Technologique: A Global Journal on Technological Developments and Scientific Innovations
2025, Vol. 5, No.1, 41 – 62 | ISSN Online: 3028-1415 | ISSN Print: 3028-1407
DOI: https://doi.org/10.62718/vmca.tech-gjtdsi.5.1.SC-0825-004

Machine Learning–Driven Framework for Sustainable Development Assessment in the Philippines

Article History:
Received: 29 July 2025
First Decision: 06 August 2025
Published: 18 September 2025

Dave Emmanuel V. Nuada[1], ORCID No. 0009-0008-7430-883X
Marvin O. Mallari[2], ORCID No. 0000-0002-6662-5105
[1]Master of Science in Computer Engineering, Polytechnic University of the Philippines, Sta. Mesa, Manila, Philippines
[2]Lecturer, Computer Engineering Graduate School, Polytechnic University of the Philippines, Sta. Mesa, Manila, Philippines

Abstract

This study develops a machine learning-optimized framework for evaluating socioeconomic and environmental factors to enhance sustainable development in the Philippines. As one of the most climate-vulnerable nations, the country faces compounding challenges, including economic inequality, environmental degradation, and resource scarcity. These issues require data-driven strategies that integrate environmental, economic, and social dimensions. This research applies Random Forest and Gradient Boosting ensemble algorithms to overcome the limitations of traditional sustainability evaluation methods based on linear analyses. By capturing nonlinear relationships, handling multicollinearity, and processing high-dimensional datasets, the proposed framework identifies and prioritizes key sustainability indicators that most significantly influence policy and technology outcomes. The study utilizes historical datasets from national and international sources, encompassing metrics such as GDP growth, energy production, CO_2 emissions, HDI, Gini coefficient, unemployment rates, and access to clean water. Model performance was evaluated using R^2 and RMSE, with results of $R^2 = 0.915534$ and RMSE = 0.566654, demonstrating strong predictive validity. Findings reveal that the integration of machine learning enables the synthesis of complex datasets into actionable, forward-looking insights that can guide evidence-based policy formulation, resource allocation, and sustainable technology adoption. These insights support balanced decision-making that fosters environmental stewardship, economic resilience, and social equity. Beyond the Philippine context, the framework offers scalability and adaptability for application in other developing nations confronting similar sustainability challenges. The study contributes to the global discourse on leveraging artificial intelligence for sustainable development by demonstrating its capacity to support the United Nations Sustainable Development Goals (SDGs) through enhanced data-driven governance and resilient socio-environmental systems.

Keywords: Machine Learning, Ensemble Learning, Sustainability Assessment, Socioeconomic Modelling, Environmental Indicators, Sustainable Development Goals, Policy Evaluation Random Forest, Gradient Boosting

INTRODUCTION

Quantitative evaluation of socioeconomic and environmental sustainability indicators through machine learning is central to advancing sustainable development, especially in countries facing climate and inequality challenges. In the Philippines, renowned for its biodiversity yet highly vulnerable to climate change, machine learning (ML) offers a means to optimize these metrics and strengthen sustainability initiatives. This study applies ML algorithms to refine technology evaluations, integrating environmental, economic, and social dimensions into a predictive framework.

While sustainability evaluation metrics exist, many current approaches cannot synthesize complex, multidimensional datasets. Traditional methods emphasize linear relationships and static indicators, failing to capture the nonlinear dynamics of sustainable development (Zhang et al., 2025; Athey, 2018; Biau & Scornet, 2016). ML addresses these limitations by modeling nonlinear patterns, managing incomplete data, and revealing hidden relationships, thereby supporting evidence-based decisions (Cuevas, 2017; Sala et al., 2015; Vinuesa et al., 2020). In the Philippine context, existing frameworks are largely descriptive, lacking predictive capacity to anticipate the combined effects of inequality, infrastructure gaps, and shifting environmental conditions (Lasco & Pulhin, 2006; Barcelon et al., 2023). This research responds by building a system capable of learning from historical data to deliver forward-looking evaluations for climate adaptation and economic planning (Sala et al., 2015; Han et al., 2021).

The study focuses on tree-based ensemble algorithms such as Random Forest and Gradient Boosting, chosen for their ability to handle multicollinearity, nonlinearity, and variable interactions (Biau & Scornet, 2016). These models surpass linear approaches in accuracy and interpretability, offering feature importance rankings and partial dependence analyses to identify critical drivers of SDG-related metrics (Athey, 2018). This study hypothesizes that ensemble models will achieve predictive accuracy (R^2) exceeding 0.85 for SDG Index prediction, significantly outperforming linear regression methods despite the limited sample size of annual national-level data.

Integrating ML into sustainability assessment aligns with the United Nations' call for data-driven strategies in achieving the Sustainable Development Goals (SDGs). Globally, AI and ML are increasingly recognized as catalysts for policy innovation, resource optimization, and resilience-building (Vinuesa et al., 2020; OECD, 2022). This study contributes to that discourse while tailoring solutions to Philippine realities, creating a model adaptable for other developing nations facing similar vulnerabilities.

The primary objective is to develop and validate a machine learning framework that accurately predicts SDG outcomes in the Philippine context using socioeconomic and environmental indicators. Specifically, the research aims to: (1) identify and prioritize metrics most influential to sustainability outcomes; (2) integrate datasets capturing environmental, economic, and social interactions; (3) select the best predictive model based on R^2 or RMSE; (4) develop a comprehensive evaluation framework; and (5) deliver actionable insights for policymaking and sustainable technology adoption.

Recent national indicators underscore the importance of such work:

- Unemployment rose to 4.0% in April 2024, up from 3.1% in December 2023 (Philippine Statistics Authority, 2024).

- Inflation slowed to 3.7% in June 2024 (Philippine Statistics Authority, 2024).

- The Department of Energy projects 4,164.92 MW of added capacity in 2024 from renewable and conventional sources (Department of Energy, 2024).

- Forty million Filipinos remain without access to potable water (Domingo, 2024).

- GDP reached $404.28 billion USD in 2022, with 5.5% growth in Q2 2025 (Philippine Statistics Authority, 2025).

- Climate change could shrink GDP by up to 13.6% by 2040 if unmitigated (European Climate Adaptation Platform, 2024).

Environmental pressures continue to mount. CO_2 emissions climbed to 150.40 million tons in 2024, a 5.36% increase (Ritchie et al., 2020), while the country ranks third globally in climate vulnerability (Sustainable Development Report, 2024). Social indicators show a Gini Coefficient of 40.2, reflecting moderate inequality (World Economics, 2025), and an HDI of 0.71 in 2022, up from 0.692 in 2021 (UNDP, 2024). Based on record, the Philippines ranked 92nd of 167 countries in the 2024 SDG Index with a score of 67.47 (Sustainable Development Solutions Network, 2024).

By leveraging ML, this study aims to balance technological progress with environmental stewardship, contributing to economic stability, resource sustainability, and climate resilience. It seeks to equip policymakers, industry leaders, and researchers with a framework for sustainable development that is rooted in the Philippine context yet aligned with global objectives for climate adaptation and socio-economic equity.

LITERATURES

While machine learning applications in sustainability have grown rapidly (Vinuesa et al., 2020; OECD, 2022; Zhang et al., 2025), limited research addresses the small-sample contexts typical of developing nations. Vinuesa et al. (2020) highlight AI's potential to advance over 130 SDG targets but assume data-rich environments (n > 1000), which often excludes countries with weaker statistical infrastructures. OECD (2022) also stresses that AI's benefits are unevenly distributed, with developing nations facing capacity gaps. Similarly, Zhang et al. (2025) demonstrate that reliance on linear approaches underestimates nonlinear interactions among SDGs. This study addresses that gap by demonstrating ensemble methods' effectiveness with a limited dataset (n = 23).

Globally, AI has been applied in climate modelling, emissions forecasting, and energy optimization (OECD, 2022; Vinuesa et al., 2020), but most studies favour well-resourced settings. By contrast, developing-country applications illustrate both promise and challenges. Mohata et al. (2024) applied regression models to state-level SDG data in India, where Random Forest achieved $R^2 = 0.727$, while Rainford et al. (2021) used Random Forest to estimate soil organic carbon in Colombia, explaining 50.28% of variance. These results provide benchmarks for ensemble learning in constrained environments and highlight the need for robust feature engineering and cautious validation.

Conflicting evidence exists on minimum sample sizes for reliable ML. Breiman (2001) suggested n > 200 for Random Forest, but more recent work shows acceptable performance in smaller datasets (Han et al., 2021; Rainford et al., 2021; Mohata et al., 2024). Han et al. (2021) demonstrated reliable results with n < 50, supporting the feasibility of this study's dataset. Such findings underscore the importance, value and requirement of testing ensemble models under limited-sample conditions, especially in countries where high-frequency data are scarce.

Traditional sustainability assessments often relied on cost-benefit analysis or expert judgment, approaches that overlook multidimensional interactions (Sala et al., 2015; Cuevas, 2017; Lasco & Pulhin, 2006). Sala et al. (2015) propose a systemic multi-criteria framework integrating environmental, economic, and social factors, emphasizing the need for dynamic, data-driven models. In the Philippine context, Cuevas (2017) identifies institutional barriers to climate adaptation, Lasco and Pulhin (2006) examine the mixed outcomes of community-based forest management, and Barcelon et al. (2023) highlight the role of human capital in driving GDP per capita. These studies collectively underscore the necessity of frameworks incorporating both socioeconomic and environmental indicators, while also revealing the limits of descriptive methods in producing forward-looking policy guidance.

Machine learning offers distinct advantages for such integration. Athey (2018) emphasizes its flexibility in high-dimensional modelling, while Biau and Scornet (2016) demonstrate the predictive accuracy and interpretability of Random Forest, particularly in capturing nonlinear and interactive effects among variables like inflation, energy use, and population growth. Recent work further supports ML's contribution to sustainability by integrating diverse datasets and revealing hidden interactions overlooked by traditional linear methods (Zhang et al., 2025; Vinuesa et al., 2020). These strengths justify the choice of ensemble algorithms in this study, applied to the Philippine context to produce a machine learning–optimized framework.

Taken together, the reviewed literature underscores both the promise and limitations of existing approaches. Global frameworks (Vinuesa et al., 2020; OECD, 2022; Zhang et al., 2025) reveal AI's transformative potential but remain biased toward data-rich contexts. Recent evidence (Mohata et al., 2024; Rainford et al., 2021; Han et al., 2021) shows ensemble learning can perform competitively in small-sample developing-country settings, directly supporting this study's dataset. Philippinefocused work (Cuevas, 2017; Lasco & Pulhin, 2006; Barcelon et al., 2023) highlights urgent institutional and socioeconomic challenges requiring robust, integrated evaluation methods. Methodological contributions (Sala et al., 2015; Athey, 2018; Biau & Scornet, 2016) emphasize the need for interpretable, nonlinear models. Building on these insights, this study advances a framework that addresses small-sample constraints while adapting global methods to the Philippine context.

METHODS

Research Design. This study employs a quantitative correlational research design using retrospective secondary data analysis with supervised machine learning modelling. These methodologies are applied to examine the relationships among socioeconomic and environmental indicators and to construct a machine learning–based framework for evaluating sustainability performance.

Each section delineates specific steps taken to ensure thoroughness and accuracy in the analysis. The overarching goal of this research is to harness the power of machine learning to significantly enhance sustainability efforts in the Philippines. By optimizing technology evaluation metrics through these methodologies, the study aims to provide valuable insights and practical solutions that contribute to environmental resilience and economic development.

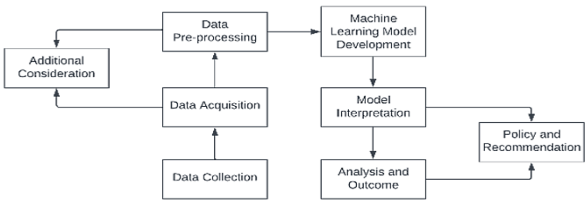


Figure 1
Methodology Diagram

This approach promises to illuminate complex relationships between technological choices and sustainability outcomes, ultimately paving the way for informed decision-making and effective policy formulation in the face of climate challenges.

Figure 1 illustrates the workflow: data collection and acquisition, preprocessing, and development and interpretation of machine learning models. The resulting outcomes are then scrutinized to guide policy formulation and recommendations.

Data Sources. This study utilizes a comprehensive set of national indicators sourced from reputable global and regional databases. The dataset comprises 253 data points (23 annual observations × 11 variables) spanning the period 2000–2022, covering key sustainability-related indicators for the Philippines. The selected indicators include Gross Domestic Product (GDP), Unemployment Rate, Inflation Rate, CO_2 Emissions, Total Energy Consumption, Clean Water Access, Total Population, Gini Coefficient, Human Development Index (HDI), and the Sustainable Development Goals (SDG) Index. These datasets form the foundation of the machine learning–based evaluation conducted in this study, supporting the analysis and modelling.

GDP data (current US$) for the Philippines was obtained from the World Bank Data. Information on unemployment rates from 1960 to 2024 was sourced from MacroTrends, which provides annual labor statistics for the country. Inflation data, specifically consumer prices expressed as an annual percentage, was accessed via the World Bank Open Data portal. For CO_2 emissions, the study referred to the "Philippines: CO_2 Country Profile" by Ritchie, Roser, and Rosado from Our World in Data.

Energy consumption statistics were taken from Enerdata's Philippines Energy Information portal, while data on access to clean water was also retrieved from MacroTrends, which tracks clean water access rates from 1960 to 2024. The total population figures were drawn from the World Bank's indicator on total population in the Philippines.

Income inequality was measured using the Gini Coefficient, with values obtained from the World Inequality Database (WID). Human Development Index (HDI) data was sourced from the United Nations Development Programme's Human

Development Reports portal, providing specific country-level information. Lastly, the Sustainable Development Goals (SDGs) Index score for the Philippines was referenced from the Sustainable Development Report 2024, which features country-level dashboards and profiles.

Ethical Considerations. This study exclusively used secondary data from reputable, publicly accessible sources (World Bank, UNDP, Our World in Data, Enerdata, MacroTrends, and the World Inequality Database). No human participants were involved, and no personally identifiable information was accessed. As the research relied solely on publicly available aggregate data, formal ethics board approval was not required. All analyses were conducted with the intention of supporting equitable, transparent, and responsible sustainability policies, and all data sources are properly cited and freely accessible to ensure compliance with ethical standards and reproducibility.

Data Analysis. The preprocessing phase involved visualizing relationships, identifying outliers, and addressing missing values, including linear interpolation for the Gini Coefficient. For model development, three supervised machine learning algorithms were used: Decision Tree, Random Forest, and Gradient Boosting.

Model performance was evaluated with a train–test split (80/20), reporting results for the hold-out set. K-fold cross-validation, though standard, was impractical for this dataset (n = 23) as folds would leave very small test sets and unstable estimates. Prior methodological studies emphasize that for sequential or small-sample data, random k-fold cross-validation can distort temporal dependencies and lead to optimistic bias, while chronological hold-out or rolling-origin validation is more appropriate (Bergmeir & Benítez, 2012; Hyndman & Athanasopoulos, 2018; Cerqueira, Torgo, & Mozetič, 2020). Accordingly, a chronological hold-out was adopted, with 2000–2017 used for training and 2018–2022 reserved for testing, and results were interpreted cautiously. To reduce overfitting, tree depth and learning rates were conservatively tuned, and ensemble methods (Random Forest and Gradient Boosting) were chosen for their robustness in small, nonlinear datasets. This ensured results reflected predictive potential while remaining transparent about small-sample limitations.

Model interpretation included analyzing feature importance and generating partial dependence plots to show the marginal effects of individual predictors. Data challenges were addressed by correcting inconsistencies, handling missing information, and examining biases that could affect fairness and accuracy. Prediction uncertainty was also assessed, and additional indicators were explored for their potential to strengthen the framework.

Statistical Analysis Software. All analyses were conducted in Python 3.12.7 (Anaconda/Jupyter). Core libraries and versions were: NumPy 1.26.4, pandas 2.2.3, scikit-learn 1.5.2, matplotlib 3.9.2, seaborn 0.13.2, and tabulate 0.9.0. A fixed random seed (42) was applied to ensure reproducibility. Complete code and environment specifications available at researchers' GitHub repository https://github.com/TimeFreeze025/.

RESULTS

Trends in Key Socioeconomic and Environmental Indicators in the Philippines. Displayed below are visual representations illustrating the patterns in Gross Domestic Product (GDP), Unemployment Rate, Inflation Rate, CO_2 Emissions, Total Energy Consumption, Clean Water Access, Total Population, Gini Coefficient, Human Development Index (HDI), and Sustainable Development Goals (SDGs) Index.

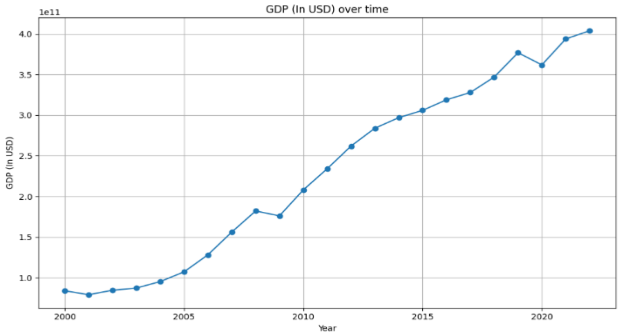


Figure 2
Line Plot of GDP Per Capita over Time in the Philippines

Figure 2 shows the Gross Domestic Product (GDP) variation in the Philippines (in USD) from 2000 to 2022. The graph illustrates a consistent upward trend, with GDP rising from approximately 80 billion USD in 2000 to over 400 billion USD by 2022. There is a substantial increase after 2010, signifying robust economic growth. The curve exhibits a slight dip around 2009, likely reflecting the global financial crisis but quickly recovers and continues its upward trajectory.

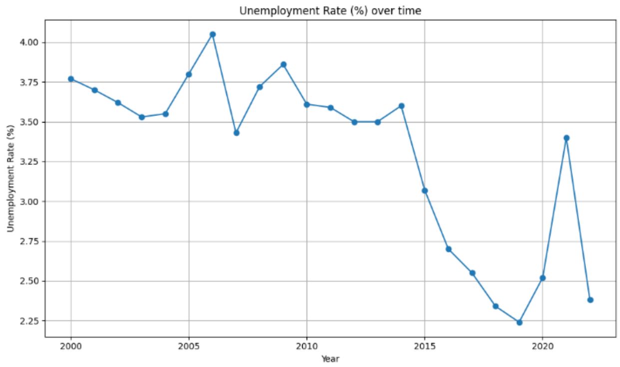


Figure 3
Line Plot of Unemployment Rate over Time in the Philippines

Figure 3 depicts the Unemployment Rate (%) in the Philippines over the same period. The graph shows considerable fluctuation, with rates generally ranging between 2.5% and 4%. Notable spikes occur around 2005 and 2020, the latter likely due to the COVID-19 pandemic. Overall, there's a gradual downward trend from 2010 to 2019, suggesting improving labor market conditions, before the sharp increase in 2020.

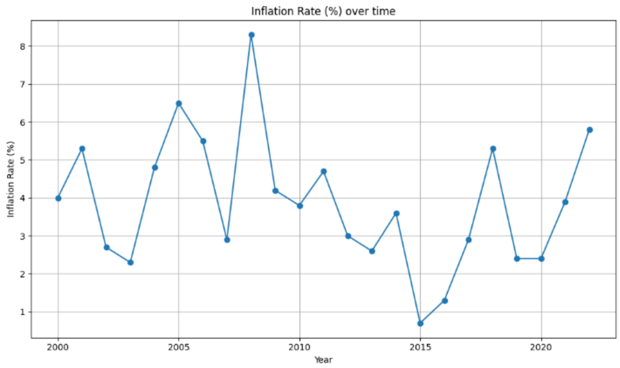


Figure 4
Line Plot of Inflation Rate over Time in the Philippines

Figure 4 shows the Inflation Rate (%) in the Philippines from 2000 to 2022, indicating notable volatility ranging between approximately 1% and 8.5%. Peaks in 2005 and 2008 coincide with global economic events, while 2015 to 2020 exhibits relatively lower and stable rates before a sharp rise in 2022.

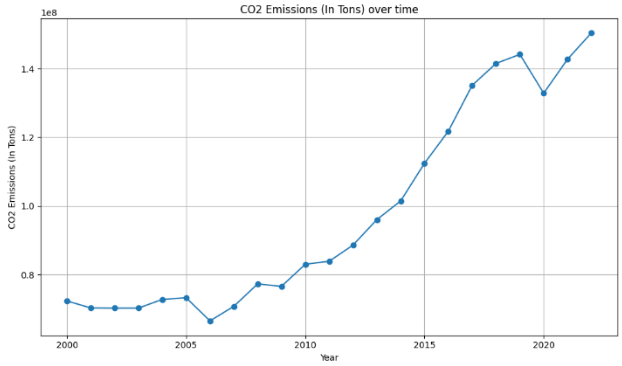


Figure 5
Line Plot of CO2 Emissions over Time in the Philippines

Figure 5 illustrates CO2 emissions (in Tons) for the Philippines over time, depicting a steady rise from approximately 70 million tons in 2000 to nearly 150 million tons in 2022. The sharpest increase follows 2010, aligning with robust GDP growth shown in the figure, with a slight downturn around 2020 possibly attributed to decreased economic activity during the COVID-19 pandemic.

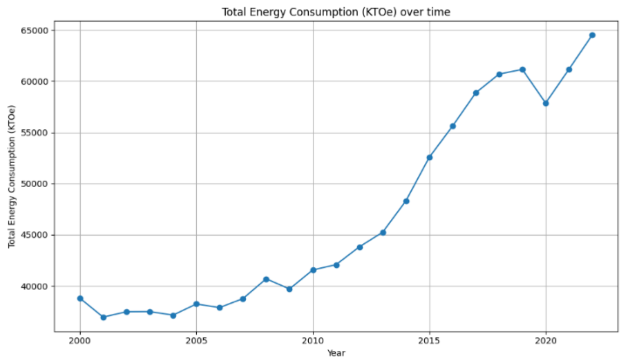


Figure 6
Line Plot of Total Energy Consumption over Time in the Philippines

Figure 6 illustrates the Total Energy Consumption (in KTOe) in the Philippines from 2000 to 2022. The graph shows a general upward trend, with consumption rising from about 38,000 KTOe in 2000 to over 64,000 KTOe by 2022. The trend is not linear, showing periods of rapid increase (particularly after 2010) and some fluctuations. This overall increase in energy consumption aligns with the country's economic growth and rising population.

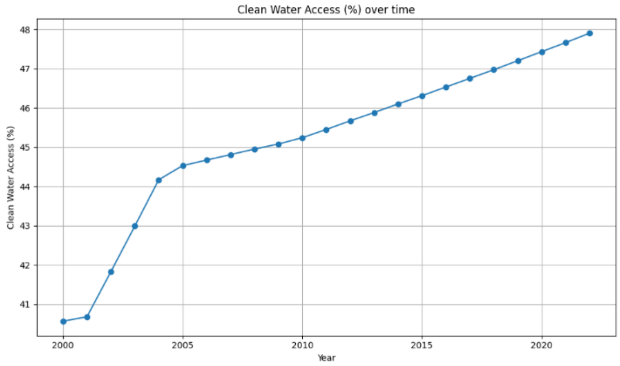


Figure 7
Line Plot of Clean Water Access over Time in the Philippines

Figure 7 presents the Clean Water Access (as a percentage) in the Philippines from 2000 to 2022. The graph shows a steady improvement, with access increasing from about 40.5% in 2000 to nearly 48% by 2022. The most rapid improvements appear to have occurred between 2000 and 2005, with a more gradual but consistent increase thereafter. This trend indicates ongoing efforts to improve water infrastructure and public health.

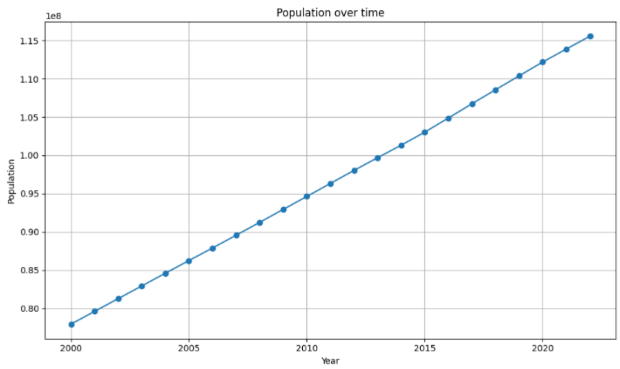


Figure 8
Line Plot of Total Population over Time in the Philippines

Figure 8 shows the Population growth in the Philippines from 2000 to 2022. The graph depicts a steady and consistent upward trend, with the population increasing from approximately 78 million in 2000 to over 115 million by 2022. This linear growth pattern suggests a stable birth rate and/or net migration, potentially putting pressure on resources and infrastructure.

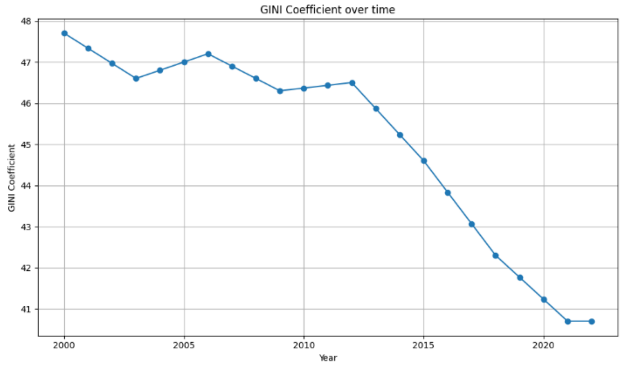


Figure 9
Line Plot of Gini Coefficient (GC) over Time in the Philippines

Figure 9 illustrates the GINI Coefficient trends in the Philippines from 2000 to 2022. The graph shows a notable downward trend, particularly after 2015. Starting from a high of about 47.8 in 2000, the coefficient decreases to approximately 40.7 by 2022. This decline indicates a reduction in income inequality over the period, suggesting that economic growth may have been accompanied by more equitable distribution of wealth.

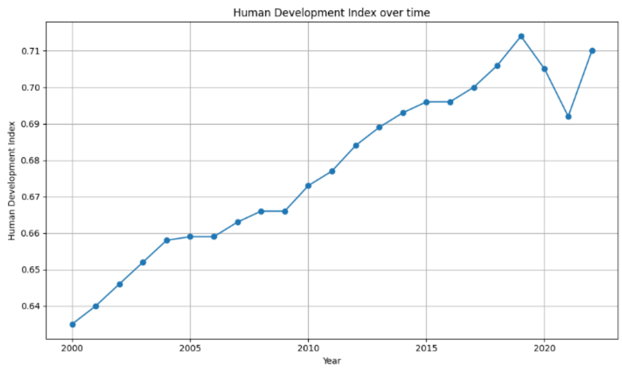


Figure 10
Line Plot of Human Development Index (HDI) over Time in the Philippines

Figure 10 presents the Human Development Index (HDI) for the Philippines from 2000 to 2022. The graph shows a general upward trend, with the HDI rising from about 0.635 in 2000 to around 0.71 in 2022. This increase reflects improvements in key areas such as education, health, and standard of living. However, the graph also shows some fluctuations, particularly a sharp drop around 2020, possibly due to the global COVID-19 pandemic, followed by a recovery.

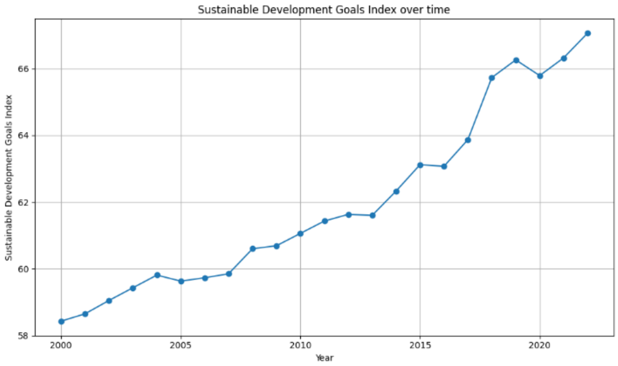


Figure 11
Line Plot of Sustainable Development Goals (SDGs) Index Score over Time in the Philippines

Figure 11 depicts the Sustainable Development Goals (SDG) Index for the Philippines from 2000 to 2022. The graph shows a clear upward trend, with the index rising from about 58.5 in 2000 to approximately 67.5 by 2022. This increase suggests steady progress towards achieving the SDGs, with particularly rapid improvements observed after 2015, coinciding with the formal adoption of the SDGs by the United Nations.

Visual Representation of Key Sustainable Development Indicators in the Philippines*.* To gain a comprehensive understanding of the key indicators influencing sustainable development in the Philippines, it is essential to first explore the descriptive statistics of the dataset used in this study.

Table 1
Descriptive Statistics of the Dataset

	Year	GDP (US$)	Unemployment Rate (%)	Inflation Rate (%)	CO2 Emissions (In Tons)
count	23	23	23	23	23
mean	2011	2.30432×10^{11}	3.30565	3.86522	9.80359×10^{7}
std	6.78233	1.13313×10^{11}	0.554033	1.75593	2.97617×10^{7}
min	2000	7.89212×10^{10}	2.24	0.7	6.65537×10^{7}
25%	2005.5	1.175×10^{11}	2.885	2.65	7.25814×10^{7}
50%	2011	2.34×10^{11}	3.53	3.8	8.3908×10^{7}
75%	2016.5	3.235×10^{11}	3.66	5.05	1.27322×10^{8}
max	2022	4.04×10^{11}	4.05	8.3	1.50396×10^{8}

(Continuation of Table 1)

	Total Energy Consumption (KTOe)	Clean Water Access (%)	Population	Gini Coefficient	Human Development Index	SDG Index Score
count	23	23	23	23	23	23
mean	46812	45.19	9.65044×10^{7}	45.1304	0.677348	61.9626
std	9753.12	2.05088	1.15977×10^{7}	2.33413	0.0235364	2.7078
min	36959	40.57	7.79582×10^{7}	40.7	0.635	58.43
25%	38502	44.6	8.70815×10^{7}	43.45	0.659	59.77
50%	42083	45.45	9.63379×10^{7}	46.3667	0.677	61.43
75%	56736	46.64	1.05807×10^{8}	46.85	0.696	63.495
max	64512	47.9	1.15559×10^{8}	47.7	0.714	67.06

This statistical overview provides valuable insights into the distribution, variability, and central tendencies of socioeconomic and environmental metrics over a 23-year period. The data encompass critical variables such as GDP, unemployment rate, inflation rate, CO_2 emissions, energy consumption, clean water access, population, income inequality (Gini coefficient), Human Development Index (HDI), and the Sustainable Development Goals (SDG) Index. These metrics collectively form the empirical foundation for the machine learning-optimized evaluation conducted in this research, enabling a data-driven approach to understanding sustainability dynamics in the Philippines.

Table 1 presents the descriptive statistics of the dataset (n=23 annual observations, 2000-2022), summarizing the count, mean, standard deviation, minimum, maximum, and quartile values for each variable studied.

- ***GDP.*** The Gross Domestic Product (GDP) of the Philippines shows considerable variation over the period covered. With a mean of 2.30432×10^{11} USD and a standard deviation of 1.13313×10^{11} USD, the economy demonstrates significant fluctuations. The minimum GDP recorded is approximately 7.89212×10^{10} USD, while the maximum reaches 4.04×10^{11} USD. This wide range suggests periods of both strong economic growth and potential contractions in the Philippine economy during the timeframe studied.

- ***Unemployment Rate.*** The unemployment rate in the Philippines averages 3.30565%, with a standard deviation of 0.554033%. The lowest recorded rate is 2.24%, while the highest is 4.05%. This relatively narrow range and moderate standard deviation indicate some stability in the labor market, though there are still notable fluctuations. The median unemployment rate of 3.53% suggests that periods of higher unemployment are balanced by times of lower joblessness.

- ***Inflation Rate.*** Inflation in the Philippines shows considerable variability, with a mean of 3.86522% and a standard deviation of 1.75593%. The minimum inflation rate is 0.7%, while the maximum reaches 8.3%. This wide range indicates periods of both price stability and more significant inflationary pressures. The high standard deviation suggests that inflation has been a dynamic factor in the Philippine economy during the period studied.

- ***CO2 Emissions.*** Carbon dioxide emissions in the Philippines average 9.80359×10^7 tons, with a substantial standard deviation of 2.97617×10^7 tons. The minimum emissions recorded are 6.65537×10^7 tons, while the maximum reaches 1.50396×10^8 tons. This wide range and high standard deviation indicate significant changes in the country's carbon footprint over time, possibly reflecting industrialization, changes in energy sources, or environmental policies.

- ***Total Energy Consumption.*** Energy consumption in the Philippines averages 46812 KTOE (Kilotons of Oil Equivalent), with a standard deviation of 9753.12 KTOE. The minimum consumption is 36959 KTOE, while the maximum reaches 64512 KTOE. This range suggests growing energy demands over time, likely correlating with economic development and population growth.

- ***Clean Water Access.*** Access to clean water in the Philippines shows improvement over time, with a mean of 45.19% and a relatively small standard deviation of 2.05088%. The minimum access rate is 40.57%, while the maximum reaches 47.9%. This upward trend indicates progress in water infrastructure and public health initiatives, though there's still room for improvement.

- ***Population.*** The Philippine population data shows significant growth, with a mean of 9.65044×10^7 and a standard deviation of 1.15977×10^7. The population ranges from a minimum of 7.79582×10^7 to a maximum of 1.15559×10^8. This substantial increase over time highlights the demographic challenges and opportunities facing the country.

- ***GINI Coefficient.*** The GINI coefficient, measuring income inequality, averages 45.1384 with a standard deviation of 2.33413. The minimum value is 40.7, while the maximum is 47.7. This range suggests fluctuations in income distribution over time, with periods of both increasing and decreasing inequality.

- ***Human Development Index.*** The Human Development Index (HDI) for the Philippines averages 0.677348, with a standard deviation of 0.0235364. The minimum HDI is 0.635, while the maximum reaches 0.714. This upward trend indicates overall improvements in education, health, and standard of living, though the pace of progress varies.

- ***Sustainable Development Goals Index.*** The SDG Index for the Philippines averages 61.9626, with a standard deviation of 2.7078. The minimum index value is 58.43, while the maximum is 67.06. This range suggests progress towards sustainable development objectives, though there's variability in the rate of advancement across different goals and time periods.

These descriptive statistics provide a foundational understanding of the economic, environmental, and social landscape of the Philippines over the studied period. They serve as vital inputs for the subsequent phases of analysis and modelling in this research, supporting the development of machine learning models that can enhance sustainable development outcomes.

Interlinkages Between Socioeconomic and Environmental Factors and the SDG Index: A Facet Plot Analysis with Outlier Removal. To deepen the understanding of how various indicators relate to sustainable development, Figure 3.10 presents a facet plot with fitted trend lines that visualizes the relationships between multiple socioeconomic and environmental variables and the Sustainable Development Goals (SDG) Index. This visualization incorporates the Z-score method to remove outliers, thereby improving the clarity and reliability of the correlations observed.

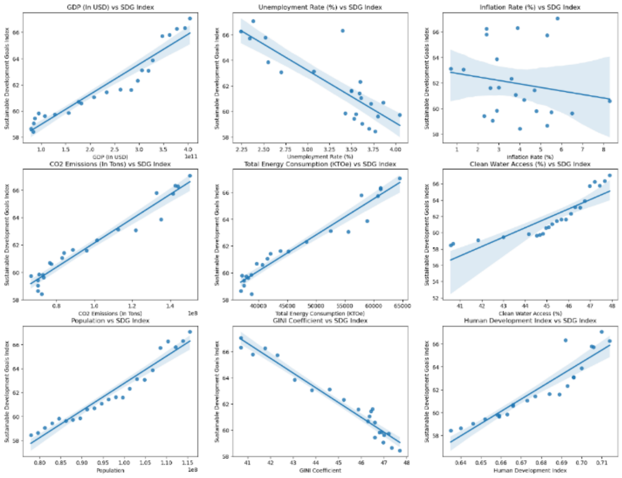


Figure 12

Facet Plot with Fitted Line and Outlier Removal Showing the Relationship between various factors and the SDG (Sustainable Development Goals) Index.

Figure 12 illustrates the distribution and correlation patterns between each factor and the SDG Index after outlier removal. This visualization highlights both positive and negative associations, providing insights into the drivers and inhibitors of sustainability performance in the Philippines.

- ***Positive Relationships.*** Several variables demonstrate strong positive correlations with the SDG Index. Notably, Gross Domestic Product (GDP) and the SDG Index show a pronounced positive relationship, underscoring the critical role of economic growth in advancing sustainability goals. Similarly, population size exhibits a positive correlation with SDG performance, which may reflect the increased resources and institutional capabilities of more populous countries. CO_2 emissions and total energy consumption also show positive associations with SDG scores, likely reflecting the current coupling of these factors with economic development. Access to clean water displays a clear positive relationship with SDG outcomes, indicating the fundamental importance of basic infrastructure in sustainable development. The Human Development Index (HDI) stands out with a particularly strong positive correlation, highlighting the intrinsic link between human development and the achievement of SDGs.

- ***Negative Relationships.*** In contrast, the facet plot identifies the unemployment rate and the Gini coefficient (a measure of income inequality) as key factors negatively correlated with the SDG Index. Higher unemployment rates correspond to lower SDG scores, indicating the detrimental impact of labor market challenges on sustainability progress. Likewise, increased income inequality, as represented by the Gini coefficient, correlates with poorer SDG performance, emphasizing the need for equitable development strategies to enhance overall sustainability.

- ***Outliers and Residuals.*** Most subplots exhibit data points tightly clustered around their trend lines, with only a few significant outliers, lending confidence to the observed correlations. However, the inflation rate subplot reveals a more scattered pattern, suggesting a weaker or potentially non-linear relationship with the SDG Index that warrants further investigation. The variability in residual patterns across factors indicates that while these variables capture much of the trend, additional unmodeled influences may also affect sustainable development outcomes.

The comprehensive facet plot analysis reveals the complex interplay of various socioeconomic and environmental factors with SDG performance. The strong positive correlations observed with economic indicators like GDP and the Human Development Index underscore their crucial role in advancing sustainable development goals. However, the negative associations with unemployment and income inequality highlight the need for balanced and equitable development strategies. The counterintuitive positive relationship between CO2 emissions and SDG scores emphasizes the ongoing challenge of decoupling economic growth from environmental degradation. These findings collectively demonstrate the multifaceted nature of sustainable development and the intricate web of global factors influencing progress towards the SDGs. Further research into these relationships could provide valuable insights for policymakers and development practitioners in formulating effective strategies to advance the global sustainable development agenda.

Correlation Analysis of Sustainability Indicators in the Philippines. Table 2 presents the correlation matrix of socioeconomic and environmental variables included in the dataset. This analysis provides foundational insights into how different factors interrelate and collectively influence sustainable development. By examining the strength and direction of these relationships, we can identify which variables most significantly support or hinder sustainability progress, especially within the Philippine context. The values in the matrix reveal both intuitive and counterintuitive patterns, offering a data-driven lens through which to understand developmental dynamics and policy priorities.

Table 2

Correlation Matrix of the Variables in the Dataset

	Year	GDP (In USD)	Unemployment Rate (%)	Inflation Rate (%)	CO2 Emissions (In Tons)	Total Energy Consumption (KTOe)
Year	1	0.99088	-0.772487	-0.237019	0.934225	0.938567
GDP (In USD)	0.99088	1	-0.773315	-0.271977	0.938533	0.943207
Unemployment Rate (%)	-0.772487	-0.773315	1	0.320314	-0.887155	-0.882443
Inflation Rate (%)	-0.237019	-0.271977	0.320314	1	-0.241054	-0.244254
CO2 Emissions (In Tons)	0.934225	0.938533	-0.887155	-0.241054	1	0.997461
Total Energy Consumption (KTOe)	0.938567	0.943207	-0.882443	-0.244254	0.997461	1
Clean Water Access (%)	0.94182	0.904781	-0.648459	-0.153485	0.799212	0.806229
Population	0.999857	0.990508	-0.777157	-0.234177	0.938369	0.94228
Gini Coefficient	-0.91756	-0.901329	0.844907	0.197974	-0.97231	-0.968187
Human Development Index	0.970699	0.967162	-0.814018	-0.288842	0.904534	0.906867
Sustainable Development Goals Index	0.96816	0.957648	-0.838482	-0.17854	0.973488	0.971524

(Continuation of Table 2)

	Clean Water Access (%)	Population	Gini Coefficient	Human Development Index	Sustainable Development Goals Index
Year	0.94182	0.999857	-0.91756	0.970699	0.96816
GDP (In USD)	0.904781	0.990508	-0.901329	0.967162	0.957648
Unemployment Rate (%)	-0.648459	-0.777157	0.844907	-0.814018	-0.838482
Inflation Rate (%)	-0.153485	-0.234177	0.197974	-0.288842	-0.17854
CO2 Emissions (In Tons)	0.799212	0.938369	-0.97231	0.904534	0.973488
Total Energy Consumption (KTOe)	0.806229	0.94228	-0.968187	0.906867	0.971524
Clean Water Access (%)	1	0.938646	-0.795207	0.935706	0.873957
Population	0.938646	1	-0.923641	0.968754	0.971707
Gini Coefficient	-0.795207	-0.923641	1	-0.856276	-0.975589
Human Development Index	0.935706	0.968754	-0.856276	1	0.930664
Sustainable Development Goals Index	0.873957	0.971707	-0.975589	0.930664	1

Table 2 provides valuable insights into the relationships between socioeconomic and environmental indicators and their association with sustainable development. The analysis reveals several strong and meaningful correlations among the variables, highlighting both positive and negative interactions. All reported correlations with $|r| > 0.413$ were statistically significant at $p < 0.05$ after Bonferroni correction for multiple comparisons (adjusted $\alpha = 0.0045$ for 11 variables).

- ***Year and SDG Index.*** With a correlation coefficient of 0.96816, there is a strong positive relationship between time and SDG performance. This suggests substantial progress in sustainable development over the years, indicating that global efforts to achieve SDGs have been largely effective.

- ***GDP and SDG Index.*** The correlation coefficient of 0.957648 demonstrates a strong positive relationship between economic output and sustainable development. This underscores the critical role of economic growth in advancing sustainability goals, potentially through increased resources for social and environmental initiatives.

- ***Unemployment Rate and SDG Index.*** The correlation matrix shows a strong negative correlation (-0.838482) between the unemployment rate and the SDG Index. This indicates that as unemployment rates decrease, SDG performance tends to improve.

- ***Inflation Rate and SDG Index.*** The correlation between the inflation rate and the SDG Index is relatively weak (-0.197592). This weak negative correlation suggests that while there's a slight tendency for higher inflation to be associated with lower SDG scores or vice-versa.

- ***CO_2 Emissions and SDG Index.*** Surprisingly, there is a strong positive correlation (0.973488) between CO_2 emissions and SDG scores. This counterintuitive relationship may reflect the current reality where economic development often coincides with increased emissions, highlighting the challenge of balancing growth with environmental sustainability.

- ***Total Energy Consumption and SDG Index.*** With a correlation of 0.971524, there's a strong positive relationship between energy consumption and SDG performance. This aligns with the CO_2 emissions correlation, further emphasizing the current link between development and resource use.

- ***Clean Water Access and SDG Index.*** A strong positive correlation (0.873597) exists between clean water access and SDG scores, highlighting the fundamental importance of basic infrastructure in achieving sustainable development goals.

- ***Population and SDG Index.*** A high correlation of 0.971707 indicates that countries with larger populations tend to have higher SDG scores. This could reflect the greater resources and capabilities of more populous nations in addressing sustainable development challenges.

- ***Gini Coefficient and SDG Index.*** A strong negative correlation (-0.975589) exists between income inequality and SDG performance. This suggests that countries with higher levels of income disparity tend to have lower SDG scores, emphasizing the importance of equitable development in achieving sustainability goals.

- ***Human Development Index and SDG Index.*** The very strong positive correlation (0.938664) between HDI and SDG scores underscores the close relationship between human development and sustainable development, indicating that progress in one area often corresponds with progress in the other.

These findings reveal the complex interplay of factors influencing sustainable development. While economic growth, larger populations, lower unemployment, increased energy use, high clean water access, and strong human development metrics are closely linked with higher SDG performance, issues like income inequality and inflation present persistent challenges. The notable positive correlations between SDG Index and CO_2 emissions and energy consumption highlight the ongoing struggle to decouple economic progress from environmental cost, underscoring the importance of integrated, balanced, and inclusive sustainability strategies.

Machine Learning Design Implications from the Correlation Heatmap. Figure 13 visualizes the same correlation matrix from Table 2 in the form of a heatmap, allowing for a more intuitive interpretation of inter-variable relationships. This visualization plays a crucial role in informing the design of machine learning (ML) models, particularly in selecting features that best capture the dynamics of sustainable development. For the Philippines, these insights can directly guide the construction of ML-optimized evaluation frameworks for technology and policy recommendations.

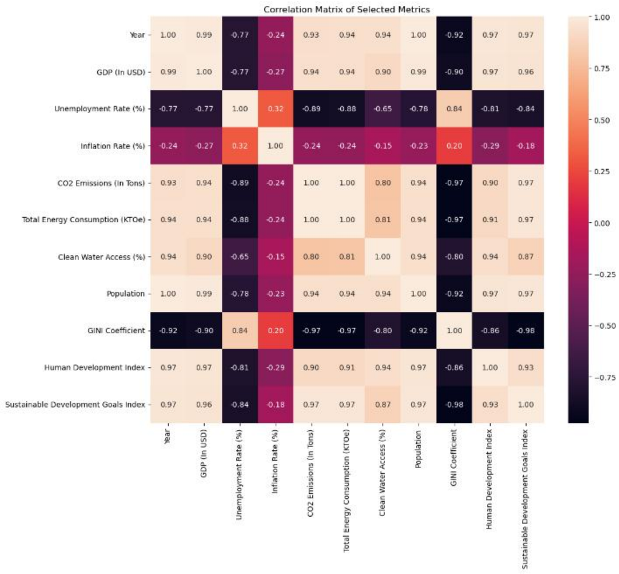


Figure 13

Correlation Matrix of Variables in the Dataset with Heatmap

In the context of the study, this correlation matrix provides crucial insights for developing effective predictive models and evaluation frameworks. The analysis reveals several key relationships that can inform the design of ML algorithms and sustainability metrics:

- ***Economic and Technological Factors.*** The strong positive correlation (0.957648) between GDP and the SDG Index and strong negative correlation (-0.838482) of Unemployment rate underscores the importance of economic growth in driving sustainable development. For the Philippines, this suggests that ML models should incorporate economic indicators as key features when evaluating technological solutions for sustainability. The models could be designed to predict how various technologies might impact both economic growth and SDG performance simultaneously.

- ***Demographic Considerations.*** The high correlation (0.971707) between population and SDG Index highlights the need for ML models to account for demographic factors. In the Philippines context, these models could be tailored to assess how different technologies might perform across various population sizes, ensuring scalable and adaptable solutions.

- ***Environmental and Economic Balance.*** The strong positive correlation (0.973488) between CO2 emissions and SDG scores, as well as the similar correlation (0.971524) for Total Energy Consumption and (0.873957) for Clean Water Access, emphasizes the challenge of balancing economic development with environmental sustainability. ML-optimized metrics for the Philippines should be designed to identify technologies that can break this pattern, promote economic growth while minimize environmental impact. This could involve multi-objective optimization algorithms that balance economic and environmental factors.

- ***Equity and Sustainability.*** The strong negative correlation (-0.975589) between the Gini coefficient and SDG Index underscores the importance of equitable development. ML models for technology evaluation in the Philippines should incorporate equity metrics, ensuring that recommended solutions contribute to both sustainability and reduced inequality. This could involve developing composite scores that weigh technological efficacy against measures of social impact and accessibility.

- ***Human Development.*** The very strong positive correlation (0.938664) between the Human Development Index and SDG Index suggests that improvements in education, health, and living standards are closely linked to sustainable development progress. ML models should consider how technologies can contribute to human development alongside environmental and economic factors.

- ***Temporal Trends.*** The strong positive correlation (0.96816) between year and SDG Index suggests a general improvement in sustainability over time. ML models could leverage this trend to forecast future sustainability scenarios for the Philippines, helping to identify which technologies are likely to have the most significant long-term impact.

These correlations provide a robust data-driven foundation for building predictive ML models tailored to the Philippine context. By integrating these variables into feature selection and model architecture, researchers can develop more precise, equitable, and impactful tools for evaluating technology and policy interventions, contributing to evidence-based progress toward national sustainability goals.

Model Performance Comparison. This section presents the results from applying various machine learning regression models to predict sustainable development outcomes based on socioeconomic and environmental indicators. The models evaluated include Decision Tree Regression, Gradient Boosting Regression, and Random Forest Regression. Their performance was assessed using two standard metrics: the R-squared (R^2) score and the Root Mean Squared Error (RMSE). These metrics provide insights into how well each model fits the data and predicts the target variable, allowing for an effective comparison of their predictive capabilities.

Table 3
R-Squared Scores of different Machine Learning Models

#	Model	R2 Score
1	Random Forest	0.915534
2	Gradient Boosting	0.906467
3	Decision Tree	0.597453

Table 4
Root Mean Square Error Scores of different Machine Learning Models

#	Model	RSME Score
1	Random Forest	0.566654
2	Gradient Boosting	0.596292
3	Decision Tree	1.237045

The dataset was utilized to train and assess several regression models, including Decision Tree Regression, Gradient Boosting Regression, and Random Forest Regression. The efficacy of each model was measured using two key metrics: the R-squared score and the RMSE (Root Mean Squared Error). The R-squared score indicates the proportion of variance in the dependent variable that can be accurately predicted from the independent variables. The RMSE score, on the other hand, provides a measure of the average magnitude of errors between the predicted and observed values, with a lower RMSE indicating a better fit of the model to the data.

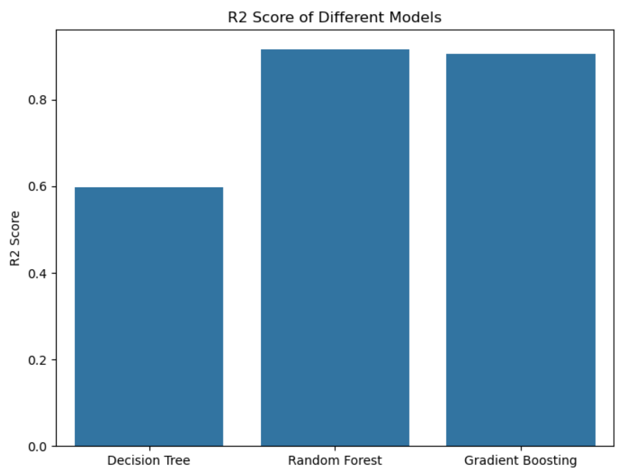


Figure 14
R-squared Scores of different Machine Learning Models

The Decision Tree model achieved the lowest R2 score of approximately 0.6. The Random Forest model shows a significant improvement, with an R2 score of about 0.93. The Gradient Boosting model performs slightly lower than Random Forest, with an R2 score just below 0.93.

These results indicate that both ensemble methods (Random Forest and Gradient Boosting) substantially outperform the simple Decision Tree in terms of explanatory power. The Random Forest model marginally edges out Gradient Boosting as the best-performing model based on R2 score.

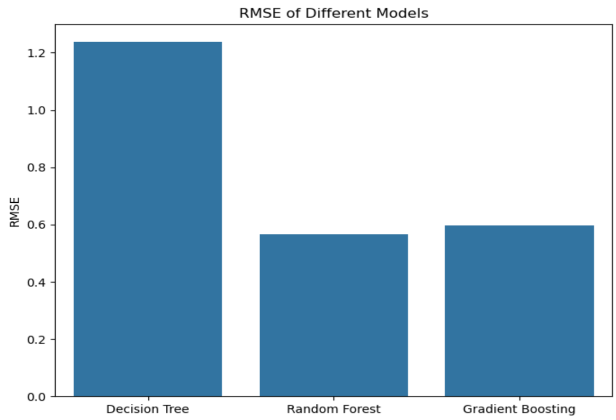


Figure 15
Root Mean Square Errors Scores of different Machine Learning Models

The Decision Tree model has the highest RMSE, at about 1.25. The Random Forest model shows a significant reduction in RMSE, dropping to approximately 0.57. The Gradient Boosting model achieves a slightly higher RMSE than Random Forest, at about 0.6.

The RMSE results corroborate the findings from the R2 score comparison. Both ensemble methods demonstrate much lower prediction error compared to the Decision Tree, with Random Forest showing a slight advantage over Gradient Boosting.

Based on both R2 score and RMSE metrics, the Random Forest model appears to be the most effective for this collected dataset and prediction task. It offers the highest explanatory power (R2 score) and the lowest prediction error (RMSE). The Gradient Boosting model performs very similarly, while the simple Decision Tree lags in both metrics.

Feature Importance Analysis. Table 5 and Figure 16 present the feature importance rankings of the employed Random Forest Regression model to analyze the relative importance of various socioeconomic and environmental factors in predicting sustainability outcomes. The model's feature importance rankings provide valuable insights into the key drivers of sustainability in the Philippine context.

Table 5
Feature Importances of Random Forest Regression Model

#	Feature	Importance
1	GINI Coefficient	0.188305
2	GDP (In USD)	0.176796
3	CO2 Emissions (In Tons)	0.155152
4	Unemployment Rate (%)	0.143923
5	Clean Water Access (%)	0.0962336
6	Total Energy Consumption (KTOe)	0.0817546
7	Population	0.0770006
8	Human Development Index	0.0732061
9	Inflation Rate (%)	0.00762868

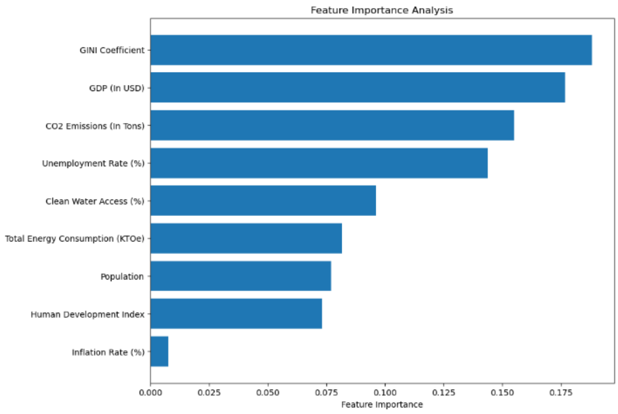


Figure 16
Bar Plot of Feature Importances of Random Forest Regression Model

The analysis reveals that income inequality, as measured by the Gini Coefficient, emerges as the most critical factor with an importance score of 0.188305. This finding underscores the significant impact of wealth distribution on sustainable development initiatives.

Closely following is the country's GDP (in USD), with an importance score of 0.176796, highlighting the crucial role of economic output in sustainability efforts. Environmental concerns are prominently represented by CO2 Emissions (in tons), ranking third with a score of 0.155152, emphasizing the need to balance economic growth with ecological considerations. The model also identifies unemployment rate (0.143923) and access to clean water (0.0962336) as significant factors, pointing to the importance of addressing basic socioeconomic needs.

Interestingly, while total energy consumption (0.0817546) and population (0.0770006) show moderate importance, the Human Development Index (0.0732061) and inflation rate (0.00762868) appear to have less direct influence on the model's sustainability predictions. These findings suggest that policies targeting income inequality, economic growth, and environmental protection could potentially yield the most significant improvements in overall sustainability performance for the Philippines.

It is crucial to note that these importance scores reflect correlations within the model and its available datasets and do not necessarily imply causation. Nevertheless, they provide a data-driven framework for prioritizing sustainability initiatives and technology evaluations in the Philippine context.

The analysis suggests that strategies targeting Income inequality reduction, sustainable economic growth, CO2 emissions reduction, and employment rate could potentially yield the most significant improvements in overall sustainability performance for the Philippines.

Feature Impact Analysis. This explores the influence of individual features on the Sustainability Index using Partial Dependence Plots derived from the Random Forest Regression model. These plots provide a clear visualization of how changes in specific socioeconomic and environmental variables affect sustainability outcomes in the Philippines, revealing complex and sometimes counterintuitive relationships that are crucial for informed policymaking.

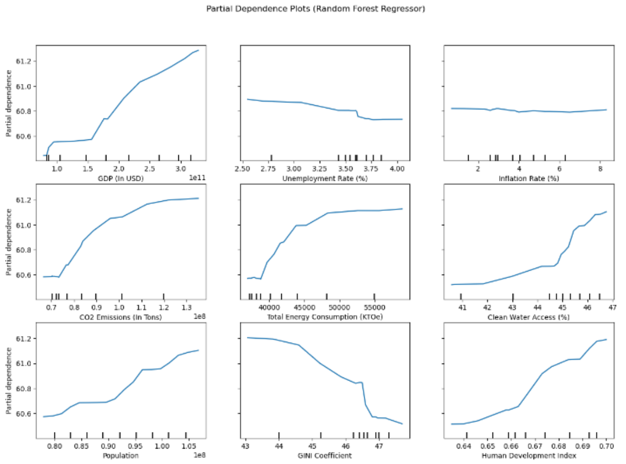


Figure 17
Partial Independence Plot of Random Forest Regression Model

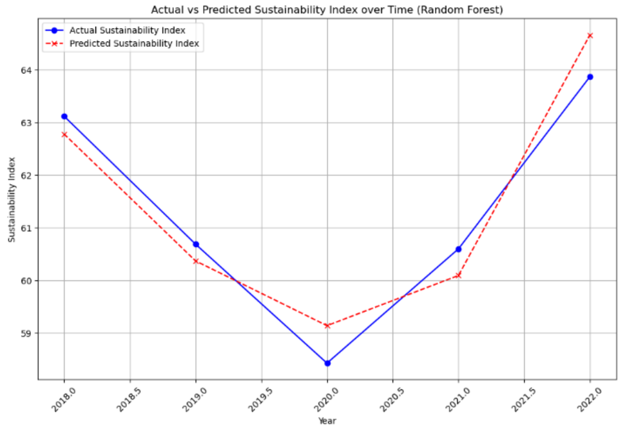
The Partial Dependence Plots derived from the Random Forest Regressor offer valuable insights into how individual features influence the Sustainability Index in the Philippines. GDP demonstrates a strong positive correlation with sustainability, albeit with diminishing returns at higher levels, underscoring the importance of economic growth. The Human Development Index and Clean Water Access show robust positive relationships, highlighting the critical role of human welfare and basic infrastructure.

Interestingly, CO2 Emissions and Total Energy Consumption also exhibit positive correlations, potentially reflecting overall development levels rather than environmental sustainability specifically. The GINI Coefficient's negative relationship emphasizes the detrimental impact of income inequality. Population size shows a positive relationship with sustainability, possibly due to increased resources and capabilities in more populous areas. The Unemployment Rate displays a slightly negative impact, indicating that higher unemployment rates marginally decrease sustainability.

Notably, the Inflation Rate appears to have minimal effect on sustainability predictions, with the plot showing a nearly flat line across different inflation levels. These nuanced relationships illuminate the multifaceted nature of sustainability, suggesting that policies aimed at enhancing sustainability in the Philippines must balance economic growth, social equity, environmental considerations, and demographic factors while managing unemployment and maintaining price stability.

Model Performance over Time. This examines the temporal predictive performance of the Random Forest Regression model by comparing actual versus predicted Sustainability Index values over the period from 2018 to 2022. This analysis provides insight into the model's strengths and limitations in capturing sustainability trends and short-term variations within the Philippine context.

Figure 18
Model Performance over Time of Random Forest Regression Model



The plot comparing Actual vs Predicted Sustainability Index over time, based on the Random Forest model, provides crucial insights into the model's predictive capabilities and limitations. The graph demonstrates that the model successfully captures the overall upward trend in sustainability in the Philippines from 2018 to 2022. However, it consistently underestimates the actual Sustainability Index values, indicating a systematic bias in the predictions. This underestimation suggests that there may be additional factors contributing to sustainability that are not fully captured by the current feature set.

Furthermore, the model struggles to accurately predict short-term fluctuations, particularly evident in the 2019-2020 period, which could be attributed to external shocks or rapid policy changes not reflected in the input features. Despite these limitations, the model's ability to track the general trend suggests its utility for long-term sustainability planning and policy formulation, while also highlighting areas for potential improvement in model accuracy and feature selection.

On the hold-out period, the Random Forest achieved $R^2 = 0.915534$ and RMSE = 0.566654. This exceeds the hypothesized threshold of $R^2 > 0.85$, supporting the study's expectation that ensemble methods can attain high predictive accuracy even under small-sample constraints. Given the limited test window, results are interpreted cautiously but provide promising evidence that Random Forest can serve as a reliable tool for sustainability evaluation in the Philippine context.

Comparative Performance. To contextualize the model's accuracy, the Random Forest results for the Philippines ($R^2 = 0.915534$, RMSE = 0.566654) were compared with similar studies in other developing-country contexts. Mohata et al. (2024) applied Random Forest to state-level SDG indicators in India, achieving $R^2 = 0.727$, while Rainford et al. (2021) reported Random Forest explaining 50.28% of the variance in soil organic carbon estimates in Colombia. Both studies demonstrated that ensemble methods perform reliably under constrained data environments but at notably lower levels of accuracy than observed here. The comparatively higher performance of this study can be attributed to two methodological choices: (1) the ensemble approach with careful hyperparameter tuning, which enhanced stability under limited observations; and (2) the inclusion of carefully selected socioeconomic and environmental predictors strongly tied to SDG outcomes, such as the Gini coefficient, GDP, and CO_2 emissions. Together, these design decisions yielded a predictive framework capable of surpassing results from other small-sample developing country applications, reinforcing the robustness of Random Forest in national-level sustainability modelling.

DISCUSSIONS

This study successfully demonstrated that machine learning can predict sustainability outcomes with 91. This study successfully demonstrated that machine learning can predict sustainability outcomes with 91.5% accuracy ($R^2 = 0.915534$) using limited annual data, providing a viable framework for data-constrained developing nations. By integrating economic, social, and environmental indicators into ensemble algorithms, the research contributes to the development of a robust, data-driven evaluation framework that captures nonlinear interactions and complex trade-offs. The findings confirm the study's central hypothesis that ensemble learning models, particularly Random Forest, can exceed predictive thresholds ($R^2 > 0.85$) even under small-sample constraints, offering both methodological and policy contributions to sustainability assessment.

The study achieved its objectives systematically. Feature importance analysis identified income inequality (Gini coefficient, 0.188305), GDP (0.176796), and CO_2 emissions (0.155152) as the strongest predictors of sustainability, thereby addressing the first objective of key metrics identification. The second objective, integrating diverse socioeconomic and environmental datasets, was met by synthesizing 11 indicators from five international databases, demonstrating the feasibility of multi-source data integration. For the third objective, model selection, Random Forest emerged as the best-performing algorithm, achieving the highest predictive accuracy ($R^2 = 0.915534$, RMSE = 0.566654). Building on this, the fourth objective was fulfilled by developing a comprehensive evaluation framework that balances predictive performance with interpretability through feature importance and partial dependence analyses. Finally, the fifth objective was realized by generating actionable policy insights, particularly in targeting inequality, promoting renewable energy, and improving access to clean water, illustrating the practical utility of machine learning models for decision-making.

The correlation analysis revealed strong and statistically significant associations between sustainability performance and several key indicators, consistent with prior findings

(Barcelon et al., 2023; Vinuesa et al., 2020; Mohata et al., 2024). GDP ($r = 0.957648$), HDI ($r = 0.930664$), clean water access ($r = 0.873957$), and population ($r = 0.971707$) were positively associated with the SDG Index, underscoring the role of inclusive growth, human capital, and infrastructure in shaping sustainability. Conversely, income inequality (Gini coefficient, $r = -0.975589$) was negatively correlated, confirming equity as a central driver of sustainable development. All correlations with $|r| > 0.413$ were statistically significant at $p < 0.05$ after Bonferroni correction, reinforcing the reliability of these findings.

A paradoxical result was the strong positive correlation between CO_2 emissions and SDG performance ($r = 0.973488$). This reflects the Environmental Kuznets Curve (EKC), in which emissions intensify during early growth before tapering at higher development stages. The Philippines remains on the EKC's upward slope, where growth is still carbon intensive. Fong, Salvo, and Taylor (2020) provide empirical evidence of the Environmental Kuznets Curve in Southeast Asia, demonstrating that emissions intensify during economic growth before easing with technological innovation and efficiency gains in more advanced stages, an observation also echoed in global studies (Ritchie et al., 2020; OECD, 2022). This underscores the urgency of scaling renewable energy, incentivizing green industries, and adopting cleaner technologies to avoid locking in carbon-heavy growth trajectories.

The predictive modeling results reinforce these insights. Random Forest outperformed other models, achieving both high accuracy and stability, consistent with prior evidence of Random Forest outperforming baseline models in sustainability contexts (Han et al., 2021; Mohata et al., 2024). Feature importance confirmed inequality, GDP, and CO_2 emissions as key sustainability drivers. This suggests Random Forest can serve as a practical policy simulation tool, enabling governments to test reforms, such as progressive taxation or renewable energy incentives, and estimate their likely impact on SDG outcomes.

This research also extends global AI–sustainability literature. Vinuesa et al. (2020) stressed AI's potential to support over 130 SDG targets but noted assumptions of data abundance. By applying ensemble learning to just 23 observations, this study shows that high predictive performance is possible even under severe data constraints, consistent with other small-sample studies. Han et al. (2021) validated Random Forest in small samples, Rainford et al. (2021) demonstrated its utility for soil carbon estimation in Colombia, and Mohata et al. (2024) achieved competitive predictive performance for state-level SDG forecasting in India.

Together, these studies support the methodological approach used in this research. Sala et al. (2015) argued for systemic, multidimensional frameworks; this study operationalizes that call with predictive machine learning. Philippine-specific studies (Cuevas, 2017; Lasco & Pulhin, 2006; Barcelon et al., 2023) highlighted institutional and socioeconomic barriers, and this research provides a quantitative framework to complement those insights.

The study also highlighted limitations. First, the small dataset constrained the reliability of validation. A chronological hold-out (2000–2017 train, 2018–2022 test) was applied, and bootstrap resampling further confirmed that predictive accuracy was associated with broad uncertainty bands (Han et al., 2021; Breiman, 2001). This reflects the sensitivity of ensemble methods under small sample conditions and underscores the need for cautious interpretation of point estimates.

Second, the models underestimated short-term shocks, particularly during 2019–2020, when COVID-19 disrupted employment, inflation, and energy consumption beyond what annual indicators could capture.

Third, Philippine statistical limitations persist, including gaps in sub-national data, inconsistent social metrics, and difficulty accounting for informal economic activity.

Despite these challenges, the policy implications are compelling. Results show inequality reduction has a disproportionately strong effect on sustainability compared to marginal GDP gains, suggesting redistributive measures may deliver higher SDG returns than growth alone. Meanwhile, the positive CO_2–SDG relationship highlights the urgency of accelerating green transitions to prevent emissions from rising in parallel with development (Fong, Salvo, & Taylor, 2020; Ritchie et al., 2020; OECD, 2022). Investments in renewable energy, clean water, and efficiency technologies therefore emerge as dual-benefit strategies with both social and environmental payoffs.

Benchmarking further highlights this study's contribution. Mohata et al. (2024) reported Random Forest R^2 = 0.727 for state-level SDG predictions in India, while Rainford et al. (2021) achieved R^2 = 0.503 for soil carbon estimation in Colombia. Both fall below the R^2 = 0.915534 achieved here, suggesting that careful feature engineering and conservative tuning can yield superior predictive performance even under small-sample constraints. This comparative strength reinforces the Philippines study as a methodological reference for other developing nations.

This research also contributes to the growing discussion of ethical data use in sustainability research. All datasets were sourced from publicly accessible repositories (World Bank, UNDP, OECD, MacroTrends, Enerdata), ensuring transparency and reproducibility. No personally identifiable information was used, and all analyses comply with open-data ethics. This transparency strengthens the credibility of the results and provides a replicable foundation for future research. Moreover, the framework demonstrates how openly available aggregate data can be harnessed effectively in resource-limited contexts to generate actionable sustainability insights.

Finally, this study proposes several forward-looking recommendations to translate insights into practice. Policies should prioritize inequality reduction as the most critical driver of sustainability outcomes, with interventions ranging from progressive tax regimes to targeted labor market reforms. Government investment in sustainable technologies should be scaled up, especially in renewable energy and green industry, to decouple growth from emissions and shift the Philippines onto the downward slope of the EKC. Clean water access and employment programs must also be prioritized, as these indicators strongly correlate with SDG performance. Policymakers should leverage machine and deep learning models to simulate alternative development scenarios, updating them as new data becomes available. Expanding sustainability data infrastructures, through improved monitoring, sub-national reporting, and real-time environmental indicators, will further strengthen predictive accuracy. Collaboration across government agencies, academia, and the private sector is essential to implement machine learning–optimized frameworks, ensuring sustainability policies are both evidence-based and contextually relevant.

In conclusion, this research provides both methodological and practical contributions to sustainability science. It validates ensemble learning as a reliable tool for small-data contexts, demonstrates how machine learning can inform integrated technology evaluation frameworks, and offers actionable insights for Philippine policymakers. Despite inherent data limitations, the study's integrated approach, combining rigorous correlation analysis, predictive modeling, and policy relevance, advances the discourse on sustainable development. By addressing inequality, accelerating green technology adoption, and strengthening data infrastructures, the Philippines can progress toward a more resilient, equitable development path. These lessons extend beyond the Philippines, offering guidance for other developing nations seeking to harness machine learning for sustainability under constrained data environments.

REFERENCES

Athey, S. (2018). The impact of machine learning on economics. In A. Agrawal, J. Gans, &

A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda* (pp. 507–547). University of Chicago Press. https://doi.org/10.7208/chicago/9780226613475.003.0021

Barcelon, J., Cabrera, T., Suin, K. (2023). The Contributions of Human Capital to the Philippine GDP Per Capita. *Millennium Journal of Humanities and Social Sciences*, 14–42. https://doi.org/10.47340/mjhss.v4i2.2.2023

Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences, 191*, 192–213. https://doi.org/10.1016/j.ins.2011.12.028

Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, 25, 197–227. https://doi.org/10.1007/s11749-016-0481-7

Breiman, L. Random Forests. *Machine Learning* 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324

Cerqueira, V., Torgo, L., & Mozetič, I. (2020). Evaluating time series forecasting models: An empirical study on performance estimation methods. *Machine Learning, 109*(11), 1997–2028. https://doi.org/10.1007/s10994-020-05910-7

Cuevas, S. C. (2017). Institutional dimensions of climate change adaptation: insights from the Philippines. *Climate Policy*, 18(4), 499–511. https://doi.org/10.1080/14693062.2017.1314245

Department of Energy. (2024). DOE expects more than 4,000 MW to come online in 2024 to boost the grid. *Retrieved from https://legacy.doe.gov.ph/press-releases/doe-expects-more-4000-mw-come-online-2024-boost-grid*

Domingo, K. (2024). 40M Filipinos do not have access to clean and potable water: DENR. *ABS-CBN News*. *Retrieved from https://news.abs-cbn.com/news/2024/5/7/40m-filipinos-do-not-have-access-to-clean-and-potable-water-denr-1320*

Enerdata. (n.d.). Philippines energy information. *Retrieved from https://www.enerdata.net/estore/energy-market/philippines/*

European Climate Adaptation Platform. (2024). Consequences of climate change. *Retrieved from https://climate.ec.europa.eu/climate-change/consequences-climate-change_en*

Fong, L. S., Salvo, A., & Taylor, D. (2020). Evidence of the environmental Kuznets curve for atmospheric pollutant emissions in Southeast Asia and implications for sustainable development: A spatial econometric approach. *Sustainable Development, 28*(5), 1441–1456. https://doi.org/10.1002/sd.2097

Han, S., Williamson, B.D. & Fong, Y. (2021). Improving random forest predictions in small datasets from two-phase sampling designs. BMC Med Inform Decis Mak 21, 322 (2021). https://doi.org/10.1186/s12911-021-01688-3

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts. Retrieved from https://otexts.com/fpp2/

Lasco, R., & Pulhin, J. (2006). Environmental impacts of community-based forest management in the Philippines. *International Journal of Environment and Sustainable Development*, 5(1), 46–56. https://doi.org/10.1504/IJESD.2006.008682

MacroTrends. (n.d.). Philippines clean water access 1960–2024. *Retrieved from https://www.macrotrends.net/global-metrics/countries/PHL/philippines/clean-water-access-statistics*

MacroTrends. (n.d.). Philippines unemployment rate 1960–2024. *Retrieved from https://www.macrotrends.net/global-metrics/countries/PHL/philippines/unemployment-rate*

Mohata, R., Chandrakar, A., Mishra, S., Sharma, V., Uhumavbi, I., & Ekatitie, E. L. (2024). Sustainability development metric validation of Indian territory using regression approach. University of Bolton Institutional Repository. *Retrieved from https://ub-ir.bolton.ac.uk/view/pdfCoverPage?download=true&filePid=13411783000 8841&instCode=44UOBO_INST*

OECD. (2022). *Measuring the environmental impacts of artificial intelligence compute and applications: The AI footprint.* OECD Digital Economy Papers, No. 339. https://doi.org/10.1787/7babf571-en

Philippine Statistics Authority. (2024). Consumer price index and inflation rate. *Retrieved from https://psa.gov.ph/price-indices/cpi-ir*

Philippine Statistics Authority. (2024). Labor force survey. *Retrieved from https://www.psa.gov.ph/statistics/labor-force-survey/node/1684063887*

Philippine Statistics Authority. (2025). GDP expands by 5.5 percent in the second quarter of 2025. *Retrieved from https://psa.gov.ph/statistics/national-accounts/sector3/Gross%20Domestic%20Product*

Rainford, S.-k., Martín-López, J. M., & Da Silva, M. (2021). Approximating Soil Organic Carbon Stock in the Eastern Plains of Colombia. Frontiers in Environmental Science, 9, 685819. https://doi.org/10.3389/fenvs.2021.685819

Ritchie, H., Roser, M., & Rosado, P. (2020a, May 11). Philippines: CO2 country profile. *Our World in Data*. *Retrieved from https://ourworldindata.org/co2/country/Philippines*

Sala, S., Ciuffo, B., & Nijkamp, P. (2015). A systemic framework for sustainability assessment. *Ecological Economics*, 119, 314–325. https://doi.org/10.1016/j.ecolecon.2015.09.015

Sustainable Development Report. (2024). Philippines, East and South Asia. *Retrieved from https://dashboards.sdgindex.org/profiles/philippines*

Sustainable Development Solutions Network. (2024). *Sustainable Development Report 2024*. *Retrieved from https://www.unsdsn.org/resources/the-sustainable-development-report-2024/*

UNDP. (2024). Human development summary capturing achievements in the HDI and complementary metrics. *Retrieved from https://hdr.undp.org/data-center/specific-country-data#/countries/PHL*

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S., Tegmark, M., & Nerini, F. F. (2020). The role of artificial intelligence in achieving the sustainable development goals. *Nature Communications*, 11(1), 233. https://doi.org/10.1038/s41467-019-14108-y

World Bank Data. (2024). GDP (current US$) – Philippines. *Retrieved from https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?end=2022&locations=PH&start=1960&view=chart*

World Bank Open Data. (n.d.). Inflation, consumer prices (annual %) – Philippines. *Retrieved from https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG?end=2022&locations=PH*

World Inequality Database. (n.d.). *Retrieved from https://wid.world/country/philippines/*

Worldometers. (2024). Philippines population. *Retrieved from https://www.worldometers.info/world-population/philippines-population/*

Zhang, J., Sun, W., Pradhan, P., Gao, S., Su, C., Skene, K. R., & Fu, B. (2025). Nonlinear and weak interactions among sustainable development goals (SDGs) drive China's SDGs growth rate below expectations. *Environmental Impact Assessment Review, 115,* 107990. https://doi.org/10.1016/j.eiar.2025.107990