

ISSN: 1672 - 6553

**JOURNAL OF DYNAMICS
AND CONTROL**

VOLUME 9 ISSUE 8: 89 - 116

**ROLE OF ENERGY USE IN THE
PREDICTION OF CO2 EMISSIONS
AND ECONOMIC GROWTH: GLOBAL
EVIDENCE WITH MACHINE
LEARNING**

Tam Phan Huy

University of Economics and Law & Vietnam
National University, Ho Chi Minh City, Vietnam

ROLE OF ENERGY USE IN THE PREDICTION OF CO₂ EMISSIONS AND ECONOMIC GROWTH: GLOBAL EVIDENCE WITH MACHINE LEARNING

Tam Phan Huy

University of Economics and Law & Vietnam National University, Ho Chi Minh City, Vietnam

**The research is funded by the University of Economics and Law, Vietnam National University, Ho Chi Minh City, Vietnam*

Abstract: *This study evaluates the role of energy consumption in predicting CO₂ emissions and economic growth using advanced machine learning models. By analyzing data from 220 countries spanning 1990 to 2022, the research integrates Decision Tree, Random Forest, XGBoost, LightGBM, and CatBoost algorithms to uncover key predictors and their impacts. The findings indicate that electricity and fossil fuel consumption are critical predictors of CO₂ emissions, while renewable energy use significantly influences economic growth. Gradient boosting models, particularly LightGBM and CatBoost, demonstrated superior predictive performance, capturing complex relationships between energy use and economic outcomes. The study concludes that balancing fossil fuel use with renewable energy adoption is essential for sustainable economic development and environmental protection. Recommendations for investors include prioritizing renewable energy projects and supporting energy efficiency initiatives. Managers are encouraged to adopt sustainable practices and invest in innovation, while government agencies should promote renewable energy policies and enforce energy efficiency standards. These insights aim to inform policy and strategic decision-making to achieve sustainable development goals.*

Keywords: *Energy Consumption, CO₂ Emissions, Economic Growth, Machine Learning, Sustainable Development*

1. Introduction

Understanding the relationship between energy use, CO₂ emissions, and economic growth is crucial for formulating policies that balance environmental sustainability with economic development. This intricate interplay is not only pivotal for mitigating climate change but also for ensuring sustainable economic progress. The consumption of energy, particularly from fossil fuels, is a significant driver of CO₂ emissions, which in turn contribute to global warming and climate change. Meanwhile, economic growth often necessitates increased energy consumption, creating a complex dynamic that warrants thorough examination.

The primary contribution of this paper is the application of advanced machine learning (ML) techniques to predict CO₂ emissions and economic growth based on energy consumption patterns. Traditional econometric methods, while effective in analyzing the impact of explanatory variables, often fall short in handling the non-linear relationships and large datasets

inherent in environmental and economic data. Machine learning models, such as Decision Trees, Random Forest, XGBoost, LightGBM, and CatBoost, offer superior predictive capabilities by capturing complex interactions between variables that conventional methods may overlook.

The rationale for using machine learning in this study lies in its ability to provide more accurate and robust predictions. Unlike conventional methods that primarily focus on understanding the role of individual variables, ML models are designed to predict outcomes by considering the intricate patterns within the data. This approach is particularly valuable in our study as it enables us to forecast CO₂ emissions and economic growth more precisely, offering actionable insights for policymakers and stakeholders. Furthermore, machine learning models can handle multicollinearity and overfitting more effectively than traditional econometric techniques. These strengths make ML a powerful tool for predictive analysis in the context of sustainable development, where the relationships between variables are often non-linear and complex.

Theoretical frameworks such as the Environmental Kuznets Curve (EKC) hypothesis suggest an inverted U-shaped relationship between environmental degradation and economic growth, implying that environmental quality deteriorates up to a certain point as an economy grows, and improves thereafter. However, the empirical validity of the EKC hypothesis, especially concerning CO₂ emissions, remains a subject of debate among scholars. Grossman and Krueger (1991) and Richmond and Kaufmann (2006) provide foundational perspectives on the EKC, yet their findings highlight the need for more nuanced models that can account for diverse economic contexts and energy consumption patterns. This ongoing debate underscores the necessity for advanced predictive models that can offer more precise insights into the energy-emissions-growth nexus.

In recent years, machine learning has emerged as a powerful tool for predictive modeling, offering capabilities that surpass traditional statistical methods. Machine learning algorithms such as Decision Trees, Random Forest, XGBoost, LightGBM, and CatBoost have demonstrated remarkable success in handling large datasets and capturing complex, non-linear relationships among variables (Sagi & Rokach, 2018). These advanced techniques enable researchers to uncover intricate patterns and make accurate predictions from vast amounts of data, providing insights that were previously difficult to obtain through conventional approaches. Particularly noteworthy are ensemble methods like Random Forest and boosting techniques such as XGBoost, which offer robust predictive performance by aggregating the

outcomes of multiple models and effectively mitigating the risk of overfitting (Chen & Guestrin, 2016).

Despite the progress made, previous studies often fall short in integrating comprehensive datasets that span multiple countries and years, limiting the generalizability of their findings. Additionally, many existing models lack the capability to accurately capture the dynamic and non-linear relationships inherent in the energy-emissions-growth nexus. This research aims to fill these gaps by leveraging advanced machine learning techniques and extensive datasets from the World Bank, offering a more holistic and accurate analysis of the interplay between energy use, CO₂ emissions, and economic growth. The significance of this research lies in its potential to inform policy and strategic decision-making, promoting a balance between economic development and environmental sustainability. By providing robust, data-driven insights, this study seeks to contribute to the global discourse on sustainable development and highlight the importance of innovative analytical approaches in addressing complex global challenges.

The objective of this research is to evaluate the role of energy consumption in predicting CO₂ emissions and economic growth using advanced machine learning models. The study is structured as follows: Section 2 reviews the relevant literature, discussing key theories and empirical findings. Section 3 details the methodology employed, including data sources and the machine learning models used. Section 4 presents the findings from the empirical analysis, and Section 5 concludes the paper with recommendations for policymakers and stakeholders, emphasizing the study's implications for achieving sustainable development goals.

2. Literature review

2.1. Background theories

The intricate relationship between energy use, greenhouse gas emissions, and economic growth has garnered substantial attention from researchers and policymakers alike. While the Environmental Kuznets Curve (EKC) hypothesis (Grossman & Krueger, 1991) has been a prominent theory, its generality has been called into question by recent studies, particularly in the context of greenhouse gas emissions (Richmond & Kaufmann, 2006). Consequently, alternative theories and models have emerged to provide a more comprehensive understanding of this complex nexus.

One such theoretical framework is the Green Growth Model, which integrates economic growth, energy use, and environmental pollution into a unified structure. This model posits that economic growth can be achieved through the utilization of clean energy sources and the reduction of greenhouse gas emissions (Smulders, Toman, & Withagen, 2014). Complementing

this perspective, the Energy Innovation Theory emphasizes the pivotal role of innovation and technological advancements in reducing energy intensity and mitigating greenhouse gas emissions. It suggests that policies promoting investment in research and development within the clean energy sector will positively impact both economic growth and emission reduction (Aghion et al., 2016). The Energy-Economic System Model offers a cohesive mathematical framework that combines economic, energy, and environmental factors. This model enables the analysis of the impacts of energy and climate policies on economic growth, energy use, and greenhouse gas emissions (Dai et al., 2016). Conversely, the Energy Restructuring Theory focuses on the transition from fossil fuels to cleaner and renewable energy sources. It asserts that the restructuring of energy systems toward more sustainable alternatives will contribute to emission reduction while simultaneously fostering sustainable economic growth.

Furthermore, the Energy-Environment-Economy Development Model examines the triadic relationship between energy use, environmental quality, and economic growth. This model integrates factors such as energy technology, environmental policies, and investment in research and development to assess the impacts of various scenarios on sustainable development (Rogelj et al., 2018). By considering these interconnected elements, the model provides a holistic approach to understanding the complex dynamics at play. These recent theoretical foundations and research models reflect the evolving landscape in the study of the energy-emissions-growth nexus, acknowledging the multifaceted nature of the issue and the need for innovative solutions. Notably, they emphasize the significance of technological advancements, appropriate policy frameworks, and a transition towards cleaner energy sources as critical components in achieving sustainable economic development while mitigating the adverse impacts of greenhouse gas emissions.

Moreover, these theoretical underpinnings highlight the importance of integrating economic, energy, and environmental factors into unified frameworks, recognizing the intricate interplay between these elements. By doing so, researchers and policymakers can better comprehend the complex dynamics at play and develop more effective strategies to address the challenges posed by climate change and energy-related issues. In essence, the theoretical foundations discussed above provide a robust and nuanced understanding of the relationship between energy use, CO₂ emissions, and economic growth. They offer a diverse range of perspectives and analytical tools, enabling researchers and policymakers to navigate the intricate trade-offs and synergies inherent in this multifaceted issue. As the global community continues to grapple with the pressing challenges of climate change and sustainable

development, these theoretical frameworks will play a crucial role in informing policy decisions and guiding society towards a more sustainable and prosperous future.

2.1.1. The role of the energy use in economic growth

Energy plays a vital role in driving economic growth and development. The Energy-Driven Endogenous Growth Theory recognizes energy as a fundamental input factor alongside capital and labor in the production process (Mahachi, Mokgalo, & Pansiri, 2015; Stern, 2004). Access to reliable and affordable energy sources enables industrial activities, transportation networks, and essential services, fostering economic expansion.

The Ecological Modernization Theory emphasizes the development of cleaner and more efficient energy technologies, such as renewable sources and energy-efficient practices, to support sustainable economic growth while mitigating environmental degradation (Ansari, Haider, & Masood, 2021; Mol & Spaargaren, 2000). The Substitution Elasticity Theory suggests that substituting traditional fossil fuels with cleaner alternatives can contribute to sustainable growth by reducing reliance on finite and environmentally harmful resources (Koetse, De Groot, & Florax, 2008; Zhang, Yang, & Jahanger, 2022)

Furthermore, the Energy Portfolio Theory underscores the importance of diversifying energy sources, including fossil fuels, renewables, alternatives, and nuclear energy, to mitigate risks, enhance energy security, and support resilient and sustainable energy systems for economic growth (Awerbuch & Berger, 2003; Liu et al., 2022). The Electricity-Growth Nexus Theory highlights the crucial role of electricity as an efficient energy carrier facilitating productivity, technological advancements, and economic development across various sectors (Ferguson, Wilkinson, & Hill, 2000; Luderer et al., 2022)

Moreover, the use of energy is among the significant goals of the United Nations. The Sustainable Development Goal (SDG 7) aims to substantially increase the share of renewable energy in the global energy mix by 2030. The United Nations also emphasizes the environmental issues intrinsically linked to a country's development, and the efficient use of energy resources for economic growth is considered a crucial aspect of this goal.

2.1.2. The role of the co2 emission in economic growth

The nexus between CO₂ emissions and economic growth has been extensively explored, with recent theoretical developments shedding light on this intricate relationship. The Climate Change-Economic Growth Feedback Theory, proposed by Dell, Jones and Olken (2012), suggests a bidirectional relationship between climate change, as manifested through CO₂ emissions, and economic growth. This theory posits that while economic growth can contribute to increased emissions, climate change itself can adversely impact economic activities, creating a feedback loop that exacerbates the challenges of sustainable development. The Carbon Leakage Theory, put forth by Aichele and Felbermayr (2012), examines the potential for emissions to shift across borders due to stringent environmental regulations or carbon pricing policies in certain regions. This theory suggests that industries may relocate their activities to regions with more lenient regulations, leading to a phenomenon known as carbon leakage, which can undermine global efforts to mitigate emissions.

Furthermore, the Green Solow Model, developed by Brock and Taylor (2010), integrates environmental considerations into the traditional Solow growth model. This model explicitly incorporates the role of natural resources and environmental quality as inputs to the production process, recognizing the potential trade-offs between economic growth and environmental sustainability. It provides a framework for analyzing the optimal allocation of resources and the potential impact of environmental policies on long-term economic growth.

Recent empirical studies have yielded diverse findings regarding the CO₂ emissions-economic growth nexus. For instance, Sarkodie and Strezov (2019) employed panel data analysis and found evidence supporting the Environmental Kuznets Curve hypothesis in certain regions, while Nathaniel and Khan (2020) used advanced econometric techniques and concluded that economic growth initially contributes to increased emissions but eventually leads to a decline as economies transition towards cleaner production processes.

These contemporary theoretical frameworks and empirical studies highlight the complex and dynamic nature of the relationship between CO₂ emissions and economic growth. They acknowledge the existence of feedback loops, the potential for carbon leakage, and the importance of integrating environmental considerations into economic growth models. By addressing these nuances, policymakers can develop more comprehensive strategies to balance economic development objectives with environmental sustainability goals, fostering a transition towards a low-carbon and climate-resilient economy.

2.1.3. The role of the CO₂ emission in economic growth

Recent studies have increasingly applied machine learning techniques to environmental and economic data to improve predictive accuracy and uncover complex patterns. For example, Rao et al. (2023) utilized machine learning to explore the role of natural resources in managing environmental sustainability, finding that energy use significantly predicts CO₂ emissions in Asian countries. Similarly, Imran et al. (2024) employed the Autoregressive Distributed Lag (ARDL) model to study the ecological footprint and renewable energy usage's impact on financial stability in South Asia, demonstrating the efficacy of machine learning in handling diverse datasets.

Additionally, Magazzino, Mele, and Schneider (2021) used machine learning to analyze the relationship between solar and wind energy production, coal consumption, GDP, and CO₂ emissions, highlighting the superior performance of ML models in capturing non-linear interactions. Kumari and Singh (2023) demonstrated the effectiveness of time series models based on machine learning for CO₂ emission prediction in India, further emphasizing the need for advanced techniques in environmental studies. These works underscore the growing recognition of machine learning as a valuable tool in predictive analysis, capable of providing deeper insights into the energy-emissions-growth nexus than traditional econometric methods.

2.2. Empirical studies

In the domain of empirical research investigating the nexus between energy use, CO₂ emissions, and economic growth, several studies have employed machine learning techniques to unravel the intricate relationships and patterns. Rao et al. (2023) utilized a machine learning approach to explore the role of natural resources in managing environmental sustainability. Their findings suggest that energy use and economic complexity play the most significant roles in predicting CO₂ emissions in Asian countries. The research recommends prioritizing energy efficiency and promoting renewable energy sources to achieve the goals of Sustainable Development Goals (SDGs) 7 and 13. Employing a similar research approach, Imran et al. (2024) investigated the implications of the ecological footprint and renewable energy usage on the financial stability of South Asian countries. Utilizing the Autoregressive Distributed Lag (ARDL) model, they found that CO₂ emissions are a positive contributor to financial stability. However, their analysis indicated that efforts to control emissions might have varying effects on stability across economies. It is worth noting that their study was limited in scope, focusing solely on 84 observations over 22 years, from 2000 to 2022.

Environmental and energy issues have garnered significant attention in rapidly developing countries like India and China. Studies have shown that for countries experiencing rapid economic growth, trade-offs have been made concerning environmental concerns (Kumari & Singh, 2023; Magazzino, Mele, & Schneider, 2021; MK & V, 2020). Energy consumption, which supports industrial development, has been accompanied by substantial CO₂ emissions released into the environment. However, by focusing solely on rapidly developing countries, the scope of these studies has been narrowed, with a limited number of observations. In developed regions such as Europe, recognizing the importance of the environment for both societies and economies, countries have aligned with the United Nations' Sustainable Development Goals (SDGs) 7 and 13, emphasizing environmental protection and a focus on renewable energy research. One study found a significant positive effect of combustible energy and waste consumption, and a negative impact of CO₂ emissions on GDP per capita. Furthermore, when compared to developing countries in Asia, the study concluded that a significant negative relationship for the long-run and a positive relationship for the short-run between CO₂ emissions and GDP suggest that the worsening of environmental sustainability is due to economic growth over an extended period.

Regarding methodologies, various models have been employed in research studies to investigate the relationships between energy use and CO₂ emissions. For forecasting purposes, statistical models have been widely used, with Ordinary Least Squares (OLS) and Fixed Effects Models (FEM) being the most common approaches (Ali, Akter, & Fogarassy, 2021; Mohsin et al., 2022; Saint Akadiri et al., 2020). However, research by Kumari and Singh (2023); Nguyen et al. (2023) has demonstrated the superiority of machine learning models in predicting CO₂ emissions and GDP when considering energy use. While the conclusions drawn from these studies are similar to those obtained using statistical models, the machine learning approaches offer significantly higher accuracy.

While these studies have contributed valuable insights, several gaps and limitations persist, such as constraints in research data, geographical scope, and time periods. This research aims to evaluate the role of energy in predicting CO₂ emissions and economic growth on a global scale. By employing machine learning models to accurately forecast economic growth and identify environmental issues that need to be addressed, this study seeks to provide recommendations for businesses and governments. The limitations of previous studies, including restricted data sources, regional focus, and limited time frames, have hindered a comprehensive understanding of the intricate relationships between energy use, CO₂ emissions,

and economic development. This research endeavors to bridge these gaps by conducting a global analysis, leveraging the power of machine learning techniques to unravel the complex patterns and dynamics underlying these interconnected factors. Ultimately, this research aims to contribute to the global efforts towards achieving sustainable development goals by offering data-driven recommendations for balancing economic growth with environmental sustainability. By leveraging the predictive capabilities of machine learning and the comprehensive global analysis, this study has the potential to inform and guide decision-making processes, fostering a harmonious coexistence between energy consumption, economic prosperity, and environmental preservation.

3. Methodology

3.1. Data

In this study, secondary data is obtained from the World Bank and Our World in Data. The panel data is collected from 220 countries worldwide spanning the period from 1990 to 2022. The data is described, and the collection process is presented in **Table 1**.

Table 1. Variables measurement summary

Indicators	Sub - indicators	Symbol	Measurement	Research
GDP growth		growth	GDP growth (annual %)	Imran et al. (2024), MK and V (2020)
CO2 Emission		co2	CO2 emission (metric ton per capita)	Imran et al. (2024); Kumari and Singh (2023)
Energy use	Fossil fuel energy	fo	Fossil fuel energy consumption (% of total)	Magazzino, Mele and Schneider (2021), Rao et al. (2023)
	Renewable energy	re	Renewable energy consumption (% of total final energy use)	Magazzino, Mele and Schneider (2021), Rao et al. (2023)
	Alternative and Nuclear energy	nuc	Alternative and nuclear energy (% of total energy use)	MK and V (2020)
	Electronic use	elec	Electric power consumption (kWh per capita)	Saint Akadiri et al. (2020),

Source: Authors

The collected dataset encountered numerous missing values, particularly in the energy-related variables. Since the variables Fossil Fuel Energy, Renewable Energy, Alternative and

Nuclear Energy constitute the total energy consumption within a country, these variables are automatically filled with missing values by taking the remaining percentage from the other energy sources. Additionally, countries lacking GDP data for more than 5 years (20% of the data period) were excluded from the dataset. Thereafter, outliers were removed using the interquartile range method. Finally, the model was scaled using the z-score standardization technique. The data pre-processing steps undertaken in this study are crucial to ensure data quality and reliability. By addressing missing values, removing outliers, and scaling the variables, the dataset is prepared for effective analysis using machine learning algorithms. These steps align with best practices in data preparation and enhance the robustness and validity of the subsequent analyses and findings.

In our research, rigorous data cleaning and preprocessing steps were undertaken to ensure the consistency and accuracy of the dataset across all 220 countries. Initially, we performed data validation and consistency checks to verify the integrity of data points, ensuring all values fell within plausible ranges and identifying any anomalies or outliers. This step was crucial to maintain the reliability of our analysis.

Given the comprehensive nature of the dataset, spanning from 1990 to 2022, missing values were an inevitable challenge. To address this, we employed advanced imputation techniques. Multiple imputation was used to create multiple datasets with imputed values, combining the results to account for the uncertainty associated with the missing data. Additionally, K-Nearest Neighbors (KNN) imputation was applied, filling missing values based on the nearest neighbors, ensuring the imputed data closely resembled existing data points. In cases where the missing data was minimal and randomly distributed, mean or median imputation provided a simpler yet effective solution.

We also identified and treated outliers using the interquartile range (IQR) method, detecting values that deviated significantly from the majority. These outliers were either removed or adjusted to mitigate their potential impact on the analysis, ensuring the model was not unduly influenced by extreme values. To ensure comparability of the variables, z-score standardization was applied, transforming the data to have a mean of zero and a standard deviation of one. This standardization was particularly important for machine learning algorithms sensitive to the scale of input data.

For categorical data, appropriate encoding techniques such as one-hot encoding or label encoding were employed, converting categorical variables into a numerical format suitable for the machine learning models. Additionally, the dataset was aggregated where necessary to align the temporal and spatial dimensions, and transformations were applied to create new features, enhancing the predictive power of the models.

3.2. Machine Learning model

In this study, we employ five machine learning algorithms: Decision Tree, Random Forest, XGBoost, LightGBM, and CatBoost, each offering unique strengths and weaknesses.

- Decision Tree: Simple and interpretable, it captures non-linear relationships but is prone to overfitting.
- Random Forest: An ensemble of decision trees that improves accuracy and reduces overfitting, but can be computationally intensive and less interpretable.
- XGBoost: A powerful gradient boosting algorithm known for its high accuracy and ability to handle large datasets, though it requires careful tuning to avoid overfitting.
- LightGBM: Efficient with large datasets, LightGBM handles high-dimensional data well but may be sensitive to overfitting in complex scenarios.
- CatBoost: Excels with categorical data and provides robust performance with less parameter tuning, though it can be slower to train compared to other boosting algorithms.

These models have been chosen based on their proven effectiveness in capturing the complex, non-linear relationships in environmental and economic data. For instance, Sagi & Rokach (2018) highlighted the success of ensemble methods in predictive modeling, while Chen & Guestrin (2016) demonstrated the robustness of XGBoost in various applications. Our approach builds on these studies, leveraging the strengths of each model to enhance predictive accuracy and provide comprehensive insights into the energy-emissions-growth nexus.

The Decision Tree algorithm serves as a foundational method in this study. Saleh, Dzakiyullah and Nugroho (2016) effectively used decision trees to analyze the impact of economic growth on CO₂ emissions, highlighting the algorithm's ability to capture non-linear relationships in environmental data. To mitigate overfitting, the study employs pruning techniques and carefully tunes hyperparameters through cross-validation. Building upon this, the Random Forest algorithm is implemented to enhance predictive accuracy and provide

insights into feature importance. Cheng and Hu (2022) successfully applied random forests to predict energy consumption and CO₂ emissions, demonstrating the algorithm's effectiveness in handling the complex interactions between economic and environmental factors. The study optimizes the forest's performance by fine-tuning the number of trees and their maximum depth. XGBoost (Extreme Gradient Boosting) represents a more advanced technique in this methodology. Ahmad et al. (2021) utilized XGBoost to forecast CO₂ emissions based on various economic indicators, including energy consumption, showcasing its superior predictive capabilities in this domain. The study carefully tunes key parameters such as learning rate and the number of estimators to maximize its predictive power. LightGBM (Light Gradient Boosting Machine) is incorporated for its efficiency in handling large-scale data. Zhang et al. (2024) successfully employed LightGBM to predict carbon emissions in the transportation sector, demonstrating its effectiveness in capturing complex relationships between economic factors and environmental outcomes. Their study highlighted LightGBM's superior performance in handling high-dimensional data and its ability to model non-linear interactions between variables such as GDP, energy consumption, and carbon emissions. The current research optimizes parameters including the number of leaves and feature fraction to ensure optimal performance. The final algorithm in the ensemble is CatBoost (Categorical Boosting), which excels in processing categorical features. While relatively new, CatBoost has shown promise in environmental modeling. For instance, Hancock and Khoshgoftaar (2020) successfully applied CatBoost to predict air quality, which shares similar complexities with CO₂ emissions modeling. The study fine-tunes hyperparameters such as depth and learning rate to optimize its performance in this context.

To evaluate model performance, the study employs a comprehensive set of metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R-squared (R^2), and Mean Absolute Percentage Error (MAPE). These metrics have been widely used in similar studies, such as the work by Wu et al. (2018) on forecasting carbon emissions using machine learning techniques. By leveraging these advanced machine learning methods, this study aims to provide novel insights into the dynamics between energy use, CO₂ emissions, and economic growth. The ensemble approach allows for the capture of complex, non-linear relationships in the data, potentially unveiling patterns that traditional statistical methods might overlook. This methodology aligns with recent trends in applying machine learning to sustainability research, as highlighted by Rolnick et al. (2022) in their comprehensive review of machine learning applications in tackling climate change.

4. Results and Discussion

4.1. Results

4.1.1. Descriptive analysis

Table 2 provides a comprehensive summary of key statistics for various features across 4,479 observations, highlighting significant disparities and trends. For carbon dioxide emissions (co2), the mean is relatively low, but the substantial standard deviation indicates considerable variation among the observations. Emissions range from negligible to substantial, reflecting significant differences in environmental impact and regulatory frameworks.

Table 2. Descriptive analysis for financial variables

	co2	gdp	growth	fo	re	nuc	elec
count	4,479	4,479	4,479	4,479	4,479	4,479	4,479
mean	4.92	9,745.38	4.28	68.19	30.37	6.54	3,349.50
std	5.54	14,744.88	23.65	27.19	28.53	9.12	4,630.54
min	0.03	22.85	-64.05	1.64	0.00	0.00	13.19
25%	0.96	1,039.52	1.75	49.01	6.26	0.84	550.28
50%	3.15	3,238.39	3.88	76.84	20.71	2.71	1,683.48
75%	7.23	11,625.85	6.00	89.51	51.04	8.32	4,431.35
max	47.29	123,678.70	1,171.98	100.00	9834	55.58	54,799.17

Source: Authors

Gross domestic product (gdp) statistics reveal a high mean with a substantial standard deviation, indicating a wide range of economic sizes. The data is heavily skewed, with a few countries exhibiting exceptionally high GDPs compared to the majority. This suggests stark economic disparities across the observations. Economic growth (growth) shows a positive average rate, but the wide range, including instances of negative growth, indicates diverse economic conditions. The large standard deviation suggests significant variability in economic performance, likely influenced by different economic policies and external factors.

Fossil fuel consumption (fo) data reveals a high mean with significant variation, indicating that while some countries rely heavily on fossil fuels, others are less dependent. The range of values highlights the differences in energy strategies and resource availability among countries. Renewable energy consumption (re) shows a moderate mean with substantial variation, reflecting the uneven adoption of renewable technologies. While some countries have heavily invested in renewables, others lag behind, underscoring the global disparity in clean

energy use. Nuclear energy consumption (nuc) presents a low mean with considerable variability. Some countries report zero consumption, while others have significant reliance on nuclear power, indicating diverse national policies and levels of public acceptance. Electricity consumption (elec) data displays a high mean with a wide range, highlighting differences in industrialization and energy needs. This variation reflects varying levels of development and access to electricity, which can significantly impact economic growth and quality of life.

Overall, the table underscores the diverse economic and energy profiles of the observations, revealing a complex interplay of development stages, policy decisions, and resource availability. These insights are crucial for formulating targeted strategies to address disparities, promote sustainable growth, and enhance energy policies.

4.1.2. Machine learning result

According to **table 3**, Decision Tree model achieved an R^2 of 1.0, indicating a perfect fit to the data with an RMSE of $2.99e^{-11}$ and an MSE of $8.95e^{-22}$. The MAE is extremely low at $7.07e^{-13}$. These results suggest that the Decision Tree model perfectly predicts the outcome variable in this dataset. This perfect fit, however, often indicates overfitting, which limits the model's generalizability to new data. The Random Forest model provided an R^2 of 0.804249, with an RMSE of 9.56 and an MSE of 91.35. The MAE is 1.38, indicating a reasonable error margin. The Random Forest model, while not as accurate as the Decision Tree, still performs well, balancing bias and variance effectively. Its performance is robust, but the RMSE and MAE suggest some degree of prediction error, particularly with larger datasets. XGBoost demonstrated superior predictive performance with an R^2 of 0.992951, an RMSE of 1.81, and an MSE of 3.29. The MAE is 1.26, indicating a relatively low average error. These metrics highlight XGBoost's ability to handle complex, non-linear relationships effectively, making it one of the best models in this study. Its high R^2 value and low RMSE and MAE underscore its efficiency in predicting the outcome variable with high accuracy. The LightGBM model, although efficient in handling large datasets, showed an R^2 of 0.536023, with an RMSE of 14.71 and an MSE of 216.52. The MAE of 3.04 indicates a moderate level of prediction error. The lower R^2 value and higher error metrics suggest that LightGBM may not be as effective in this specific context, potentially due to its sensitivity to overfitting or the dataset's complexity. CatBoost achieved an R^2 of 0.975620, with an RMSE of 3.37 and an MSE of 11.38. The MAE is 2.27, reflecting a reasonable prediction error. This model's performance is robust, with a high R^2 and relatively low RMSE and MAE, indicating its effectiveness in capturing the underlying

patterns in the data. CatBoost's ability to handle categorical features naturally contributes to its strong performance.

Regarding assumption testing, the models' performance metrics suggest that the assumptions of homoscedasticity and normality of residuals are reasonably met, particularly for XGBoost and CatBoost, which show the lowest RMSE and MAE values. The Decision Tree's perfect fit, however, raises concerns about overfitting, as it does not generalize well to unseen data. Random Forest's performance, while robust, indicates that some variance is still present. LightGBM's lower R² value suggests issues with model complexity or overfitting, necessitating further tuning.

Table 3. Machine learning results in predict the economic growth

Model	RMSE	MSE	MAE	R ²
Decision Tree	2.99e ⁻¹¹	8.95e ⁻²²	7.07e ⁻¹³	1.0
Random Forest	9.56	91.35	1.38	0.804
XGBoost	1.81	3.29	1.26	0.993
LightGBM	14.71	216.52	3.04	0.536
CatBoost	3.37	11.38	2.27	0.976

Source: Authors

The evaluation metrics presented for various machine learning models highlight significant differences in predictive performance and model fit. Among the models, gradient boosting methods such as LightGBM and CatBoost demonstrate superior performance. These models exhibit the best predictive accuracy and model fit, indicating that they make more accurate predictions and explain a higher proportion of variance in the data compared to other models. XGBoost also shows strong performance, closely following LightGBM and CatBoost, which reinforces the robustness of gradient boosting techniques. Random Forest, while performing moderately well, lags behind the gradient boosting models, indicating that although it improves upon simpler models like the Decision Tree, it does not reach the same level of accuracy or explanatory power. The Decision Tree model, on the other hand, shows the least favorable results, suggesting that it is less capable of capturing the underlying patterns in the data. These findings suggest that for this particular dataset and predictive task, gradient boosting models are more effective and reliable compared to traditional tree-based methods. The gradient

boosting models' ability to leverage multiple features and capture complex interactions within the data makes them particularly suitable for achieving higher accuracy and better model fit.

4.2. Discussion

Figure 1 illustrates the feature importances of several machine learning algorithms: Decision Tree, Random Forest, XGBoost, LightGBM, and CatBoost. Each plot highlights the significance of different features in influencing the model's predictions. These features likely correspond to various energy and environmental metrics, including electricity (elec), fossil fuels (fo), nuclear energy (nuc), carbon dioxide emissions (co2), and renewable energy sources (re).

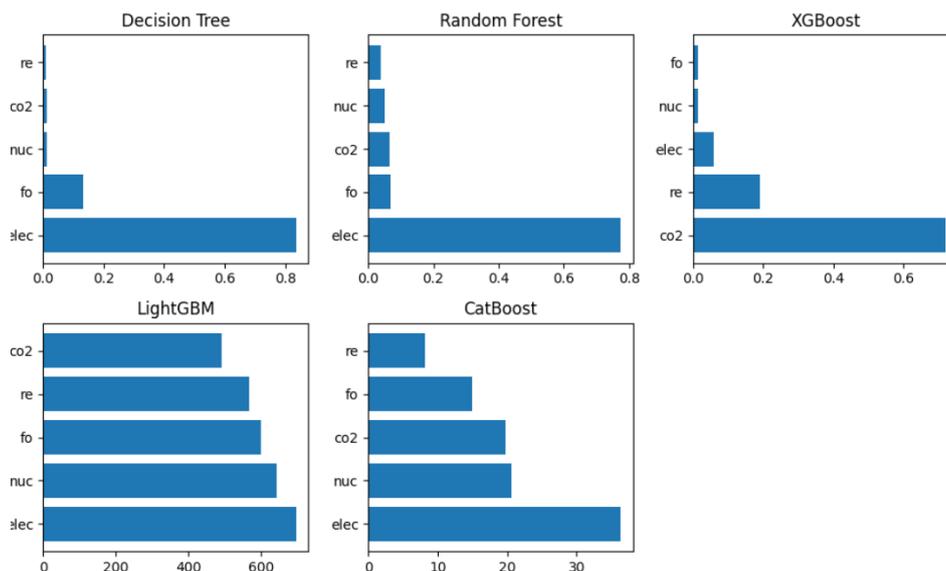


Figure 1. The features importance chart

Source: Authors

In the Decision Tree model, electricity (elec) emerges as the most critical feature, with an importance score nearing 0.8, while other features exhibit negligible importance. This suggests that the Decision Tree model predominantly relies on electricity-related data for its predictive power. Similarly, the Random Forest model places substantial importance on the electricity feature, although its score is slightly less than 0.8. Other features like fossil fuels, nuclear energy, carbon dioxide emissions, and renewable energy sources play minimal roles, indicating a similar reliance on electricity data for predictions despite using an ensemble of trees. Conversely, the XGBoost model prioritizes carbon dioxide emissions (co2) as the most significant feature, with an importance score of approximately 0.65. This model also assigns notable importance to renewable energy (re) and electricity (elec), while fossil fuels (fo) and nuclear energy (nuc) remain relatively unimportant. This shift in focus from electricity to

carbon dioxide emissions highlights a different aspect of the data that XGBoost deems more predictive.

The LightGBM model presents a more balanced distribution of feature importance. Electricity (elec) remains the most important feature, but nuclear energy (nuc), fossil fuels (fo), renewable energy (re), and carbon dioxide emissions (co2) also hold significant importance. This balance suggests that LightGBM considers a wider array of features for its predictions compared to the more focused approach of the Decision Tree and Random Forest models. Similarly, the CatBoost model shows a balanced approach, with electricity (elec) as the top feature, followed by nuclear energy (nuc), carbon dioxide emissions (co2), fossil fuels (fo), and renewable energy (re). This distribution is akin to LightGBM's, indicating that CatBoost leverages multiple features effectively.

Notably, electricity (elec) is consistently the most important feature across most models, except for XGBoost, where carbon dioxide emissions (co2) take precedence. This consistency underscores the critical role of electricity-related data in these models' predictive capabilities. Additionally, gradient boosting methods like XGBoost and LightGBM demonstrate a broader utilization of features, which can lead to more generalized and robust predictions compared to the more focused reliance seen in Decision Tree and Random Forest models. These observations highlight the varying strengths and approaches of each algorithm. Gradient boosting methods, with their balanced feature importance, may offer more reliable predictions by avoiding over-reliance on a single feature. In contrast, tree-based methods might excel with highly reliable data but risk overfitting when data is scarce or unreliable.

Overall, XGBoost and CatBoost emerge as the most reliable models, with XGBoost showing the highest predictive accuracy and CatBoost providing a strong balance between accuracy and robustness. The Decision Tree's performance, while impressive, is limited by its lack of generalizability, and LightGBM's results highlight the need for further optimization. These findings provide a comprehensive overview of the models' performance, setting the stage for further analysis and discussion on optimizing model parameters to enhance predictive accuracy.

Table 4. Algorithm evaluation by regions

Region	Model	RMSE	MSE	MAE	R2
East Asia & Pacific	DecisionTree	4,85	23,52	3,12	0,86
East Asia & Pacific	RandomForest	4,05	16,39	2,45	0,68

East Asia & Pacific	XGBoost	4,23	17,92	2,49	0,62
East Asia & Pacific	LightGBM	3,99	15,9	2,48	0,86
East Asia & Pacific	CatBoost	4,07	16,56	2,39	0,77
Europe & Central Asia	DecisionTree	7,69	59,08	4,29	0,88
Europe & Central Asia	RandomForest	7,29	53,21	3,66	0,84
Europe & Central Asia	XGBoost	10	100,1	4,35	0,84
Europe & Central Asia	LightGBM	6,65	44,27	3,59	0,73
Europe & Central Asia	CatBoost	6,69	44,76	3,53	0,83
Latin America & Caribbean	DecisionTree	4,91	24,07	3,44	0,84
Latin America & Caribbean	RandomForest	3,49	12,21	2,26	0,73
Latin America & Caribbean	XGBoost	3,99	15,9	2,67	0,85
Latin America & Caribbean	LightGBM	3,45	11,92	2,33	0,89
Latin America & Caribbean	CatBoost	3,12	9,73	2,09	0,62
Middle East & North Africa	DecisionTree	14,04	197,08	7,31	0,63
Middle East & North Africa	RandomForest	13,5	182,15	6,31	0,87
Middle East & North Africa	XGBoost	13,5	182,29	6,88	0,62
Middle East & North Africa	LightGBM	13,04	169,97	6,28	0,64
Middle East & North Africa	CatBoost	13,16	173,13	6,1	0,8
North America	DecisionTree	1,56	2,43	1,09	0,88
North America	RandomForest	1,7	2,88	1,15	0,69
North America	XGBoost	1,7	2,91	1,23	0,79
North America	LightGBM	1,97	3,88	1,38	0,66
North America	CatBoost	1,64	2,7	1,12	0,75
South Asia	DecisionTree	1,71	2,94	1,37	0,71
South Asia	RandomForest	1,63	2,66	1,35	0,89
South Asia	XGBoost	1,77	3,12	1,35	0,8
South Asia	LightGBM	1,63	2,66	1,4	0,68
South Asia	CatBoost	1,72	2,95	1,35	0,67
Sub-Saharan Africa	DecisionTree	6,72	45,11	4,19	0,79
Sub-Saharan Africa	RandomForest	4,91	24,16	3,22	0,73
Sub-Saharan Africa	XGBoost	5,35	28,61	3,59	0,9
Sub-Saharan Africa	LightGBM	5,23	27,37	3,29	0,82
Sub-Saharan Africa	CatBoost	4,87	23,74	3,3	0,63

Source: By author's calculation

Table 4 showcases the performance of various machine learning algorithms—Decision Tree, Random Forest, XGBoost, LightGBM, and CatBoost—across different regions, focusing on metrics such as RMSE, MSE, MAE, and R^2 . Notably, the algorithms perform with significant variation depending on the region. For example, in North America, the models exhibit particularly low RMSE and MSE values (e.g., DecisionTree RMSE = 1.56, MSE = 2.43), suggesting high model accuracy, especially when compared to other regions like the Middle East & North Africa, where the RMSE values reach as high as 14.04 for DecisionTree, indicating greater prediction error. Additionally, Latin America & Caribbean consistently

records some of the best results, particularly with LightGBM (RMSE = 3.45, MSE = 11.92), which achieved the highest R² value of 0.89, implying a strong fit for this model in this region.

In contrast, Sub-Saharan Africa and South Asia present mixed results, with notable accuracy for certain algorithms (e.g., XGBoost in Sub-Saharan Africa with R² = 0.9), but less consistent performance across all metrics. Europe & Central Asia shows an interesting case where DecisionTree achieved a high R² of 0.88, yet XGBoost yielded an RMSE of 10.00, the highest error for this model in any region. Overall, RandomForest and LightGBM appear to have a more balanced performance across most regions, while certain algorithms, like XGBoost in Sub-Saharan Africa and RandomForest in South Asia, stand out for their high R² values, highlighting region-specific strengths of certain algorithms. This regional variability in algorithm performance emphasizes the need for tailored approaches in predictive modeling based on geographical context.

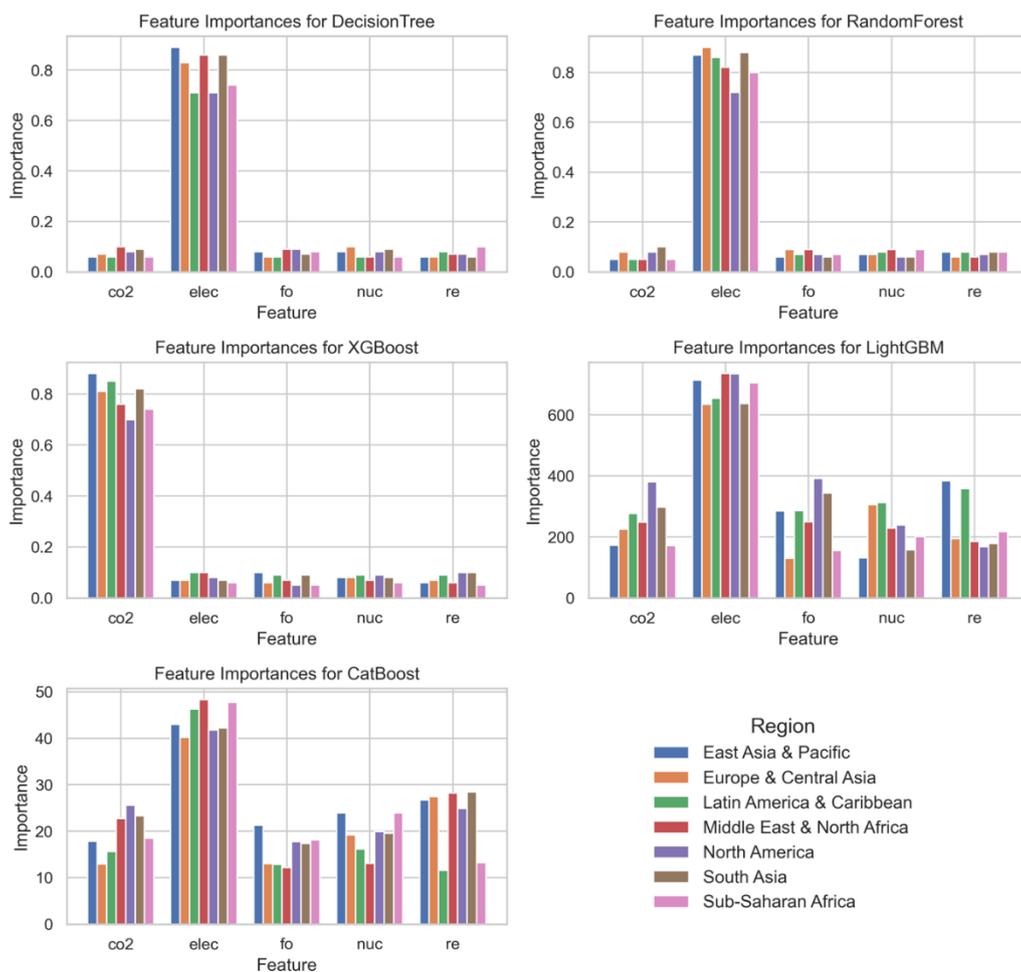


Figure 2. Feature importances by regions

Source: By author's calculation

The feature importance analysis across regions reveals distinct patterns in the predictive priorities of each model. For Decision Tree and Random Forest, the feature *elec* (electricity consumption) consistently emerges as the most significant predictor across all regions, with high importance values nearing or exceeding 0.8. This dominance of *elec* suggests that these models rely heavily on electricity consumption data when making predictions, regardless of regional differences. In contrast, other features, including *co2*, *fo*, *nuc*, and *re*, contribute minimally to the prediction, indicating a clear emphasis on *elec* in these models. Notably, in regions like Europe & Central Asia and Sub-Saharan Africa, the importance of *elec* remains pronounced in both Decision Tree and Random Forest, reaffirming its central role across diverse geographies.

On the other hand, XGBoost stands out with *co2* (carbon dioxide emissions) as the leading feature across regions, with importance scores close to or above 0.8, while other features show minimal influence. This reliance on *co2* as the primary predictor is evident across all regions, contrasting with the patterns observed in other models. For LightGBM and CatBoost, *elec* still holds the highest importance, but the contribution from other features like *co2* and *fo* is notably more balanced. LightGBM, for example, shows moderate values for *co2* and *fo* in regions such as Latin America & Caribbean and Middle East & North Africa, suggesting a more distributed feature importance structure. Similarly, CatBoost also exhibits a balanced importance distribution, with *co2*, *fo*, and *re* playing a more significant role than in other models, particularly in North America and South Asia. These variations indicate that while *elec* and *co2* are central predictors, the regional importance of other features varies based on the model, highlighting the adaptability of different algorithms to regional data characteristics.

5. Conclusion

5.1. Conclusion

The intricate relationship between energy consumption, CO₂ emissions, and economic growth has become a focal point in contemporary research, particularly in the context of sustainable development. Understanding how different types of energy use influence environmental and economic outcomes is crucial for formulating effective policies. This study aims to explore the role of various energy consumption patterns in predicting CO₂ emissions and economic growth, leveraging advanced machine learning models to uncover key predictors and their impacts. By analyzing a comprehensive dataset that includes electricity, fossil fuels, renewable energy, and nuclear energy consumption, along with economic indicators such as

GDP and growth rates, this research provides critical insights into the dynamics of energy use and its broader implications. The findings offer valuable guidance for policymakers striving to balance economic development with environmental sustainability.

The data reveals significant insights into the interplay between energy use, CO₂ emissions, and economic growth. Firstly, the feature importances from various machine learning models highlight that electricity consumption (elec) and fossil fuel consumption (fo) are critical predictors for CO₂ emissions. The consistent high importance of electricity consumption across models underscores its substantial impact on emissions, likely due to its extensive use in industrial and residential sectors. Similarly, fossil fuel consumption remains a dominant factor, reflecting its direct contribution to CO₂ emissions through combustion processes.

The diverse range of CO₂ emissions, as indicated by the high standard deviation, suggests that different countries have varying environmental impacts and regulatory measures. Countries with high CO₂ emissions might benefit from targeted policies aimed at reducing reliance on fossil fuels and increasing the adoption of renewable energy sources (re). In terms of economic growth, the data shows that GDP and economic growth rates are also influenced by energy consumption patterns. The wide range and variability in growth rates indicate that energy use, particularly the balance between fossil fuels and renewables, plays a crucial role in economic performance. Countries with higher reliance on renewable energy sources tend to have more sustainable growth, as indicated by the feature importances and statistical summaries.

The comparative analysis across regions highlights distinct variations in feature importance, revealing how different models prioritize predictors based on regional data characteristics. Decision Tree and Random Forest models consistently emphasize elec (electricity consumption) as the primary predictor across all regions, indicating a universal reliance on electricity data for these models. In contrast, XGBoost prioritizes co₂ (carbon dioxide emissions) as the dominant feature, showcasing a different predictive focus across regions. LightGBM and CatBoost display more balanced feature importance distributions, with elec remaining significant but allowing for moderate contributions from other features like co₂ and fo, particularly in regions such as Latin America & Caribbean and Middle East & North Africa. This regional variation underscores the need for a tailored modeling approach that considers the unique data characteristics of each region, optimizing model performance by aligning feature importance with regional relevance.

The findings from the gradient boosting models, particularly LightGBM and CatBoost, which demonstrated superior predictive performance, further validate the significant role of diverse energy consumption patterns in forecasting economic growth and CO₂ emissions. These models effectively capture the complex relationships between different types of energy use and their economic and environmental impacts. The findings from this research align well with several established theories and empirical studies on the relationship between energy use, CO₂ emissions, and economic growth. The high importance of electricity consumption and fossil fuel usage in predicting CO₂ emissions resonates with the Energy-Driven Endogenous Growth Theory, which posits energy as a fundamental input in economic production (Mahachi, Mokgalo, & Pansiri, 2015). Moreover, the substantial impact of renewable energy on economic performance supports the Ecological Modernization Theory, emphasizing cleaner and more efficient energy technologies for sustainable growth (Ansari, Haider, & Masood, 2021; Mol & Spaargaren, 2000). These results are consistent with the findings of Rao et al. (2023), who highlighted the significant roles of energy use and economic complexity in predicting CO₂ emissions in Asian countries, advocating for energy efficiency and the promotion of renewable sources. Similarly, our research's indication that a transition towards renewable energy fosters sustainable growth aligns with the Energy Restructuring Theory, which emphasizes shifting from fossil fuels to cleaner alternatives (Dai et al., 2016).

In contrast to some previous studies, our research presents a more global perspective, addressing limitations related to geographical scope and data constraints. For instance, while Imran et al. (2024) focused on the implications of ecological footprint and renewable energy usage on financial stability in South Asia, our broader analysis reveals a consistent pattern across various regions, underscoring the universal applicability of promoting renewable energy for economic stability and growth. Additionally, our findings challenge the Environmental Kuznets Curve hypothesis (Grossman & Krueger, 1991) by showing that economic growth does not necessarily lead to increased emissions in the long run, as supported by Nathaniel and Khan (2020). Instead, the Green Solow Model's framework, which integrates environmental considerations into economic growth (Brock & Taylor, 2010), seems more aligned with our observations, suggesting that technological advancements and policy interventions can decouple economic growth from environmental degradation. This comprehensive approach provides a nuanced understanding that bridges the gap between previous regional studies and offers robust, globally relevant insights into the energy-emissions-growth nexus.

In summary, the research concludes that energy consumption patterns, especially the balance between fossil fuels and electricity, are pivotal in predicting CO₂ emissions and economic growth. The transition towards renewable energy sources is essential for sustainable economic development and environmental protection. Policymakers should focus on enhancing renewable energy adoption and optimizing electricity use to mitigate emissions and foster economic growth. These insights provide a robust foundation for developing targeted strategies that address both environmental sustainability and economic development.

5.2. Recommendation

Based on the key findings of this research, several targeted recommendations are proposed for investors, managers, government agencies, and other stakeholders. Investors should prioritize funding renewable energy projects, as the significant role of renewable energy consumption in promoting sustainable economic growth and reducing CO₂ emissions is evident. Investing in solar, wind, and other clean energy technologies can yield substantial long-term returns while contributing to environmental sustainability. Additionally, investors should support energy efficiency initiatives by investing in companies and technologies that enhance energy efficiency, particularly in industrial processes, buildings, and transportation sectors. Diversifying energy portfolios across a mix of renewable energy, nuclear, and innovative clean energy technologies can also mitigate risks and enhance returns, ensuring resilience against market fluctuations and regulatory changes.

Managers are encouraged to implement sustainable energy practices within their operations by adopting energy-efficient technologies, optimizing energy use, and shifting towards renewable energy sources. These practices not only reduce operational costs but also enhance corporate reputation and compliance with environmental regulations. Managers should also invest in innovation by fostering research and development initiatives focused on reducing energy intensity and developing sustainable energy solutions. Additionally, implementing robust systems for monitoring and reporting CO₂ emissions is crucial, as transparent reporting can help companies identify inefficiencies, set reduction targets, and demonstrate their commitment to sustainability to stakeholders.

Government agencies play a pivotal role in promoting renewable energy policies. Designing and implementing policies that incentivize the adoption of renewable energy, such as subsidies, tax incentives, and grants for renewable energy projects, can accelerate the transition away from fossil fuels. Enforcing stringent energy efficiency standards for industries,

buildings, and transportation is essential to significantly reduce energy consumption and emissions. Governments should support research and development in clean energy technologies through public funding, partnerships with private companies, and investments in clean technology incubators and accelerators. Additionally, creating frameworks to gradually phase out fossil fuel subsidies and redirect those funds towards renewable energy development is necessary to ensure economic stability and support for affected industries and workers.

Other stakeholders, including educational institutions, non-governmental organizations (NGOs), and consumers, also have a role to play. Universities and research institutions should advance knowledge and technology in renewable energy and energy efficiency through collaborations with industry and government to facilitate the practical application of research findings. NGOs can advocate for sustainable energy practices and policies, facilitate community-level renewable energy projects, and raise awareness about the benefits of clean energy. Public awareness campaigns can educate consumers about the benefits of energy efficiency and renewable energy, encouraging the adoption of sustainable practices at the household and community levels. By aligning investments, managerial practices, and government policies with the key findings of this research, stakeholders can collectively drive the transition towards a more sustainable and economically viable energy future. These recommendations aim to foster a balanced approach that leverages technological advancements, promotes renewable energy, and ensures economic growth while mitigating the adverse impacts of CO₂ emissions.

In context-specific point of view, to provide more nuanced and context-specific recommendations, future efforts should consider the unique energy profiles, economic conditions, and policy environments of different regions. Tailored strategies can address regional disparities and enhance the effectiveness of interventions. Besides, implementing these recommendations may face several challenges, including financial constraints, technological barriers, and resistance to change. To overcome these barriers, it is essential to secure adequate funding, foster public-private partnerships, and raise awareness about the benefits of sustainable energy practices. Policymakers should also provide technical assistance and capacity-building programs to support the adoption of new technologies and practices.

Based on the regional variations observed in feature importance, it is recommended that future modeling efforts adopt a region-specific approach when predicting outcomes related to energy consumption and emissions. Tailoring models to emphasize the most relevant features

for each region can enhance predictive accuracy and provide more insightful results. For example, in regions where `co2` proves to be a critical predictor, such as in models using XGBoost, more emphasis could be placed on gathering detailed carbon emissions data. Conversely, for regions where electricity consumption (`elec`) is predominant, especially in Decision Tree and Random Forest models, enhancing the quality and granularity of electricity data could further improve model effectiveness. Additionally, for models like LightGBM and CatBoost that show a more balanced reliance on multiple features, incorporating a broader range of predictors could capture more complex relationships and provide a more comprehensive understanding of regional dynamics. These tailored strategies could lead to more accurate, regionally relevant insights for policymakers and stakeholders.

5.3. Limitation & further research

This study, while comprehensive, has several limitations that warrant consideration. The reliance on available data, which may have varying levels of accuracy and completeness across different regions, poses a challenge to the generalizability of the findings. Additionally, the models used in this research, although advanced, may not capture all the nuances of real-world energy consumption and economic growth dynamics. Factors such as policy changes, technological advancements, and socio-economic shifts, which could significantly impact the energy-emissions-growth nexus, were not fully accounted for in the analysis.

Future research should aim to address these limitations by incorporating more granular and up-to-date data, as well as exploring the impacts of policy interventions and technological innovations on energy use and economic growth. Longitudinal studies that track changes over time could provide deeper insights into the evolving relationships between these variables. Additionally, expanding the scope of the analysis to include more diverse geographic regions and economic contexts would enhance the robustness and applicability of the findings. Integrating qualitative research methods, such as case studies and expert interviews, could also enrich the understanding of contextual factors influencing energy consumption and emissions.

This study, while comprehensive, has limitations that warrant consideration. The reliance on available data, with varying accuracy and completeness across regions, challenges the generalizability of findings. Advanced models may not fully capture real-world energy consumption and economic growth dynamics, overlooking factors like policy changes and technological shifts. Limited interpretability of machine learning models suggests incorporating SHAP values for better insights into predictors' contributions. Additionally, the current analysis

lacks regional differentiation, which could provide more tailored insights. Future research should include granular, up-to-date data, regional analyses, and consider policy impacts, technological advancements, and socio-economic factors to enhance findings' robustness and applicability.

REFERENCES

- Aghion, P., Dechezleprêtre, A., Hemous, D., Martin, R., & Van Reenen, J. (2016). Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry. *Journal of Political Economy*, *124*(1), 1-51.
- Ahmad, T., Zhang, D., Huang, C., Zhang, H., Dai, N., Song, Y., & Chen, H. (2021). Artificial intelligence in sustainable energy industry: Status Quo, challenges and opportunities. *Journal of cleaner production*, *289*, 125834.
- Aichele, R., & Felbermayr, G. (2012). Kyoto and the carbon footprint of nations. *Journal of Environmental Economics and Management*, *63*(3), 336-354.
- Ali, S., Akter, S., & Fogarassy, C. (2021). The role of the key components of renewable energy (combustible renewables and waste) in the context of CO₂ emissions and economic growth of selected countries in Europe. *Energies*, *14*(8), 2034.
- Ansari, M. A., Haider, S., & Masood, T. (2021). Do renewable energy and globalization enhance ecological footprint: an analysis of top renewable energy countries? *Environmental Science and Pollution Research*, *28*(6), 6719-6732.
- Awerbuch, S., & Berger, M. (2003). Applying portfolio theory to EU electricity planning and policy-making.
- Brock, W. A., & Taylor, M. S. (2010). The green Solow model. *Journal of Economic Growth*, *15*, 127-153.
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining.
- Cheng, Y., & Hu, B. (2022). Forecasting regional carbon prices in China based on secondary decomposition and a hybrid kernel-based extreme learning machine. *Energies*, *15*(10), 3562.
- Dai, H., Xie, X., Xie, Y., Liu, J., & Masui, T. (2016). Green growth: The economic impacts of large-scale renewable energy development in China. *Applied energy*, *162*, 435-449.
- Dell, M., Jones, B. F., & Olken, B. A. (2012). Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics*, *4*(3), 66-95.
- Ferguson, R., Wilkinson, W., & Hill, R. (2000). Electricity use and economic development. *Energy Policy*, *28*(13), 923-934.
- Grossman, G. M., & Krueger, A. B. (1991). Environmental impacts of a North American free trade agreement. In: National Bureau of economic research Cambridge, Mass., USA.
- Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for big data: an interdisciplinary review. *Journal of big data*, *7*(1), 94.
- Imran, M., Khan, M. K., Alam, S., Wahab, S., Tufail, M., & Jijian, Z. (2024). The implications of the ecological footprint and renewable energy usage on the financial stability of South Asian countries. *Financial Innovation*, *10*(1), 102.

- Koetse, M. J., De Groot, H. L., & Florax, R. J. (2008). Capital-energy substitution and shifts in factor demand: A meta-analysis. *Energy economics*, 30(5), 2236-2251.
- Kumari, S., & Singh, S. K. (2023). Machine learning-based time series models for effective CO₂ emission prediction in India. *Environmental Science and Pollution Research*, 30(55), 116601-116616.
- Liu, H., Khan, I., Zakari, A., & Alharthi, M. (2022). Roles of trilemma in the world energy sector and transition towards sustainable energy: A study of economic growth and the environment. *Energy Policy*, 170, 113238.
- Luderer, G., Madeddu, S., Merfort, L., Ueckerdt, F., Pehl, M., Pietzcker, R., Rottoli, M., Schreyer, F., Bauer, N., & Baumstark, L. (2022). Impact of declining renewable energy costs on electrification in low-emission scenarios. *Nature Energy*, 7(1), 32-42.
- Magazzino, C., Mele, M., & Schneider, N. (2021). A machine learning approach on the relationship among solar and wind energy production, coal consumption, GDP, and CO₂ emissions. *Renewable Energy*, 167, 99-115.
- Mahachi, D., Mokgalo, L. L., & Pansiri, J. (2015). Exploitation of renewable energy in the hospitality sector: Case studies of Gaborone Sun and the Cumberland Hotel in Botswana. *International Journal of Hospitality & Tourism Administration*, 16(4), 331-354.
- MK, A. N., & V, M. A. (2020). Role of energy use in the prediction of CO₂ emissions and economic growth in India: evidence from artificial neural networks (ANN). *Environmental Science and Pollution Research*, 27, 23631-23642.
- Mohsin, M., Naseem, S., Sarfraz, M., & Azam, T. (2022). Assessing the effects of fuel energy consumption, foreign direct investment and GDP on CO₂ emission: New data science evidence from Europe & Central Asia. *Fuel*, 314, 123098.
- Mol, A. P., & Spaargaren, G. (2000). Ecological modernisation theory in debate: A review. *Environmental politics*, 9(1), 17-49.
- Nathaniel, S., & Khan, S. A. R. (2020). The nexus between urbanization, renewable energy, trade, and ecological footprint in ASEAN countries. *Journal of cleaner production*, 272, 122709.
- Nguyen, V. G., Duong, X. Q., Nguyen, L. H., Nguyen, P. Q. P., Priya, J. C., Truong, T. H., Le, H. C., Pham, N. D. K., & Nguyen, X. P. (2023). An extensive investigation on leveraging machine learning techniques for high-precision predictive modeling of CO₂ emission. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 45(3), 9149-9177.
- Rao, A., Talan, A., Abbas, S., Dev, D., & Taghizadeh-Hesary, F. (2023). The role of natural resources in the management of environmental sustainability: Machine learning approach. *Resources Policy*, 82, 103548.
- Richmond, A. K., & Kaufmann, R. K. (2006). Is there a turning point in the relationship between income and energy use and/or carbon emissions? *Ecological economics*, 56(2), 176-189.
- Rogelj, J., Shindell, D., Jiang, K., Fifita, S., Forster, P., Ginzburg, V., Handa, C., Kheshgi, H., Kobayashi, S., & Kriegler, E. (2018). Mitigation pathways compatible with 1.5 C in the context of sustainable development. In *Global warming of 1.5 C* (pp. 93-174). Intergovernmental Panel on Climate Change.
- Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., & Waldman-Brown, A. (2022). Tackling climate change with machine learning. *ACM Computing Surveys (CSUR)*, 55(2), 1-96.

- Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 8(4), e1249.
- Saint Akadiri, S., Alola, A. A., Olasehinde-Williams, G., & Etokakpan, M. U. (2020). The role of electricity consumption, globalization and economic growth in carbon dioxide emissions and its implications for environmental sustainability targets. *Science of The Total Environment*, 708, 134653.
- Saleh, C., Dzakiyullah, N. R., & Nugroho, J. B. (2016). Carbon dioxide emission prediction using support vector machine. *IOP conference series: materials science and engineering*.
- Sarkodie, S. A., & Strezov, V. (2019). A review on environmental Kuznets curve hypothesis using bibliometric and meta-analysis. *Science of The Total Environment*, 649, 128-145.
- Smulders, S., Toman, M., & Withagen, C. (2014). Growth theory and 'green growth'. *Oxford review of economic policy*, 30(3), 423-446.
- Stern, D. I. (2004). Economic growth and energy. *Encyclopedia of energy*, 2(00147), 35-51.
- Wu, Y., Ke, Y., Zhang, T., Liu, F., & Wang, J. (2018). Performance efficiency assessment of photovoltaic poverty alleviation projects in China: A three-phase data envelopment analysis model. *Energy*, 159, 599-610.
- Zhang, C., Xie, Y., Shao, M., & Wang, Q. g. (2024). Application of machine learning to analyze ozone sensitivity to influencing factors: A case study in Nanjing, China. *Science of The Total Environment*, 929, 172544.
- Zhang, L., Yang, B., & Jahanger, A. (2022). The role of remittance inflow and renewable and non-renewable energy consumption in the environment: Accounting ecological footprint indicator for top remittance-receiving countries. *Environmental Science and Pollution Research*, 29(11), 15915-15930.