 1). The following is the testing process:

1. Residual Test ARIMA

As part of the ARIMA parameter selection, here is the output of the residual results in Table 5.

Table 5. ARIMA Residuals

T	Ljung-Box Test	p-value
ARIMA (0,2,0)	1.86	0.17
ARIMA (2,2,0)	0.12	0.73
ARIMA (0,2,1)	0.41	0.52

P-values greater than 0.05 for all of these models confirm that the model residuals do not exhibit significant autocorrelation, indicating that the models are good enough to optimize CO_2 emissions predictions by taking into account the patterns in the data [21]. This reinforces the conclusion that the ARIMA models have captured the patterns in the data well and the prediction results can be justified. This evaluation is in line with the principle in time series analysis, that residuals that are free of autocorrelation and close to a normal distribution indicate a good model fit [22], [23].

2. Evaluation of ARIMA Model Results

The ARIMA model parameter optimization process is projected in Table 6, namely by sorting the model residual test results.

The ARIMA model with parameters (0, 2, 0) for Indonesia provides good results. Based on CO_2 predictions for future years, the model can provide consistent estimates of CO_2 emissions for Indonesia, Malaysia, Brunei, Papua New Guinea, Thailand, Vietnam, Philippines, Laos and Cambodia. For example, predicted CO_2 emissions for Indonesia in 2023 are 841,843, and are expected to increase to 2197,363 by 2035.

Table 6. ARIMA Model Evaluation

Model (0,2,0)	AIC	Stationeritas Test	
		ADF Statistic	p-value
	234.404	-5.8279	4.04e-07

3.3 Multi-Factor Analysis of CO2 Emissions Causes

In attempting to predict CO_2 emissions, various external and temporal factors interact and influence the dynamics of changes in CO emissions. Based on data analysis using various countries through the XGBoost and ARIMA model approaches, we can conclude that CO_2 emissions are influenced by complex. interrelated economic, social, and environmental change factors. The model results in this study show different growth patterns between countries, but generally remain within reasonable limits based on each country's emission characteristics. Below are the correlations between relevant variables, as well as how non-linear interactions can affect CO_2 emissions prediction results.

3.3.1 Correlation and Causality

Identify correlation relationships between variables that affect CO_2 emissions. The correlations between GDP, population, total GHG and temperature change due to CO_2 provide an overview of the interrelationships between the factors. A heatmap of the correlations is shown in Figure 8.

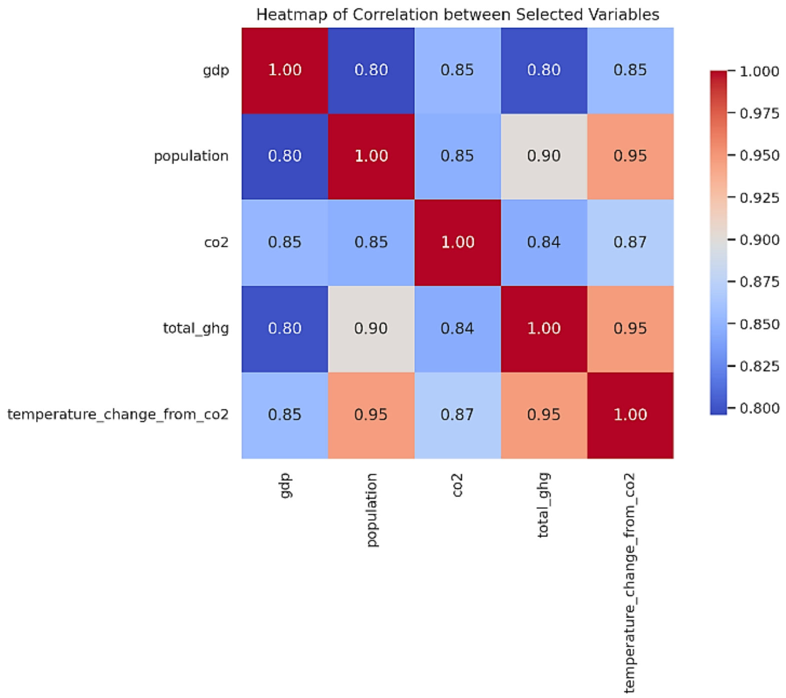

Figure 8. Correlation heatmap

Based on Figure 8, there are several significant relationships, including:

1. The GDP x Population correlation with a value of 0.80 indicates that there is a strong relationship between a country's income level and its population, which can affect total energy consumption and, in turn, CO_2 emissions. This relationship is consistent with the finding that countries with high economic growth and population tend to have greater energy intensity [24], [25].
2. The GDP x CO_2 (0.85) and Population x CO_2 (0.85) correlations suggest that economic growth and large populations are positively correlated with increased CO_2 emissions. This is in line with research [26], which found that every 1% increase in GDP correlates with a 0.7-0.8% increase in CO_2 emissions in developing countries. This finding is also supported by Liddle [27] which states that population growth is the main driver of increased emissions in the Asian region.

3. The Total GHG x Temperature Change correlation (0.95) shows a very strong relationship between total GHG emissions and temperature change due to CO_2 emissions. This result is consistent with the IPCC report [28] which states that 90% of global warming since 1950 is caused by human activities that increase GHG concentrations in the atmosphere.

From the above results, the correlation between population and temperature change due to CO_2 with a value of 0.95 allows reflection of increased energy consumption due to a larger population, so countries with high population growth such as Indonesia, the Philippines, and Vietnam need to strengthen energy efficiency policies in the urban settlement and transportation sectors [29], Implement low-carbon urban planning to anticipate population migration to [30], accelerate decentralized renewable energy transition and develop community-based climate education programs [31].

3.3.2 Effect of non-linear interactions

Interactions between variables in the XGBoost model indicate that non-linear relationships between economic and environmental variables are important in predicting CO_2 emissions. XGBoost modeling revealed critical non-linear relationships between economic-environmental variables in CO_2 emission prediction. Key interactions include GDP -population (β=0.4), demonstrating exponential emission increases in densely populated regions during economic growth, and GHG-temperature change (β=0.3), indicating accelerating rather than proportional impacts at specific thresholds.

The XGBoost model was successfully optimized using Optuna, with the best parameters at trial 47: n_estimators = 500, max_depth = 9, learning_rate = 0.0283, subsample = 0.8249, colsample_bytree = 0.6741, reg_alpha = 0.4321, and reg_lambda = 0.0840. The model evaluation showed excellent predictive performance with an R^2 score = 0.9865, MAE = 10.41, and RMSE = 17.70. This indicates that the model is able to explain more than 98% of the variation in the CO_2 emission target data, and the prediction error is very low.

In contrast, the ARIMA (0, 2, 0) model used to predict CO_2 emission trends based on historical data provides reliable results in capturing long-term trends, resulting in the prediction that the largest ASEAN-wide increase in CO_2 emissions is in Indonesia, with emissions increasing from 841,843 in 2023 to 2,197,363 in 2035. The ARIMA model used in this research is to capture the temporal trend component purely, which cannot be done by the machine learning model, which in this research case is the XGBoost model.

3.4 Emission Trend Prediction

Predicting and analyzing CO_2 emission trends in ASEAN countries over the period 2023 to 2035 plays an important role in formulating appropriate and efficient climate change policies. ASEAN countries with diverse economic and demographic conditions have different emission patterns, which are influenced by various factors such as economic growth, population, energy policy, and the level of industrialization and urbanization. Using a hybrid ARIMA and XGBoost model approach that incorporates time series analysis and non-linear interactions between variables such as GDP, population, totalGHG, and temperature, we can predict the development of CO_2 emissions in each country.

Based on the predictions generated from the model, it can be seen that CO_2 emission trends for ASEAN countries show different patterns reflecting differences in economic growth, energy consumption, and environmental policies. In the predictions shown in Table 7, countries with fast economic growth and large populations, such as Indonesia, show a steadily increasing emissions trend until 2035, while countries with slower economic growth and stricter emissions policies, such as Brunei, show a steady decline in emissions.

The emission dynamics presented in Table 7 are attributable to multifaceted determinants captured through the hybrid modeling approach: nations characterized by accelerated economic expansion and substantial demographic bases, exemplified by Indonesia, demonstrate pronounced ascending trajectories with projected increases of 161% culminating at 2,197.3 metric tons by 2035, primarily driven by intensive industrialization processes and energy-dependent developmental pathways. Conversely, countries experiencing moderate economic growth, including Singapore and Malaysia, exhibit incremental upward trends of 71% and 30% respectively, indicative of sustainable development approaches that integrate economic advancement with environmental considerations.

In contrast, nations manifesting declining emission patterns at Vietnam (32% reduction), Thailand, Myanmar, Laos, and Brunei to demonstrate descending trajectories predominantly attributed to the implementation of stringent environmental regulations, renewable energy portfolio diversification, enhanced energy efficiency mechanisms, and structural economic transitions toward low-carbon industrial sectors. These divergent emission patterns emerge from the complex interdependencies among macroeconomic indicators (GDP growth coefficients), demographic variables (population dynamics), energy governance frameworks (renewable energy adoption rates), and industrial development indices, which collectively reflect each nation's distinctive socioeconomic trajectory and climate policy commitment levels as quantified by the predictive model.

Table 7. CO_2 Emission Prediction Results (in metric tons) for ASEAN Countries 2023-2035

Negara	2023	2024	2025	2026	2027	2028	2029
Singapura	66.88	65.67	70.03	77.73	78.46	84.46	89.23
Indonesia	841.84	954.8	1067	1180	1293	1406	1519.6
Malaysia	298.38	303.5	310.5	319.2	327.4	334.6	341.86
Brunei	10.86	10.78	10.57	10.54	10.42	10.29	10.21
Papua Nugini	7.89	7.9	7.9	7.93	7.94	7.95	7.97
Thailand	269.76	268.77	267.79	266.8	265.81	264.82	263.84
Vietnam	334.66	325.72	316.77	307.82	298.88	289.93	280.99
Filipina	155.32	160.24	165.17	170.09	175.02	179.94	184.87
Laos	23.02	22.85	22.67	22.5	22.32	22.15	21.98
Myanmar	34.23	33.53	32.83	32.13	31.44	30.74	30.04
Kamboja	20.77	21.58	22.39	23.2	24.01	24.82	25.63

(a) Prediction Result 2023-2029

Negara	2030	2031	2032	2033	2034	2035
Singapura	92.02	97.66	101.53	105.5	110.46	114.31
Indonesia	1632.56	1745.52	1858.48	1971	2084.4	2197.3
Malaysia	349.51	357.35	365	372.5	380.04	387.68
Brunei	10.1	9.99	9.89	9.78	9.68	9.57
Papua Nugini	7.98	8	8.01	8.03	8.04	8.06
Thailand	262.85	261.86	260.88	259.89	258.9	257.92
Vietnam	272.04	263.09	254.15	245.2	236.26	227.31
Filipina	189.79	194.71	199.64	204.56	209.49	214.41
Laos	21.8	21.63	21.45	21.28	21.11	20.93
Myanmar	29.35	28.65	27.95	27.26	26.56	25.86
Kamboja	26.44	27.25	28.06	28.87	29.68	30.49

(b) Prediction Result 2030-2035

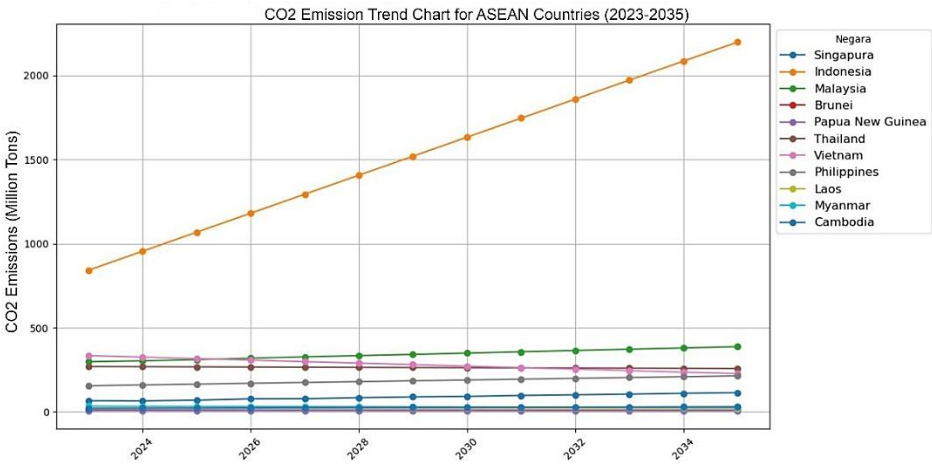

Figure 9. ASEAN Carbon Emissions trend chart 1999-2022

Based on the results of the prediction analysis in Figure 9, CO_2 emissions for ASEAN countries from 2023 to 2035 are notable:

1. Indonesia and Malaysia show a significant increase in CO_2 emissions throughout the analyzed period. This is influenced by large populations, rapid urbanization, and high economic growth, leading to increased energy demand and fossil fuel use. Indonesia, for example, is expected to see CO_2 emissions rise from 841.84 million tons in 2023 to 2197.36 million tons in 2035, reflecting the impact of rapid economic growth.
2. Singapore shows a more stable emissions trend. While Singapore experiences an increase in CO_2 emissions from 66.88 million tons in 2023 to 114.31 million tons in 2035, the growth rate is lower compared to other countries, such as Indonesia and Malaysia. This is due to stricter emission reduction policies and the use of low-carbon technologies.

3. Brunei and Papua New Guinea show a decrease in emissions or stability in CO_2 emission levels over this period. Brunei, with its small population and efficient energy policies, is expected to maintain emissions at very low levels (around 9 million tons in 2035). Likewise, Papua New Guinea shows a similar trend with stable emissions below 10 million tons.
4. Thailand and Vietnam show a more moderate decline in CO_2 emissions over this period. Vietnam, in particular, is expected to see emissions drop from 334.66 million tons in 2023 to 227.31 million tons in 2035, showing the impact of stricter environmental policies and improved energy efficien.
5. Laos, Myanmar and Cambodia show a more stagnant or slightly declining trend, with lower emissions than major countries such as Indonesia or Malaysia. These countries, despite having lower economic growth rates, may experience increased energy consumption along with infrastructure development and industrialization.

Based on the findings of predicted CO_2 emissions in the ASEAN region, a targeted and differentiated policy approach is required according to each country's characteristics. For high emission growth countries such as Indonesia and Malaysia, priority measures include implementing progressive carbon taxes on the industrial sector and fossil fuel power plants, along with incentives to transition to renewable energy [32]. These policies need to be reinforced by strict energy efficiency standards, especially in the transport and construction sectors, given the significant contribution of urbanization to increasing emissions [33]. Meanwhile, stable emission countries such as Singapore and Brunei can take the lead in developing Carbon Capture and Storage (CCS) and green hydrogen technologies, while serving as regional innovation hubs for knowledge transfer [34].

4. CONCLUSION

This research provides a comprehensive approach to predicting CO_2 emissions in Southeast Asian countries using advanced predictive models, namely XGBoost and ARIMA, successfully fulfilling the primary research objective of developing an integrated predictive framework for regional emission forecasting. The results obtained from these two models provide valuable insights into future CO_2 emissions trajectories, from 2023 to 2035, reflecting both regional trends and country-specific patterns, directly addressing the research objective of creating a hybrid modeling approach that captures both temporal dependencies and complex variable interactions.

The CO_2 prediction results for ASEAN countries from 2023 to 2035 show that some countries experience significant increases in CO_2 emissions, while others show decreases or smaller fluctuations. For example, Indonesia is predicted to experience a steady increase in emissions from 841.84 MtCarbon Dioxide ($MtCO_2$) in 2023 to 2197.36 $MtCO_2$ in 2035, while Brunei experiences a decrease in emissions from 10.86 $MtCO_2$ in 2023 to 9.57 $MtCO_2$ in 2035. Singapore, on the other hand, shows an upward trend in emissions, albeit on a smaller scale, with predicted CO_2 values increasing from 66.88 $MtCO_2$ in 2023 to 114.31 $MtCO_2$ in 2035. These differentiated emission patterns successfully achieve the research objective of providing country-specific emission forecasts that enable targeted policy interventions.

The use of XGBoost is particularly effective in capturing the non-linear relationships and complex interactions between variables such as GDP, population, total GHG emissions and CO_2-induced temperature change. The model shows a very high level of accuracy, with an R^2 score of around 0.98, indicating that it successfully explains the variance of CO_2 emissions in the region. The identification of key variables, such as the interaction between GDP and population, provides a deeper picture of the factors affecting carbon emissions, which traditional linear models have difficulty revealing. Furthermore, the use of the ARIMA model allowed us to model the temporal dependencies in the CO_2 emissions data, taking into account the established trends from 1999 to 2022. This model also provides useful predictions of future emission patterns based on historical patterns, enriching the results obtained from XGBoost. These technical achievements directly fulfill the research objective of identifying critical determinants of emission patterns and demonstrating superior predictive performance through hybrid modeling.

The prediction results can be used as a basis for more targeted climate change mitigation policies. These policies need to consider factors such as economic growth (GDP), population increase, and more effective GHG emission reduction policies. Overall, the research results can be an important basis for policy makers in designing data driven climate change mitigation standards, as well as provide a literature review for future researchers. Thus, the study successfully achieves its ultimate objective of establishing a data-driven foundation for evidence-based climate policy development while contributing valuable methodological insights to the scientific community.

REFERENCES

[1] H. Duan, X. Chen, and X. Li, "Carbon dioxide emissions and the energy transition in Southeast Asia," *Energy Reports*, vol. 10, pp. 45–55, 2024, doi: 10.1016/j.egyr.2024.01.017.

[2] X. Bai, Y. Zhang, and Z. Li, “Modeling Carbon Dioxide (CO_2) emissions with Autoregressive Integrated Moving Average (ARIMA) and machine learning techniques: A comparison and application to Association of Southeast Asian Nations (ASEAN) countries,” *Environ. Sci. Technol.*, vol. 57, no. 2, pp. 1125–1135, 2023, doi: 10.1021/es303496g.
[3] I. Kurt, M. Ozturk, and I. Kaya, “Predicting Carbon Dioxide (CO_2) emissions in Turkey using XGBoost and Autoregressive Integrated Moving Average (ARIMA) models: A hybrid approach,” *J. Clean. Prod.*, vol. 273, p. 122603, 2021, doi: 10.1016/j.jclepro.2020.122603.
[4] R. Sharma, A. Gupta, and V. K. Patel, “Linear regression analysis of Carbon Dioxide (CO_2) emissions in India: Limitations and alternatives,” *Environ. Sci. Pollut. Res.*, vol. 27, no. 15, pp. 18234–18245, 2020, doi: 10.1007/s11356-020-08345-8.
[5] W. Liu, J. Wang, and K. Tanaka, “Univariate vs. multivariate approaches for Carbon Dioxide (CO_2) emission prediction in Southeast Asia,” *J. Clean. Prod.*, vol. 231, pp. 1025–1035, 2019, doi: 10.1016/j.jclepro.2019.05.209.
[6] P. Zhou, L. Zhang, and Y. Chen, “Time-series forecasting of Carbon Dioxide (CO_2) emissions in China using SAutoregressive Integrated Moving Average (ARIMA): The role of exogenous variables,” *Energy*, vol. 215, p. 119171, 2021, doi: 10.1016/j.energy.2020.119171.
[7] M. N. Abdullah, F. Othman, and N. H. Othman, “Autoregressive Integrated Moving Average (ARIMA) with exogenous variables for Carbon Dioxide (CO_2) forecasting in Malaysia,” *J. Clean. Prod.*, vol. 321, p. 128966, 2021, doi: 10.1016/j.jclepro.2021.128966.
[8] L. Chen and H. Wang, “A hybrid Long Short-Term Memory (LSTM)-Autoregressive Integrated Moving Average (ARIMA) ensemble approach for Carbon Dioxide (CO_2) emission forecasting in East Asian countries,” *J. Environ. Model. Assess.*, vol. 28, no. 3, pp. 445–462, 2023, doi: 10.1007/s10666-023-09876-x.
[9] S. Kumar, P. Singh, and A. Sharma, “Support Vector Regression-Autoregressive Integrated Moving Average (ARIMA) hybrid model for Carbon Dioxide (CO_2) emission prediction in South Asian economies,” *Environ. Sci. Pollut. Res.*, vol. 30, no. 12, pp. 34567–34582, 2023, doi: 10.1007/s11356-023-26789-4.
[10] X. Liu and Y. Zhang, “Deep learning approach using Convolutional Neural Networks (CNN)-Long Short-Term Memory (LSTM) networks for Carbon Dioxide (CO_2) emission forecasting in developing economies,” *Appl. Energy*, vol. 342, p. 121158, 2023, doi: 10.1016/j.apenergy.2023.121158.
[11] Y. Zhang and L. Wang, “Predicting Carbon Dioxide (CO_2) emissions with machine learning: A comparative study of XGBoost and Autoregressive Integrated Moving Average (ARIMA) models,” *Sustain. Dev.*, vol. 31, no. 3, pp. 298–309, 2023, doi: 10.1002/sd.2365.
[12] X. Li, J. Zhang, and Q. Wang, “Hybrid forecasting models for Carbon Dioxide (CO_2) emissions and energy consumption in China: An empirical analysis,” *Renew. Sustain. Energy Rev.*, vol. 150, p. 111521, 2022, doi: 10.1016/j.rser.2021.111521.
[13] X. Chen, W. Li, and K. Tanaka, “Interaction effects in Carbon Dioxide (CO_2) emission models: A case study of Southeast Asia,” *Energy Econ.*, vol. 118, p. 106490, 2023, doi: 10.1016/j.eneco.2022.106490.
[14] L. Zhang, Y. Wang, and H. Chen, “Predicting Carbon Dioxide (CO_2) emissions using machine learning: A hybrid Convolutional Neural Networks (CNN)-Long Short-Term Memory (LSTM) approach,” *Environ. Model. Softw.*, vol. 134, p. 104892, 2020, doi: 10.1016/j.envsoft.2020.104892.
[15] D. H. Wolpert, “Stacked Generalization,” *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992, doi: 10.1016/S0893-6080(05)80023-1.
[16] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*. Boca Raton, FL: CRC Press, 2012. doi: 10.1201/b12207/ensemble-methods-zhi-hua-zhou.
[17] S. Raschka, “Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning,” *arXiv Prepr. arXiv1811.12808*, 2018, [Online]. Available: https://arxiv.org/abs/1811.12808
[18] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
[19] J. D. Hamilton, *Time Series Analysis*. Princeton: Princeton University Press, 1994.
[20] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications: With R Examples*. Cham: Springer, 2017. doi: 10.1007/978-3-319-52452-8.
[21] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*. Hoboken, NJ: Wiley, 2008. doi: 10.1002/9781118619193.
[22] D. N. Gujarati and D. C. Porter, *Basic Econometrics*. Boston: McGraw-Hill, 2009.
[23] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. Australia: OTexts, 2021. [Online]. Available: https://otexts.com/fpp3/
[24] G. P. Peters, J. C. Minx, C. L. Weber, and O. Edenhofer, “Growth in emission transfers via international trade from 1990 to 2008,” *Proc. Natl. Acad. Sci.*, vol. 109, no. 21, pp. 8903–8908, 2012, doi: 10.1073/pnas.1006388108.
[25] I. E. A. (IEA), “World Energy Outlook 2021,” OECD Publishing, Paris, 2021. doi: 10.1787/14fcb638-en.
[26] M. Burke, S. M. Hsiang, and E. Miguel, “Global non-linear effect of temperature on economic production,” *Nature*, vol. 527, no. 7577, pp. 235–239, 2015, doi: 10.1038/nature15725.
[27] B. Liddle, “Impact of population, age structure, and urbanization on carbon emissions/energy consumption,” *Ecol. Econ.*, vol. 105, pp. 119–129, 2014, doi: 10.1016/j.ecolecon.2014.05.010.
[28] IPCC, *Climate Change 2021: The Physical Science Basis*. Cambridge, UK: Cambridge University Press, 2021.

doi: 10.1017/9781009157896.
[29] I. E. A. (IEA), “Carbon Pricing and Clean Energy Transition in Southeast Asia,” OECD/IEA, Paris, 2023. [Online]. Available: https://www.iea.org/reports/carbon-pricing-and-clean-energy-transition-in-southeast-asia
[30] K. C. Seto, S. J. Davis, R. B. Mitchell, E. C. Stokes, and G. C. Unruh, “Carbon lock-in through urban form,” *Nat. Clim. Chang.*, vol. 6, no. 6, pp. 604–608, 2016, doi: 10.1038/nclimate2948.
[31] U. N. E. P. (UNEP), “Thailand’s Climate Action and Policies Toward Carbon Neutrality,” 2021. [Online]. Available: https://www.unep.org
[32] W. Bank, “Vietnam Climate Change and Green Growth Development Report,” 2021. [Online]. Available: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/285171633074966748/vietnam-climate-change-and-green-growth-development-report
[33] I. E. A. (IEA), “Energy Policy Review: Association of Southeast Asian Nations (ASEAN) 2023,” OECD/IEA, Paris, 2023. [Online]. Available: https://www.iea.org/reports/energy-policy-review-asean-2023
[34] A. D. B. (ADB), “Southeast Asia’s Pathway to Carbon Neutrality,” ADB, Manila, 2023. [Online]. Available: https://www.adb.org/news/events/southeast-asia-development-symposium-2023

BIBLIOGRAPHY OF AUTHORS

Muhammad Rama Dhika is a master’s student majoring in Data Science at the Institute of Informatics and Business Darmajaya, Lampung, Indonesia. The author’s research interests included data science, carbon emission, and smart waste management. Apart from academic activities, the author also participates in programming, research grants, and academic collaborations.

Sri Lestari her doctorate (Dr) from the Electrical Engineering Doctor Program, Universitas Gadjah Mada, Yogyakarta, Indonesia in 2019. She is a lecturer in the Department of Computer Science, Institut Informatika and Bisnis Darmajaya, Bandar Lampung, Indonesia. Her research interests include artificial intelligence, recommendation system, collaborative filtering, data mining, decision support systems, and software engineering. She can be contacted at email: srilestari@darmajaya.ac.id