International Journal of Energy Economics and Policy

ISSN: 2146-4553

available at http: www.econjournals.com

International Journal of Energy Economics and Policy, 2025, 15(4), 511-524.

Antecedents and Forecasting of Carbon Emissions Using Machine Learning Algorithms: Insights from the Top Ten Carbon-Emitting Nations

Mousa Gowfal Selmey[1]*, Bassam A. El bialy[2], Ahmed Hassanein[3,4], Abdalqader Ahmed Baker[5], Wael Mohamed Ali[6], Nagi Rashed Aboushadi[7], Abdullah Abdulaziz Alhumud[2], Elsayed Farrag Elsaid Mohamad[8,9]

[1]Department of Economics, Faculty of Commerce, Mansoura University, Mansoura, Egypt, [2]Department of Business Administration, College of Business, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia, [3]Gulf University for Science and Technology, Mubarak Al-Abdullah, Kuwait, [4]Mansoura University, Mansoura, Egypt, [5]Department of Economics, College of Law and Economics, Islamic University of Madinah, Medina, Saudi Arabia, [6]Department of Basic Sciences, Higher Future Institute for Specialized Technological Studies, Obour, Egypt, [7]Department of Economics, Faculty of Commerce, Mansoura University, Mansoura, Egypt, [8]Department of Economics, Faculty of Commerce, Damietta University, New Damietta, Egypt, [9]Department of Economics, College of Law and Economics, Islamic University of Madinah, Saudi Arabia.

*Email: mousa_gowfal@mans.edu.eg

Received: 03 February 2025 **Accepted:** 26 May 2025 **DOI:** https://doi.org/10.32479/ijeep.19622

ABSTRACT

This paper presents an analysis of predicting annual carbon emissions (CO_2 emissions) from 1990 to 2023 in the top ten high-carbon emission source countries using machine learning algorithms. The research employed a random forest algorithm, logistic regression, support vector machines (SVM), K-nearest neighbors (KNN), and gradient boosting to predict CO_2 emissions based on a set of selected features. The performance of these models was evaluated using root mean square error (RMSE), mean absolute error (MAE), R-squared, accuracy, precision, recall, F1 score, area under the curve (ROC AUC), and confusion matrix accuracy. The paper finds that Primary energy consumption is the main Antecedent and most influential factor, followed by population size and gross domestic product (GDP). The results also reveal that trade openness, urbanization, and renewable energy consumption have relatively minor impacts on the model's predictions. Furthermore, the results indicate that the Random Forest algorithm achieves near-perfect performance across all evaluation metrics for the prediction of carbon emission. The paper provides significant implications for policymakers and scholars in reducing CO_2 emissions.

Keywords: Carbon Emissions, Primary Energy Consumption, Top Ten Carbon-Emitting Nations, Machine Learning Algorithms
JEL Classifications: O24, Q4, Q5

1. INTRODUCTION

Environmental degradation is one of the most urgent environmental issues facing the world today. Environmental challenges and risks pose a significant obstacle to the achievement of sustainable development goals. Environmental risks arise from environmental degradation, which is linked to biodiversity loss, increased air pollution, natural resource degradation, and global warming (Nurgazina et al., 2022; Ali et al., 2023). Such degradation poses a significant threat to both the environmental and economic systems. Carbon emissions stand out as one of these challenges, contributing to the exacerbation of global warming and climate change and its associated environmental and economic repercussions.Carbon emissions stand out as one of these challenges, contributing to the exacerbation of global warming and climate change and its associated environmental and economic repercussions. Studies by Adebayo et al. (2021) and Acheampong and Opoku (2023) confirmed this finding. They found that rising carbon emissions are one of the main causes of climate change, which is what these risks are all about. It is anticipated that carbon dioxide emissions will persist in contributing to global warming. According to global data, one of the main causes of climate change between 1750 and 2005 was carbon dioxide (Balogh and Jámbor, 2017).

By releasing carbon dioxide and other greenhouse gases into the atmosphere, human activity has been a major contributor to climate change since the start of the Industrial Revolution. Consequently, industrial growth has exacerbated the ecological environment and accelerated the phenomenon of global warming, leading to the erosion of environmental quality (Adom et al., 2012). From the beginning of the 21st century until 2019, global greenhouse gas emissions showed an increasing trend, primarily due to the rise in emissions from China, India, the United States, major industrial countries, and other emerging economies, which produced the highest share of carbon emissions (European Commission, Joint Research Centre, 2023). In 2020, global emissions decreased by 3.7% compared to 2019 levels due to the COVID-19 pandemic, interrupting the continuously increasing trend for more than a decade. However, global greenhouse gas emissions restarted their growth immediately after the peak of the pandemic, rising by 2.3% in 2022 compared to 2019 and by 1.4% compared to 2021 (European Commission, Joint Research Centre, 2023).

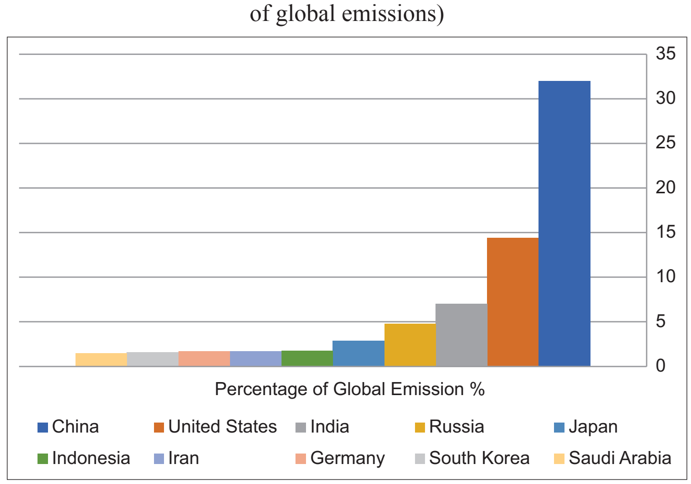
The 2022 Emissions Gap Report indicated that countries around the world are not on track to achieve the Paris Agreement's aim of limiting the global temperature from surpassing 1.5°C over pre-industrial levels, or 2°C. This level is considered the upper limit to prevent the worst potential consequences of climate change, which include catastrophic effects such as floods, storms, droughts, and rising sea levels (Adebayo et al., 2021; Emissions Gap Report, 2022). Present plans predict a temperature increase of 2.8°C by the end of the current century. Therefore, global low-carbon transformations are necessary to reduce projected greenhouse gas emissions by 28% by 2030, which would put the world on track towards a 2°C increase in worldwide temperature, and by 42% by 2023, which would put the world on track towards a 1.5°C rise in worldwide temperature, thereby averting a global catastrophe (Emissions Gap Report, 2023). Large industrial countries are responsible for most of these emissions because they use a lot of fuel-based energy to power their fast economic growth. This has been shown by many global reports and databases, such as emission gap reports, world population review data, the World Development Indicators database, and others. The world's greenhouse gas emissions hit a new high of 57,400 million tons of CO_2 equivalent in 2022. About two-thirds of these emissions come from burning fossil fuels and using electricity in factories (Emissions Gap Report, 2023; World Bank Database; World Population Review, 2024).

The top ten economies contributing to carbon emissions account for around 70% of the total global carbon emissions. Leading the list is China, which alone contributes about 32% of global carbon emissions, amounting to 12,667 million tons. Following China are the United States at approximately 14.4% with 4,853 million tons, India at 7% with 2,693 million tons, Russia at 4.8% with 1,909 million tons, Japan at 2.9% with 1,083 million tons, Indonesia at 1.75% with 692 million tons, Iran at 1.73% with 686 million tons, Germany at 1.7% with 674 million tons, South Korea at 1.6% with 635.5 million tons, and Saudi Arabia at 1.5% with 608 million tons (World Population Review, 2024, CO_2 Emissions), as shown in Figure 1.

Figure 1 shows that the 10 countries that are the largest sources of carbon emissions have a significant responsibility, as their contribution alone represents approximately 70% of the total carbon emissions worldwide. Therefore, it is the responsibility of those countries that lead in carbon emissions to reduce the level of global carbon emissions. these countries face challenges when carbon dioxide emissions arise from energy generation and its use for the industrialization process, particularly as the use of traditional energy sources increases in economic activities such as industrial production (Adebayo et al., 2021). In such cases, reducing carbon dioxide emissions could negatively impact economic growth, a concern that governments often hesitate to address. This is due to the fact that these nations have the biggest economies in the world and are always looking to boost their economic performance through manufacturing, urbanization, and the production and consumption of energy. Given that trade frequently entails the movement of manufacturing industries, which raises pollution levels, as well as the transportation of goods and pollution across borders, these governments also aim to maximize economic growth through international investment and exchange (Balogh and Jámbor, 2017). According to a number of studies, such as Adedoyin and Bekun (2020), Adebayo et al. (2021), Li et al. (2023), and Selmey et al. (2024), which indicated that the lack of political measures may incite an upward trend in emissions as a result of the growing demand for energy worldwide. This will have a detrimental effect on the environment because fossil fuels mostly cause greenhouse gas emissions.

Finally, carbon emissions are a huge environmental problem in the world at the current time, creating significant climate changes that may harm life as we know it. According to the data, the countries in question account for a significant portion of global emissions,

Figure 1: Carbon emissions contribution of the top ten countries (70% of global emissions)



Source: The authors

with a combined contribution of around 70% of global carbon emissions. To better understand the causes of the issue and offer workable remedies, it is necessary to look into the factors that contribute to these emissions. According to this viewpoint, the study attempts to draw attention to the environmental, social, and economic elements that support the rise in these emissions in the top ten nations that release carbon. Therefore, understanding the determinants of carbon emissions is crucial for guiding environmental and economic policies in line with the right path to achieving sustainable low-carbon economic growth goals. The study uses machine learning models to forecast carbon emissions to achieve this goal. Subsequently, it gives decision-makers the knowledge and policies they need to develop plans that promote economic growth while also protecting the environment.

The contribution of the current study is that it differs from previous literature in several essential aspects. Firstly, previous studies primarily examined the determinants of carbon emissions in specific countries or regions, during specific times, or using different analysis tools. while The current study focuses on the top ten carbon-emitting countries, which account for a large proportion of up to 70% of global carbon emissions. It highlights the crucial role of these countries in shaping global carbon emission trends. Furthermore, it underscores the importance of studying the determinants of carbon emissions and forecasting them in the top ten countries that emit them. Secondly, this study differs from others because it uses machine learning models to analyze large and complex data. This approach differs from traditional studies that rely on conventional econometric methods. Furthermore, this approach is particularly significant given the intricate interplay between economic, social, and environmental relationships, as well as the impact of policy uncertainty on carbon emissions in these countries. This methodology allows us to obtain more reliable patterns and predictions. Thus, the methodology used in the study provides a deeper understanding of the determinants of carbon emissions. Third, this study focuses on economic growth, trade openness, and inflation as economic and uncertainty factors without neglecting the impact of urbanization, population growth, and energy use on environmental quality. Finally, this study offers various perspectives and ideas to help policymakers create policies that promote economic growth while maintaining environmental quality. It also offers actionable recommendations that can boost economic growth with reduced emissions.

The study is structured as follows: Section 2 reviews the relevant literature and outlines the hypotheses. Section 3 describes the research methodology. Section 4 presents the findings and discusses their implications. Finally, Section 5 concludes the study.

2. LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

The literature on the factors influencing CO_2 emissions is highlighted in this section. We shall examine the following topics in this section:

2.1. Economic Growth and CO_2 Emissions

The relationship between economic growth and the quality of the environment has attracted interest from scholars because of the harmful influence of CO_2 emissions on climate change (Mikayilov et al., 2018; Mladenović et al., 2016; Osobajo et al., 2020; Idroes et al., 2023). Empirical evidence continues to show conflicting findings despite the large number of studies that have looked at the causal relationship between environmental quality and economic performance (Lapinskienė et al., 2017; Namahoro et al., 2021; Ayhan et al., 2023; Jebabli et al., 2023; Tsimisaraka et al., 2023; Ghazouani and Maktouf, 2024; Mohammed et al., 2024).

The findings of the many studies that have looked at the connection between carbon dioxide emissions and economic growth are mixed. Khan et al. (2024), for instance, investigated the relationships between technical innovation, carbon dioxide emissions, and economic growth in 35 countries along the Belt and Road between 1985 and 2019. Economic growth negatively impacts environmental quality, according to the findings. In addition, Khan et al. (2023) found that there is a positive correlation between economic growth and carbon dioxide. Raihan et al. (2023) explored the dynamic effects of economic growth on CO_2 emissions in Egypt from 1990 to 2019. The results concluded that economic growth in Egypt contributes to an increase in carbon emissions. In the same vein, Selmey and Elamer (2023) examined the impact of economic growth on environmental damage in Egypt between 1990 and 2018 using the ARDL methodology. The results revealed a positive relationship between economic growth and environmental damage in Egypt in both the short and long run. Acheampong (2018) investigated the connection between energy use, CO_2 emissions, and economic growth in 116 nations between 1990 and 2014. The results showed that throughout the Caribbean and Latin America, economic growth reduces CO_2 emissions. However, the results found no significant linkage between economic growth and CO_2 emissions in several areas, including both developed and emerging nations. Shuai et al. (2018) explored the main factors influencing China's carbon emissions. The findings demonstrated that real GDP per capita is the primary driver of carbon emissions.

Thus, the empirical evidence remains mixed despite numerous studies on the linkage between economic growth and environmental damage. One of the goals of this paper is to add to the body of research by looking into the cause-and-effect link between economic growth and environmental damage in a certain country or region. This will assist policymakers in formulating more effective policies for the environment that foster sustainable economic growth. Thus, we hypothesize the following:

H_1: There is a link between economic growth and carbon dioxide emissions.

2.2. Energy Consumption and CO_2 Emissions

One of the most important environmental issues that occupy researchers and policymakers worldwide is the linkage between the consumption of primary energy and carbon dioxide emissions. Since energy is the primary driver of economic activities, the heavy reliance on primary energy, such as fossil fuels, as the major source of energy leads to an increase in emissions of greenhouse gases, especially carbon dioxide, which exacerbates the climate change problem (Akpan and Akpan, 2012; Wang et al., 2016; Sarkodie & Ozturk, 2020; Rahman et al., 2021; Yuaningsih et al., 2021). Worldwide environmental agencies and leading environmental experts have indicated that the consumption of energy is the major element of pollution and environmental damage (Islam et al., 2021).

Since Kraft and Kraft's pioneering work in 1978, the linkage between consumption of energy and economic growth has garnered significant attention in environmental science and energy economics (Kraft and Kraft, 1975). Several studies conducted in recent decades have explored the dynamic linkage between energy, the economy, and the environment. For instance, Chen et al. (2016) utilized a panel co-integration and vector error-correction model to analyze the dynamic linkage among the economy, energy, and environment for 188 nations from 1993 to 2010. The empirical findings demonstrated the existence of long-term associations between energy usage and carbon emissions across all nations. In addition, a unidirectional causal relationship between energy consumption and carbon dioxide emissions in emerging and advanced economies. This demonstrates how changes in primary energy use have a direct effect on carbon emissions in every country. Liu et al. (2023) used panel co-integration tests and pooled mean group (PMG-ARDL) methods to examine the connection between energy consumption and carbon emissions in China between 1995 and 2020. According to the results, energy use, both immediately and over time, considerably worsens environmental impact. Selmey et al. (2024) studied the impact of primary energy consumption on carbon emissions across the BRICS economies (Brazil, Russia, India, China, and South Africa) between 1991 and 2023. The Panel Pooled Mean Group-Autoregressive Distributed Lag Model (PMG-ARDL) was used to show that the primary energy used has a significant and positive effect on carbon emissions in both the short and long term.

From the above, it is clear that many studies indicate a solid relationship between primary energy consumption and increased carbon emissions and environmental degradation. Regardless of the differences in the methods used and the periods studied, these studies agree that energy consumption plays a vital role in accelerating global climate change, and this has been corroborated by numerous studies (e.g., Anwar et al., 2020: Nathaniel and Adeleye, 2021). Thus, we hypothesize the following:

H_2: There is a link between primary energy consumption and carbon dioxide emissions.

2.3. Urbanization and CO_2 Emissions

The scholarly literature recognizes three major theories regarding urbanization and the environment: modernized ecological, urban transition, and compacted city theories (Poumanyvong and Kaneko, 2010; Kongkuah et al., 2022). Ecological modernization theory posits that urbanization is crucial in the economic transformation of the sectoral structure from agriculture to industry to services, which can mitigate environmental degradation resulting from economic growth. The urban transition hypothesis posits an association between wealth and environmental considerations in urban areas. It contends that as cities attain middle-income levels, their production, population, and consumption have been more concentrated, leading to increasing industrial pollution. Environmental regulations, technical advancements, and the transition from industrialization to service fields mitigate pollution as urban areas progress. Conversely, increased urban growth may elevate individuals' income and usage of energy-intensive goods, thereby harming the environment. The compact city theory is based on the premise that increased urbanization is advantageous. It asserts that urbanization, via the enhancement of existing urban areas, as opposed to suburban expansion, provides an economy of scale for infrastructure in the community, including transport and electricity generation, thereby reducing environmental damage (Burton, 2000; Lee and Lim, 2018; Kongkuah et al., 2022).

According to concepts mentioned previously, the impact of urbanization on the quality of the environment is ambiguous, as indicated by contradictory empirical results. Poumanyvong and Kaneko (2010) demonstrate that urbanization elevates carbon dioxide emissions in ninety-nine different economies. El-Aasar and Hanafy (2018) noticed that urbanization positively influenced carbon emissions in Sub-Saharan African economies. Kwakwa et al. (2018) found that urbanization raises the consumption of fossil fuels in Ghana, Kenya, and South Africa. Wang et al. (2017) and Grodzicki and Jankiewicz (2022) indicated that urbanization will result in rising CO_2 emissions. Moreover, adopting compact urban development forms would contribute to the reduction of carbon emissions. On the other hand, Ali et al. (2017) discovered a significant negative impact of urbanization on carbon emissions in Singapore during 1970-2015. Additionally, Khan et al. (2019) showed that urbanization has a negative impact on CO_2 emissions. Huo et al. (2020) showed that urbanization is a crucial determinant of building carbon emissions. Additionally, results indicate that the urban population and urban building floor space adversely affect carbon emissions in the urban building sectors. Kongkuah et al. (2022) revealed that urbanization significantly lowers long-term CO_2 emissions pollution. Based on the above, experimental studies have agreed that urbanization is one of the main determinants affecting carbon emissions, although the results varied. Thus, our study hypothesizes the following:

H_3: There is a link between urbanization and carbon dioxide emissions.

2.4. Renewable Energy and CO_2 Emissions

The investigation into the relationship between renewable energy (RE) and carbon emissions is essential, as it provides policymakers with additional insights for accomplishing the Sustainable Development Goals (SDGs). Overall, renewable energy may be an essential instrument in mitigating CO_2 emissions due to its reduced carbon footprint (Ansari. 2022; Elamer et al., 2022; Mirziyoyeva and Salahodjaev. 2022). Despite the variety of studies that have addressed the association between renewable energy usage and carbon emissions, the vast majority of them indicate a strong inverse relationship between them, confirming the role of renewable energy usage in reducing carbon emissions. For case in point, Li et al. (2025) conducted an examination of the impact of renewable energy usage on environmental quality in Costa Rica from 1990 to 2020. The results found that renewable energy usage improves environmental quality and decreases ecological footprint. Anwar et al. (2022) studied the relation between renewable sources of energy and carbon emissions in selected Asian economies from 1990 to 2014. The results showed that using renewable energy lowers carbon emissions.

However, numerous experimental studies have yielded results that diverge from the previously mentioned studies. Rehman et al. (2023) investigated the relationship between renewable energy use and carbon emissions at the worldwide level between 1985 and 2020. The outcomes revealed that the consumption of renewable energy does not affect carbon emission levels. Chen et al. (2022) explored the influence of renewable energy on carbon dioxide emissions per capita between 1995 and 2015 in 97 countries. The findings confirmed that using more renewable energy will not reduce per capita carbon emissions unless countries reach a certain threshold of consumption of renewable energy. According to Li and Haneklaus (2021), China's rapidly developing clean energy industry leads to contemporaneous increases in carbon dioxide emissions. A 1% increase in the use of renewable energy results in a short-term increase in CO_2 emissions of 0.285% to 0.288%.

Based on the above, experimental studies have shown varying results regarding the nature ofthe association between renewable energy usage and carbon emissions. Therefore, we can make the following hypothesis:

H_4: There is a link between Renewable Energy usage and carbon dioxide emissions.

2.5. population size and CO_2 Emissions

The population of the world is expanding swiftly, resulting in a continually rising need for energy consumption. Their heavy reliance on traditional energy sources, such as fossil fuels, leads to increases in CO_2 emissions and climate change through increased human activity. Studies like Wang et al. (2017), which studied the influence of population on carbon emissions in China between 1997 and 2012. The study's outcomes indicate that the population size significantly influences carbon emissions. Dong et al. (2018) explored the relation between carbon emissions and population growth in 128 economies between 1990 and 2014. The findings showed that population size has a significant impact on CO_2 emissions. Rehman and Rehman (2022) studied the linkage between carbon emissions and population growth in Asian countries during the period between 2001 and 2014. According to the findings of the grey relational analysis (GRA) processes, India's population growth is a major contributor to carbon emissions. Lee and Zhao (2023) examined the factors affecting carbon emissions in 96 countries from 2000 to 2020. The experimental results concluded that population growth exacerbates carbon emissions.

In contrast, certain studies have produced findings that differ from those of the aforementioned studies, including Begum et al. (2015), which looked at Malaysia's population growth and CO_2 emissions between 1970 and 1980. The findings showed that population growth has no discernible impact on carbon emissions per person. Sulaiman and Abdul-Rahim (2018) looked into the relationship between Nigeria's population increase and CO_2 emissions. The outcome demonstrated that, over the long run, population increase was not a driver of CO_2 emissions in Nigeria.

Based on the findings of earlier research, we must further explore this link, pinpoint the primary factors influencing the risks of climate change, and create policies that can achieve carbon neutrality in order to mitigate the dangers of environmental deterioration. As a result, we can make the following hypothesis:

H_5: There is a correlation between population size and carbon dioxide emissions.

3. RESEARCH METHODOLOGY

Five machine learning models were utilized in this study: random forest, logistic regression, gradient boosting, support vector machines (SVM), and K-nearest neighbors (KNN). Analyzing the training data is how all supervised machine learning models create their prediction function data.

This study employs machine learning models to evaluate the factors influencing CO_2 emissions and forecast those emissions in the top 10 nations based on annual data from 1990 to 2023. we rely on five models to evaluate their accuracy, then use the most accurate model to forecast CO_2 emissions while also identifying the primary factors that influence CO_2 emissions. Metrics such as accuracy, precision, recall, F1-score, and area under the curve (ROCAUC) analyze the performance comparison of the algorithms (Itoo et al., 2021). this paper implemented all machine learning algorithms using the Scikit-Learn package in the Python programming language.

3.1. Machine Learning Models

We used a variety of machine-learning techniques for comparison.

3.1.1. logistic regression

Logistic regression is a supervised machine-learning technique designed for learning classification problems. A classification problem with learning occurs when the desired variable is categorical. The objective of logistic regression is to establish a function that correlates the dataset's attributes with the targets, enabling the prediction of the probability that a new instance corresponds to one of the target classes (Bisong, 2019).

The logistic function, referred to as the logit or sigmoid function, constrains the output of the cost function to yield a probability between 0 and 1. The sigmoid or logit function is expressed as (Bisong, 2019):

$$h(t) = \frac{1}{1+e^{-t}} \tag{1}$$

Utilizing the probability logistic function πi, we derive the relevant logistic regression approach as follows:

$$\ln \underbrace{\left[\frac{\pi_i}{1-\pi_i}\right]}_{\textit{logit function}} = \beta_0 + \beta_1 x \tag{2}$$

β_0 represents the intercept, while β_1 denotes the impact of explanatory variables (EC, GDP, REC, URB, TO, Pop, EPU, INF) on the dependent variable (CO_2 emissions).

3.1.2. Random forest

Random forest is an assembly machine learning approach that integrates the outcomes of numerous models of decision trees (Hopp, 2024).

This approach involves the construction of decision trees utilizing variables from a matrix X and a stochastic selection of features. Furthermore, it entails the random selection of data subsets from X with replace to train every tree in the ensemble, differentiating it from other tree-based methodologies. Every tree produces a forecast of the desired variable (in this instance, annual CO_2 emissions), and the final model chooses the most frequently predicted outcome from the ensemble of trees (Tenorio and Pérez, 2023). Tiffin (2016) says that Random Forest effectively combines predictions from many trees and gives more weight to those with lower errors. This lessens the effect of possible individual errors as long as there isn't much of a link between the trees. This approach iteratively partitions the data in X into optimized sections and uses variable-based criteria to predict the desired variable; thereafter, it computes the dependent variable as the average of these regions (Tiffin, 2016; Tenorio and Pérez. 2023). Random forest models are based on the following equation (Tiffin, 2016):

$$F_0(x) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \gamma\right)^2 \quad (3)$$

Where γ is the predicted value, and y_i represents the actual value.

3.1.3. Gradient boosting trees

Friedman introduced the gradient boosting approach as an ensemble machine learning technique in 2001(Friedman, 2001). The main premise of the gradient-boosting approach is to amalgamate several weak learners to enhance the accuracy and robustness of the final algorithm (Yoon, 2021). the gradient-boosting procedure begins with the establishment of a single leaf and the generation of regression trees. A regression tree is a decision tree specifically constructed to estimate a continuous factual-valued function rather than to classify data. The gradient boosting approach initiates with the creation of a single leaf and the construction of regression trees. The regression tree is created using a method that repeatedly divides the data into nodes or splits it into smaller sets of data. Initially, every observation is categorized into a single category. The data is subsequently divided into two parts, employing every conceivable split on each accessible predictor. The predictor that splits the tree divides the observations most clearly into two groups and decreases the residual error (Yoon, 2021).

Boehmke and Greenwell sum up this method using the following formula (Tenorio and Pérez, 2023):

$$F(x) = \sum_{z=1}^{Z} F_z(x) \quad (4)$$

Where z represents the total number of trees that cumulatively collect the errors from all preceding trees. The first tree is represented as y = F1(x), followed by the second tree defined as F2(x) = F1(x) + e1, and this process continues consecutively to minimize F(x) as expressed below (Tenorio and Pérez, 2023):

$$L = \sum_{z} L\left(y_z, F_z(x)\right) \quad (5)$$

Consequently, as new decision trees add up, the accuracy of the ultimate predict progressively enhances, yielding more precise forecasts for annual CO_2 emissions.

3.2. Performance Evaluation Metrics for Algorithms

This paper initially used the mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R2) values to check the accuracy of the model (Yu et al., 2024). These measures are the most commonly utilized for regression issues. Reduced values for these markers signify superior model fits. It is denoted by the following equations (Güleryüz and Özden, 2020; Yu et al., 2024):

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}\left(x_i\right)\right)^2} \quad (6)$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{f}\left(x_i\right)\right| \quad (7)$$

MAE measures the average absolute difference between predicted and actual values.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{f}\left(x_i\right)\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{f}\left(x_i\right)\right)^2} \quad (8)$$

The term n denotes the number of observations, y_i denotes actual value, $\hat{f}\left(x_i\right)$ denotes the predicted value and $\bar{f}\left(x_i\right)$ denote the mean value. Larger R^2 denotes better model fitting, but larger MSE and MAPE values suggest higher prediction errors. RMSE values might range from 0 to ∞, with lower values signifying better performance. An RMSE value of 0 indicates that the technique has incurred no errors (Prayudani et al., 2019).

Additionally, this paper relies on important measures of different algorithms' performance, such as accuracy, precision, recall, and F1-score. The metrics above were calculated for both training and testing databases to ensure an accurate assessment of model performance. The confusion matrix shows performance across different classes, making it easier to identify true positives, false positives, true negatives, and false negatives. Furthermore, visual instruments like ROC curves were produced to demonstrate the trade-off between sensitivity and specificity, providing insights into the performance of models at different thresholds (Jain and Srihari. 2024).

3.3. Variable Importance

Feature importance indicates the ranking of each feature's importance in predicting carbon dioxide emissions. It is executed via a Python program. In an ensemble with decision trees, is computed as (Yu et al. 2024):

$$\operatorname{Im} p\left(x_j\right)=\frac{1}{K}\sum_{k=1}^{K}\sum_{z\in\varphi_k} I\left(j_z=j\right)\left[\frac{n_z}{N}\Delta i(z,s)\right] \tag{9}$$

Where z indicates the z^{th} non-terminal node of the decision tree φ_k, j_z indicates the feature identifier utilized to split node z, $I()$ denotes the indicator function, n_z represent the number of samples arriving at node z, N denotes the total number of samples, and $\Delta i(z,s)$ refers to the reduction in the Gini coefficient at the z^{th} node after s-splitting.

In addition, Lundberg and Lee (2017) introduced SHAP as a technique for elucidating the mechanisms of a particular instance by measuring the contribution of each feature. Many experts are using SHAP to understand social and environmental events (Stojić et al., 2019). According to Orji and Ukwandu (2024), SHAP is founded on game theory and its tenets, which include assigning zero attributions to variables that did not influence the model's prediction. Thus, SHAP seeks to interpret a model f at a certain point x^* using a value function e_S, where S is a subset of S ⊆ {1., p}. This function is generally defined as the expected value for a conditional distribution applied to each variable S is expressed as (Orji and Ukwandu, 2024):

$$e_S = E\left[f(x) \mid x_S = x_S^*\right] \tag{10}$$

According to Holzinger et al. (2022), the contribution of a variable j is represented by ϕj and computed as the weighted average across all potential S, as illustrated below (Orji & Ukwandu, 2024):

$$\phi_j(\mathrm{val}) = \sum_{S\subseteq\left\{x_1,\ldots,x_p\right\}\setminus\left\{x_j\right\}} \frac{|s|!\left(p-|s|-1\right)!}{p!}\left(\mathrm{val}\left(S\cup\left\{x_j\right\}\right)-\mathrm{val}\left(S\right)\right) \tag{11}$$

Where x denotes the feature value matrix for a case in the model being explained, p denotes the number of features, S denotes a subset of these features, and $val(S)$ denotes the prediction for the feature values inside the set S.

By figuring out the contribution of each variable and computing every potential combination of the variables that could account for the effect, SHAP evaluates the role of each component. The final contribution of the traits is then calculated by averaging these contributions (Orji and Ukwandu, 2024). According to Teoh et al. (2023), the Shapley value makes it easier to comprehend how characteristics contribute to the model's outcome and how much of an impact they have on the machine learning model's judgments. The absolute SHAPley values for every variable are determined using the following formula:

$$I_j = \sum_{i=1}^{n}\left|\phi_j^{(i)}\right| \tag{12}$$

Where $\phi_j^{(i)}$ denotes the SHAP value of the j-th feature for in this case i.

3.4. Sample and Variable Measurements

This study utilizes annual time-series data from 1990 to 2023 to analyze the key determinants of carbon dioxide emissions in the top ten highest-emitting countries. The sample includes both developed and emerging economies, allowing for a comparative analysis of emission patterns across different economic structures. Table 1 shows the sample countries. Developed nations in the sample—such as the United States, Japan, Germany, and South Korea—tend to have advanced industrial sectors and stringent environmental policies. Conversely, emerging economies like China, India, Indonesia, and Iran exhibit rapid industrialization and population growth, contributing to their high emission levels. The dataset consists of 370 observations, covering a mix of developed and emerging economies. These countries were selected based on their substantial contribution to global CO_2 emissions, making them central to understanding emission trends and the underlying factors influencing climate change.

Table 1 indicates that China, the United States, and India are the three largest emitters, collectively accounting for over half of global emissions from this group. China, classified as an emerging economy, contributes the highest share, representing 32% of the total emissions in the sample. The United States, a developed country, follows with 14.4%, while India, another emerging economy, accounts for 7%. Other significant contributors include Russia (4.8%), Japan (2.9%), and Indonesia (1.75%). By incorporating this diverse set of countries, the study provides a comprehensive perspective on global CO_2 emissions, highlighting the differences between industrialized and developing economies. This approach helps in identifying common factors influencing emissions while considering country-specific characteristics that may affect environmental policies and sustainability efforts.

This study incorporates a set of key economic, social, and environmental variables to analyze the determinants of carbon dioxide emissions. The variables, their symbols, measurement units, and data sources are outlined in Table 2.

By incorporating these variables, the study aims to assess how economic, demographic, and policy factors contribute to CO_2 emissions. The combination of macroeconomic and environmental indicators allows for a comprehensive analysis of emission trends

Table 1: Sample

Countries	Country classification	Observations	CO_2 emissions (%)
China	Emerging Country	37	32
United States	Developed Country	37	14.4
India	Emerging Country	37	7
Russia	Emerging Country	37	4.8
Japan	Developed Country	37	2.9
Indonesia	Emerging Country	37	1.75
Iran	Emerging Country	37	1.73
Germany	Developed Country	37	1.7
South Korea	Developed Country	37	1.6
Saudi Arabia	Emerging Country	37	1.5

Source: The authors

and potential mitigation strategies across different economic contexts.

4. RESULTS AND DISCUSSION

This study presents a novel approach to forecast annual CO_2 emissions in the top ten high-emission source nations from 1990 to 2023 using machine learning methods. The machine learning models generate robust predictions for the annual CO_2 emissions. In this part of our study, we present the main findings and determine the accuracy of the algorithm. We then compare these algorithms to determine which one is the most reliable

4.1. Performance Measurements

The paper examines the comparative effectiveness of five distinct predictive models: Random Forest Regressor, Logistic Regression, SVM, KNN, and Gradient Boosting. The objective is to compare the outcomes of each method in order to determine which model is the most accurate and efficient. For the sake of achieving this purpose, the research initially employs four validated statistical tools, as delineated in Equations 1, 2, and 3 (Adewale et al., 2024).

Table 3 compares the performance of five machine learning models—Random Forest, Logistic Regression, SVM, KNN, and Gradient Boosting—using three evaluation metrics: RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), and R-squared. Each metric provides insights into how well the models predict the target variable, with lower RMSE and MAE values indicating better performance and higher R-squared values (closer to 1), suggesting a better fit to the data. Random Forest emerges as one of the top-performing models, with an RMSE of 0.054 and an MAE of 0.039, indicating minimal prediction errors. Its R-squared value of 0.997 is exceptionally high, meaning it explains nearly all the variance in the data. The Gradient Boosting algorithm came next, displaying next values of 0.057606599, 0.041715394, and 0.997121027 for RMSE, MAE, and R-squared, respectively. The Random Forest model is clearly the best choice for this task, as it delivers highly accurate predictions and fits the data almost perfectly.

Table 2: Variable measurements

Variables	Symbol	Measurement	Source
Carbon Dioxide Emissions from Energy	CO_2	Million tonnes of carbon dioxide	Statistical Review of World Energy
Primary energy consumption	EC	(Exajoules)	Statistical Review of World Energy
GDP per capita	GDP	GDP (Constant 2015 US$)	World Development Indicators
Trade openness	TO	Trade (% of GDP)	World Development Indicators
Population size	Pop	Population, total	World Development Indicators
Urbanization	Urb	Urban population (% of total population)	World Development Indicators
Inflation	inf	Inflation, consumer prices (annual %)	World Development Indicators
Economic policy uncertainty	EPU		World Uncertainty Index

Source: The authors

Table 3: Performance measures of algorithms

Metrics	RMSE	MAE	R-squared
Random Forest	0.054137772	0.038820091	0.997457307
Logistic Regression	0.132591819	0.065575066	0.929616753
SVM	0.648633409	0.511449311	−0.684358255
KNN	0.380910366	0.286877296	0.874125224
Gradient Boosting	0.057606599	0.041715394	0.997121027

Source: Output by Python

Logistic Regression and KNN show moderate performance. Logistic regression has an RMSE of 0.133 and an MAE of 0.066, with an R-squared of 0.930, indicating a good but not outstanding fit. KNN performs slightly worse, with an RMSE of 0.381, an MAE of 0.287, and an R-squared of 0.874. Despite not being as robust as Random Forest or Gradient Boosting, these models still yield reasonable predictions and could be helpful in scenarios that prioritize simplicity or interpretability.

On the other hand, SVM performs poorly across all metrics. Its RMSE of 0.649 and MAE of 0.511 are significantly higher than the other models, indicating large prediction errors. Additionally, its R-squared value of −0.684 is particularly concerning, as a negative value suggests that the model performs worse than a simple horizontal line (baseline model). This makes SVM unsuitable for this task. In conclusion, Random Forest is the best-performing model for this dataset, offering near-perfect accuracy and fit. However, while the other models are viable alternatives, they fall short of the top performers. However, because SVM performs poorly, it should be avoided. The particular requirements of the task, such as the need for accuracy, interpretability, or computational efficiency, ultimately determine which model is best.

Using a number of important metrics, including accuracy, precision, recall, F1 score, ROC AUC, and confusion matrix accuracy, Table 4 assesses the performance of five machine learning models: Logistic Regression, Random Forest, SVM, Naive Bayes, KNN, and Gradient Boosting. These measures offer a thorough understanding of each model's overall predictive strength, false positive and false negative balance, and data classification accuracy. With flawless scores on every criterion, Random Forest and Gradient Boosting are clearly the best-performing models. With an accuracy of 1, both models accurately categories every incident in the dataset. They strike the ideal balance between detecting genuine positives and reducing false positives and false negatives because their precision, recall, and F1 scores are all 1. They are the greatest options for this classification task because of their outstanding performance, which is further supported by their ROC AUC of 1 and confusion matrix accuracy of 1.

Logistic regression and SVM also perform very well, though not perfectly. Both models achieve an accuracy of 0.985, with precision, recall, and F1 score values close to 1. Logistic Regression has a slight edge in Recall (1 compared to SVM's 0.970), meaning it correctly identifies all positive instances. In

Table 4: Performance measures of algorithms

Metrics	Accuracy	Precision	Recall	F1 Score	ROC AUC	Confusion matrix accuracy
Logistic Regression	0.985294118	0.970588235	1	0.985074627	1	34.49253731
Random Forest	1	1	1	1	1	1
SVM	0.985294118	1	0.96969697	0.984615385	0.998268398	1
Naive Bayes	0.911764706	0.909090909	0.909090909	0.909090909	0.967965368	1
KNN	0.985294118	1	0.96969697	0.984615385	0.997402597	1
Gradient Boosting	1	1	1	1	1	1

Source: output by Python

contrast, SVM has a slight edge in Precision (1 compared to Logistic Regression's 0.971), meaning it avoids false positives entirely. Their ROC AUC scores are nearly perfect (1 for logistic regression and 0.998 for SVM), and their confusion matrix accuracy is also strong. These models are highly reliable but fall just short of the top performers.

KNN performs similarly to logistic regression and SVM, with an accuracy of 0.985, a precision of 1, and a recall of 0.970. Its F1 Score of 0.985 and ROC AUC of 0.997 indicate strong performance, though not quite as excellent as Random Forest or Gradient Boosting. Like SVM, KNN achieves a confusion matrix accuracy of 1, meaning it correctly classifies all instances in the confusion matrix. It is a solid choice but not the best.

Naive Bayes is the weakest performer in this comparison. While its accuracy of 0.912 is still respectable, its precision, recall, and F1 score are all 0.909, indicating a higher rate of misclassification compared to the other models. Its ROC AUC of 0.968 is decent but significantly lower than the others. The other models clearly outperform Naive Bayes in this task despite achieving a confusion matrix accuracy of 1.

Based on the provided assessment measures, the Random Forest algorithm is the most appropriate of the five.

4.2. The Importance of Features in a Random Forest Algorithm

Table 5 and Figure 2 display a SHAP (Shapley Additive Explanations) summary plot. A machine-learning model uses it to determine the importance of different features. SHAP values provide a unified measure of feature importance by quantifying the contribution of each feature to the model's predictions. The plot ranks several features based on their mean absolute SHAP values, thereby reflecting their overall impact on the model's output. The results of Table 5 and Figure 2 help us determine the importance of the factors that impact carbon emissions. Energy consumption appears to be the most significant factor, with a relative importance of 83.4%. This result indicates that primary energy consumption, such as fossil fuels, has a substantial impact on carbon emissions. Consequently, energy consumption is the largest and most crucial determinant of increased carbon emissions in the studied countries; therefore, changes in energy usage significantly affect the anticipated CO_2 emissions.

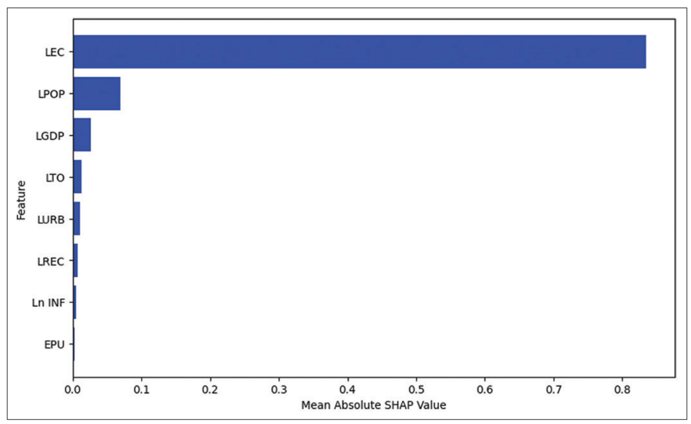
Many rapidly growing emerging economies, such as China and India, are classified among the highest carbon-emitting countries due to their reliance on non-renewable energy sources like coal,

Table 5: Indicators of feature importance

Features	Importance (Mean absolute SHAP value)
Energy Consumption	0.834055
Population Size	0.069203
Gross Domestic Product	0.02679
Trade Openness	0.013405
Urbanization	0.010171
renewable energy consumption	0.007035
Inflation	0.004613
Economic Policy Uncertainty	0.003354

Source: output by Python

Figure 2: Features importance using SHAP (Random forest)



Source: The authors

gas, and fossil fuels for economic growth. Particularly, following the implementation of the reform and opening-up policy, China has experienced rapid economic growth (Li et al., 2017). By 2013, China had become the world's largest source of carbon emissions, surpassing the United States (Mohammed et al., 2019). This is due to the Chinese economy's heavy reliance on primary energy consumption, particularly in the industrial sector, which accounts for approximately 80% of total energy consumption, as most industries produce energy-intensive products. This aligns with the findings of Mohammed et al. (2019), which showed that high energy consumption in China, particularly in the industrial sector, is the primary factor contributing to increased energy use and carbon dioxide emissions in the country. Moreover, for example, Saudi Arabia is a big oil producer that requires a larger consumption of energy. Several studies (e.g., Osobajo et al., 2020; Aller et al., 2021; Rehman and Rehman, 2022; Li et al., 2023; Yu et al., 2024; Zhong et al., 2024; Al-Azizi et al., 2025; Duran and Demirkale, 2025; Shaheen et al., 2025) have also come to the same conclusion: energy use is the main and most important cause of carbon emissions.

Therefore, it is important to offer insights into formulating policies that focus on enhancing energy efficiency, transitioning to clean energy sources, and assisting investments in sectors consisting of energy, transportation, production, and renewable energy sources.

The decomposition findings show the population size effect is the second most vital influencing component on CO_2 emissions in the top ten carbon-emitting nations, after the effect of primary energy intake, with a mean absolute SHAP value of approximately 7%. Still, it is difficult to ignore the importance of population size in influencing carbon emissions. Population size might indirectly influence carbon emissions since increasing demand for natural resources and energy results from population growth. The rising population size significantly influenced CO_2 emissions throughout the study period, highlighting the crucial impact of population growth on the escalation of CO_2 emissions, particularly in China, the US, India, and Saudi Arabia (Mohammed et al., 2019). These results align with the findings of Wu et al. (2016) and Du & Xia (2018), indicating that population size influences CO_2 emissions.

Furthermore, high population densities and rapid economic growth in economies like China and India cause substantial demand for energy to meet industrial, transportation, and household needs, producing more carbon emissions. On the other hand, other nations with rather low population densities, such as the United States, have seen high carbon emissions result from high per capita energy consumption, reflecting the consumerist character of the population's lifestyle. Therefore, understanding the relationship between population dynamics and carbon emissions is crucial to addressing climate change by developing effective population policies to reduce these emissions through improved education, women's empowerment, and the provision of reproductive health services.

The decomposition findings show the gross domestic product effect is the third most important influencing factor, after the influence of population size, which has an average absolute SHAP value of about 2.7%. Carbon dioxide emissions resulting from economic performance were directly linked to the growth rate of the gross domestic product. Industrialisation processes contributed to an increase in energy intensity in China, the United States, India, Japan, Iran, and Saudi Arabia, indicating a decline in energy usage efficiency due to a high demand for energy-intensive products (Mohammed et al., 2019). This, in turn, led to rapid and sustained growth in primary energy consumption and an increase in energy intensity.

Similar findings by Wang et al. (2015) indicated that accelerated economic growth played a vital role in increasing energy intensity. This, in turn, led to an increase in carbon emissions. This outcome aligns with previous research that has shown that growth positively affects the level of carbon dioxide emissions in China and other economies (e.g., Riti et al., 2017; Aslam et al., 2021; Zhang and Zhang, 2021).

Trade openness has a weak impact, with a mean absolute SHAP value of around 1.4%. While it contributes to the model's predictions, its influence is less pronounced compared to the top three features, such as energy consumption, population size, and GDP. This trend may be due to the fact that trade openness is linked to industrial activity in those countries, which relies on energy-intensive industries and thereby increases emissions. Policymakers should establish green trade policies, such as promoting environmentally friendly export industries or imposing restrictions on emissions from manufacturing and transportation.

The remaining variables, further down the list, including urbanization, renewable energy consumption, inflation, and economic policy uncertainty, exert weak significant influence, and their importance does not surpass 2% as a whole. Thus, this analysis helps identify which factors are most significant for the model, providing valuable insights for decision-making and further refinement.

Although trade openness, urbanisation, renewable energy consumption, inflation, and economic policy uncertainty are of little significance in influencing carbon emissions in this study, that does not necessarily mean that these variables are unimportant. Rather, their weak impact may indicate that current policies related to these factors are inadequate or that their impact is indirect. For example, renewable energy may have a limited impact due to its small share in the current energy mix, whereas urbanisation may increase emissions if it is not accompanied by sustainable policies. On the other hand, the weak impact of these variables can be viewed as an opportunity to direct efforts toward more influential factors, such as industrial growth and energy intensity, making policies more effective. However, we should not overlook these factors.

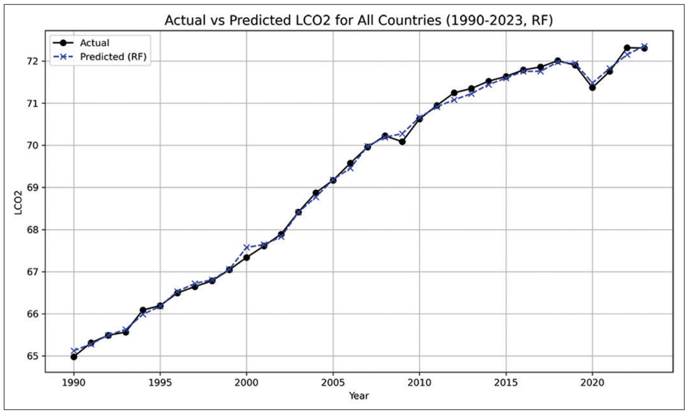
Table 6 and Figure 3 show that the random forest algorithm's predicted values for carbon dioxide emissions are very close to their real values in the study countries. In most years, the differences between the actual and expected values were minimal.

Table 6: Random Forest prediction for actual CO_2 emissions

Year	Ln CO_2 emissions actual values	Random Forest Prediction Values	Year	Ln CO_2 emissions actual values	Random Forest Prediction Values
1990	64.981	65.02485	2007	69.95966	69.98588
1991	65.31286	65.28138	2008	70.22466	70.09573
1992	65.49042	65.33997	2009	70.08702	70.20079
1993	65.5651	65.60343	2010	70.62546	70.57793
1994	66.08907	65.92903	2011	70.9453	71.03004
1995	66.19459	66.02802	2012	71.24643	71.11736
1996	66.49567	66.62056	2013	71.34774	71.30538
1997	66.64427	66.75611	2014	71.52371	71.48932
1998	66.78326	66.91322	2015	71.6349	71.61998
1999	67.04495	67.00887	2016	71.78761	71.72777
2000	67.3393	67.45607	2017	71.85906	71.83324
2001	67.60794	67.62111	2018	72.00818	72.13565
2002	67.89245	67.88594	2019	71.90228	71.96073
2003	68.4161	68.41946	2020	71.36908	71.62085
2004	68.87164	68.86475	2021	71.75538	71.83584
2005	69.17046	69.15989	2022	72.31559	72.19112
2006	69.57677	69.46385	2023	72.30778	72.18438

Source: output by Python

Figure 3: Actual vs predicted LCO_2 for All Countries (1990-2023, RF)



Source: Output by Python

This means that the predictions were very accurate and of high quality. Additionally, carbon emissions showed an upward trend during the study period from 1990 to 2023, with the exception of 2009 and 2020. The decrease in carbon emissions in 2009 and 2020 can be explained by global economic and social events that impacted economic activity and energy consumption, leading to a temporary decline in carbon emissions. The global crisis that began in 2008 caused a major economic contraction, particularly in heavily industrialised countries, such as the United States and the European Union. These developments resulted in a decrease in energy consumption and a deceleration in global trade, thereby impacting the air and sea transport sectors, which are major contributors to carbon emissions. The COVID-19 pandemic was the primary reason for the decline in carbon emissions in 2020. The decline was due to the slowdown in industrial and economic activity, as well as lockdown measures that halted air and sea transport, leading to a significant decline in energy consumption and carbon emissions. Therefore, these decreases were temporary, which underscores the need to develop long-term policies that lead to sustainable reductions in carbon emissions. Ultimately, forecasting carbon emissions alone is not sufficient; it must be accompanied by effective policies to reduce CO_2 emissions and address the negative impacts of climate change.

5. CONCLUSION

This study examined the main determinants of carbon emissions and predicted the CO_2 emissions in the top ten carbon-emitting economies globally, which include China, the United States, India, Russia, Japan, Indonesia, Iran, Germany, South Africa, and Saudi Arabia from 1990 to 2023. It used a novel methodology: machine learning algorithms like random forest, logistic regression, SVM, KNN, and gradient boosting. The study predicted the carbon emissions and identified the significant factors affecting them in the studied countries. The study results indicated that the random forest algorithm is the most accurate and reliable model with the lowest RMSE and MAE values and the highest value of R^2. In addition, the random forest model achieved perfect scores in accuracy, precision, recall, F1 scores, and confusion matrix accuracy.

The study found that primary energy consumption is the most important factor in determining carbon emissions in the ten countries that emit the most carbon. The mean absolute SHAP value of primary energy consumption is approximately 83.4%, indicating its significant role in explaining carbon emissions. This result confirms that economic growth and industrial activity heavily depend on primary energy, and thus, it is the main driver of increased carbon emissions in the sample countries. This result is consistent with Shaheen et al. (2025) and Li et al. (2023); they found that energy consumption is a major determinant of carbon emissions.

The results also revealed that the variable of population size has a significant impact on carbon emissions, with a mean absolute SHAP value of about 7%. The result indicates that the size of the population significantly contributes to explaining the changes in carbon emissions, although it ranks second after the major variable of primary energy consumption. These results suggest that population growth plays a fundamental role in driving the rise in carbon emissions. Population growth leads to an indirect increase in the demand for fossil fuels by raising the demand for goods, services, and environmentally polluting transportation. As for the variable of gross domestic product (GDP), the results indicated that the mean absolute SHAP value is 2.7%. This conclusion implies that although gross domestic product (GDP) is a factor influencing carbon emissions, its influence is rather smaller than that of other determinants such as population increase and energy consumption. This conclusion implies that GDP plays a less important role in determining variances in carbon emissions. In light of this, policymakers should take economic growth into account within carbon reduction policies by focusing on achieving low-carbon sustainable growth. The results indicated that the mean absolute SHAP value reached 1.3% and 1% for the variables of urbanization and trade openness, respectively. This conclusion means that urbanization and trade openness contribute little to explaining carbon emissions. Although its influence is small compared to other factors, the increase in urban populations may help explain higher emissions since urban energy consumption increases. However, policymakers should promote sustainable urban planning to ensure a reduction in carbon emissions in the future while adopting green trade policies, such as improving transport efficiency and encouraging less polluting exports. With regard to the other factors—renewable energy, inflation, and policy uncertainty—their combined mean absolute SHAP value for explaining carbon emissions is about 1.5%. The conclusion showed that renewable energy has not much effect on carbon emissions; it may be due to its low share in the energy consumption mix. It has to raise the percentage of renewable energy in the energy mix if we are to facilitate the change to a low-carbon economy. Furthermore, changes in government policies and political and economic ones could affect investments in energy and infrastructure, therefore indirectly affecting carbon emissions.

In conclusion, this paper successfully demonstrates the capability of machine learning models for CO_2 emissions prediction, particularly the Random Forest Regressor. The findings offer the policymakers a robust framework and valuable insights into the main drivers of emissions and offer clear policies that aim to alleviate climate change.

FUNDING

This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) (grant number IMSIU-DDRSP2502)

REFERENCES

Acheampong, A.O. (2018), Economic growth, CO_2 emissions and energy consumption: What causes what and where? Energy Economics, 74, 677-692.

Acheampong, A.O., Opoku, E.E.O. (2023), Environmental degradation and economic growth: Investigating linkages and potential pathways. Energy Economics, 123, 106734.

Adebayo, T.S., Akinsola, G.D., Odugbesan, J.A., Olanrewaju, V.O. (2021), Determinants of environmental degradation in Thailand: Empirical evidence from ARDL and wavelet coherence approaches. Pollution, 7(1), 181-196.

Adedoyin, F.F., Bekun, F.V. (2020), Modelling the interaction between tourism, energy consumption, pollutant emissions and urbanization: Renewed evidence from panel VAR. Environmental Science and Pollution Research, 27(31), 38881-38900.

Adewale, M.D., Ebem, D.U., Awodele, O., Sambo-Magaji, A., Aggrey, E.M., Okechalu, E.A., & Oluyide, O.P. (2024), Predicting gross domestic product using the ensemble machine learning method. Systems and Soft Computing, 6, 200132.

Adom, P.K., Bekoe, W., Amuakwa-Mensah, F., Mensah, J.T., Botchway, E. (2012), Carbon dioxide emissions, economic growth, industrial structure, and technical efficiency: Empirical evidence from Ghana, Senegal, and Morocco on the causal dynamics. Energy, 47(1), 314-325.

Akpan, U.F., Akpan, G.E. (2012), The contribution of energy consumption to climate change: A feasible policy direction. International Journal of Energy Economics and Policy, 2(1), 21-33.

Al-Azizi, Z.H., Rahmawati, F., Rokhmani, L. (2025), Analysis of the effect on economic growth, foreign investment, population and energy consumption on Co_2 in asean economies. Journal of Applied Business Taxation and Economics Research, 4(3), 378-394.

Ali, H.S., Abdul-Rahim, A.S., Ribadu, M.B. (2017), Urbanization and carbon dioxide emissions in Singapore: Evidence from the ARDL approach. Environmental Science and Pollution Research, 24, 1967-1974.

Ali, U., Guo, Q., Nurgazina, Z., Sharif, A., Kartal, M.T., Depren, S.K., Khan, A. (2023), Heterogeneous impact of industrialization, foreign direct investments, and technological innovation on carbon emissions intensity: Evidence from Kingdom of Saudi Arabia. Applied Energy, 336, 120804.

Aller, C., Ductor, L., and Grechyna, D. (2021), Robust determinants of CO_2 emissions. Energy Economics, 96, 105154

Ansari, M.A. (2022), Re-visiting the environmental kuznets curve for ASEAN: A comparison between ecological footprint and carbon dioxide emissions. Renewable and Sustainable Energy Reviews, 168, 112867.

Anwar, A., Sinha, A., Sharif, A., Siddique, M., Irshad, S., Anwar, W., Malik, S. (2022), The nexus between urbanization, renewable energy consumption, financial development, and CO_2 emissions: Evidence from selected Asian countries. Environment Development and Sustainability, 24, 6556-6576.

Anwar, A., Younis, M., Ullah, I. (2020), Impact of urbanization and economic growth on CO_2 emission: A case of far East Asian Countries. International Journal of Environmental Research and Public Health, 17(7), 2531.

Aslam, B., Hu, J., Majeed, M.T., Andlib, Z., Ullah, S. (2021), Asymmetric macroeconomic determinants of CO_2 emission in China and policy approaches. Environmental Science and Pollution Research, 28, 41923-41936.

Ayhan, F., Kartal, M.T., Kılıç Depren, S., Depren, Ö. (2023), Asymmetric effect of economic policy uncertainty, political stability, energy consumption, and economic growth on CO_2 emissions: Evidence from G-7 countries. Environmental Science and Pollution Research, 30(16), 47422-47437.

Balogh, J.M., Jámbor, A. (2017), Determinants of CO_2 emission: A global evidence. International Journal of Energy Economics and Policy, 7(5), 217-226.

Begum, R.A., Sohag, K., Abdullah, S.M.S., Jaafar, M. (2015), CO_2 emissions, energy consumption, economic and population growth in Malaysia. Renewable and Sustainable Energy Reviews, 41, 594-601.

Bisong, E. (2019), Building machine learning and deep learning models on Google cloud platform. Berkeley, CA: Apress. p. 59-64.

Burton, E. (2000), The compact city: Just or just compact? A preliminary analysis. Urban studies, 37(11), 1969-2006.

Chen, C., Pinar, M., Stengos, T. (2022), Renewable energy and CO_2 emissions: New evidence with the panel threshold model. Renewable Energy, 194, 117-128.

Chen, P.Y., Chen, S.T., Hsu, C.S., Chen, C.C. (2016), Modeling the global relationships among economic growth, energy consumption and CO_2 emissions. Renewable and Sustainable Energy Reviews, 65, 420-431.

Dong, K., Hochman, G., Zhang, Y., Sun, R., Li, H., Liao, H. (2018), CO_2 emissions, economic and population growth, and renewable energy: Empirical evidence across regions. Energy Economics, 75, 180-192.

Du, W.C., and Xia, X.H. (2018), How does urbanization affect GHG emissions? A cross-country panel threshold data analysis. Applied energy, 229, 872-883.

Duran, N.İ., Demirkale, Ö. (2025), Examining the relationship between climate change, energy consumption, and the financial system: The case of E7 Countries. International Journal of Advances in Engineering and Pure Sciences, 37(UYIK 2024 Special Issue), 77-87.

El-Aasar, K.M., Hanafy, S.A. (2018), Investigating the environmental Kuznets curve hypothesis in Egypt: The role of renewable energy and trade in mitigating GHGS. International Journal of Energy Economics and Policy, 8(3), 177-184.

Elamer, A.A., Elbialy, B.A., Alsaab, K.A., Khashan, M.A. (2022), The impact of COVID-19 on the relationship between non-renewable energy and Saudi stock market sectors using wavelet coherence approach and neural networks. Sustainability, 14(21), 14496.

European Commission, Joint Research Centre. (2023), EDGAR - Emissions Database for Global Atmospheric Research: CO_2 Emissions. Available from: https://www.unep/org/resources/emissions-gap-report-2023

Friedman, J.H. (2001), Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189-1232.

Ghazouani, T., Maktouf, S. (2024), Impact of natural resources, trade openness, and economic growth on CO_2 emissions in oil□exporting countries: A panel autoregressive distributed lag analysis. Natural Resources Forum, 48(1), 211-231.

Grodzicki, T., Jankiewicz, M. (2022), The impact of renewable energy and urbanization on CO_2 emissions in Europe -spatio-temporal approach. Environmental Development, 44, 100755.

Güleryüz, D., Özden, E. (2020), The prediction of brent crude oil trend using LSTM and Facebook prophet. Avrupa Bilim ve Teknoloji Dergisi, (20), 1-9.

Holzinger, A., Saranti, A., Molnar, C., Biecek, P., and Samek, W. (2022), Explainable AI methods-a brief overview. In International workshop on extending explainable AI beyond deep models and classifiers Cham: Springer International Publishing. p. 13-38.

Hopp, D. (2024), Benchmarking econometric and machine learning methodologies in nowcasting GDP. Empirical Economics, 66(5), 2191-2247.

Huo, T., Li, X., Cai, W., Zuo, J., Jia, F., Wei, H. (2020), Exploring the impact of urbanization on urban building carbon emissions in China: Evidence from a provincial panel data model. Sustainable Cities and Society, 56, 102068.

Idroes, G.M., Hardi, I., Noviandy, T.R., Sasmita, N.R., Hilal, I.S., Kusumo, F., Idroes, R. (2023), A deep dive into Indonesia's CO_2 emissions: The role of energy consumption, economic growth and natural disasters. Ekonomikalia Journal of Economics, 1(2), 69-81.

Islam, M.M., Khan, M.K., Tareque, M., Jehan, N., Dagar, V. (2021), Impact of globalization, foreign direct investment, and energy consumption on CO_2 emissions in Bangladesh: Does institutional quality matter? Environmental Science and Pollution Research, 28(35), 48851-48871.

Itoo, F., Meenakshi, E., Singh, S. (2021), Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection. International Journal of Information Technology, 13(4), 1503-1511.

Jain, M., Srihari, A. (2024), Comparison of machine learning algorithm in intrusion detection systems: A review using binary logistic regression. International Journal of Computer Science and Mobile Computing, 13, 45-53.

Jebabli, I., Lahiani, A., Mefteh-Wali, S. (2023), Quantile connectedness between CO_2 emissions and economic growth in G7 countries. Resources Policy, 81, 103348.

Khan, H., Weili, L., Khan, I., Zhang, J. (2023), The nexus between natural resources, renewable energy consumption, economic growth, and carbon dioxide emission in BRI countries. Environmental Science and Pollution Research, 30(13), 36692-36709.

Khan, I., Zhong, R., Khan, H., Dong, Y., Nuţă, F.M. (2024), Examining the relationship between technological innovation, economic growth and carbon dioxide emission: Dynamic panel data evidence. Environment Development and Sustainability, 26(7), 18161-18180.

Khan, M.K., Teng, J.Z., Khan, M.I., Khan, M.O. (2019), Impact of globalization, economic factors and energy consumption on CO_2 emissions in Pakistan. Science of the Total Environment, 688, 424-436.

Kongkuah, M., Yao, H., Yilanci, V. (2022), The relationship between energy consumption, economic growth, and CO_2 emissions in China: The role of urbanisation and international trade. Environment Development and Sustainability, 24, 4684-4708.

Kraft, J., Kraft, A. (1978), On the relationship between energy and GNP. The Journal of Energy and Development, 3, 401-403.

Kwakwa, P.A., Adu, G., Osei-Fosu, A.K. (2018), A time series analysis of fossil fuel consumption in Sub-Saharan Africa: Evidence from Ghana, Kenya and South Africa. International Journal of Sustainable Energy Planning and Management, 17, 31-44.

Lapinskienė, G., Peleckis, K., Slavinskaitė, N. (2017), Energy consumption, economic growth and greenhouse gas emissions in the European Union countries. Journal of business economics and management, 18(6), 1082-1097.

Lee, C.C., Zhao, Y.N. (2023), Heterogeneity analysis of factors influencing CO_2 emissions: The role of human capital, urbanization, and FDI. Renewable and Sustainable Energy Reviews, 185, 113644.

Lee, J.H., Lim, S. (2018), The selection of compact city policy instruments and their effects on energy consumption and greenhouse gas emissions in the transportation sector: The case of South Korea. Sustainable Cities and Society, 37, 116-124.

Li, B., Haneklaus, N. (2021), The role of renewable energy, fossil fuel consumption, urbanization and economic growth on CO_2 emissions in China. Energy Reports, 7, 783-791.

Li, J., Irfan, M., Samad, S., Ali, B., Zhang, Y., Badulescu, D., Badulescu, A. (2023), The relationship between energy consumption, CO_2 emissions, economic growth, and health indicators. International Journal of Environmental Research and Public Health, 20(3), 2325.

Li, N., Zhang, X., Shi, M., Zhou, S. (2017), The prospects of China's long-term economic development and CO_2 emissions under fossil fuel supply constraints. Resources Conservation and Recycling, 121, 11-22.

Li, X., Aghazadeh, S., Liaquat, M., Nassani, A.A., Eweade, B.S. (2025), Transforming costa rica's environmental quality: The role of renewable energy, rule of law, corruption control, and foreign direct investment in building a sustainable future. Renewable Energy, 239, 121993.

Liu, H., Wong, W.K., Cong, P.T., Nassani, A.A., Haffar, M., Abu-Rumman, A. (2023), Linkage among urbanization, energy consumption, economic growth and carbon emissions. Panel data analysis for China using ARDL model. Fuel, 332, 126122.

Lundberg, S.M., and Lee, S.I. (2017), A unified approach to interpreting model predictions. Advances in neural information processing systems, 30, 4765-4774.

Mikayilov, J.I., Galeotti, M., Hasanov, F.J. (2018), The impact of economic growth on CO_2 emissions in Azerbaijan. Journal of Cleaner Production, 197, 1558-1572.

Mirziyoyeva, Z., Salahodjaev, R. (2022), Renewable energy and CO_2 emissions intensity in the top carbon intense Countries. Renewable Energy, 192, 507-512.

Mladenović, I., Sokolov-Mladenović, S., Milovančević, M., Marković, D., Simeunović, N. (2016), Management and estimation of thermal comfort, carbon dioxide emission and economic growth by support vector machine. Renewable and Sustainable Energy Reviews, 64, 466-476.

Mohammed, S., Gill, A.R., Ghosal, K., Al-Dalahmeh, M., Alsafadi, K., Szabó, S., Oláh, J., Alkerdi, A., Ocwa, A., Harsanyi, E. (2024), Assessment of the environmental kuznets curve within EU-27: Steps toward environmental sustainability (1990-2019). Environmental Science and Ecotechnology, 18, 100312.

Mohammed, A., Li, Z., Arowolo, A.O., Su, H., Deng, X., Najmuddin, O., Zhang, Y. (2019), Driving factors of CO_2 emissions and nexus with economic growth, development and human health in the top ten emitting Countries. Resources Conservation and Recycling, 148, 157-169.

Namahoro, J.P., Wu, Q., Zhou, N., Xue, S. (2021), Impact of energy intensity, renewable energy, and economic growth on CO_2 emissions: Evidence from Africa across regions and income levels. Renewable and Sustainable Energy Reviews, 147, 111233.

Nathaniel, S.P., and Adeleye, N. (2021), Environmental preservation amidst carbon emissions, energy consumption, and urbanization in selected African countries: Implication for sustainability. Journal of Cleaner Production, 285, 125409.

Nurgazina, Z., Guo, Q., Ali, U., Kartal, M.T., Ullah, A., Khan, Z.A. (2022), Retesting the influences on CO_2 emissions in China: Evidence from dynamic ARDL approach. Frontiers in Environmental Science, 10, 868740.

Orji, U., Ukwandu, E. (2024), Machine learning for an explainable cost prediction of medical insurance. Machine Learning with Applications, 15, 100516.

Osobajo, O.A., Otitoju, A., Otitoju, M.A., Oke, A. (2020), The impact of energy consumption and economic growth on carbon dioxide emissions. Sustainability, 12(19), 7965.

Poumanyvong, P., Kaneko, S. (2010), Does urbanization lead to less energy use and lower CO_2 emissions? A cross-country analysis. Ecological Economics, 70(2), 434-444.

Prayudani, S., Hizriadi, A., Lase, Y.Y., Fatmi, Y. (2019), Analysis accuracy of forecasting measurement technique on random K-nearest neighbor (RKNN) using MAPE and MSE. Journal of Physics Conference Series, 1361(1), 012089.

Rahman, M.M., Nepal, R., Alam, K. (2021), Impacts of human capital, exports, economic growth and energy consumption on CO_2 emissions of a cross-sectionally dependent panel: Evidence from the newly industrialized countries (NICs). Environmental Science and Policy, 121, 24-36.

Raihan, A., Ibrahim, S., Muhtasim, D.A. (2023), Dynamic impacts of economic growth, energy use, tourism, and agricultural productivity on carbon dioxide emissions in Egypt. World Development Sustainability, 2, 100059.

Rehman, A., Alam, M.M., Ozturk, I., Alvarado, R., Murshed, M., Işık, C., Ma, H. (2023), Globalization and renewable energy use: How are they contributing to upsurge the CO_2 emissions? A global perspective. Environmental Science and Pollution Research, 30(4), 9699-9712.

Rehman, E., Rehman, S. (2022), Modeling the nexus between carbon emissions, urbanization, population growth, energy consumption, and economic development in Asia: Evidence from grey relational analysis. Energy Reports, 8, 5430-5442.

Riti, J.S., Song, D., Shu, Y., Kamah, M. (2017), Decoupling CO_2 emission and economic growth in China: Is there consistency in estimation results in analyzing environmental Kuznets curve? Journal of Cleaner Production, 166, 1448-1461.

Sarkodie, S.A., Ozturk, I. (2020), Investigating the environmental Kuznets curve hypothesis in Kenya: A multivariate analysis. Renewable and Sustainable Energy Reviews, 117, 109481.

Selmey, M.G., Elamer, A.A. (2023), Economic policy uncertainty, renewable energy and environmental degradation: Evidence from Egypt. Environmental Science and Pollution Research, 30(20), 58603-58617.

Selmey, M.G.S.G., Elkhodary, Y.F.Y., Elsayed, E.F.E.M. (2024), Economic policy uncertainty, energy consumption, trade openness and CO_2 emissions: Evidence from BRICS Countries. International Journal of Energy Economics and Policy, 14(6), 554-565.

Shaheen, A., Sheng, J., Arshad, S., Muhammad, H., Salam, S. (2025), Forecasting the determinants of environmental degradation: A gray modeling approach. Energy Sources Part A Recovery Utilization and Environmental Effects, 47(1), 1084-1104.

Shuai, C., Chen, X., Wu, Y., Tan, Y., Zhang, Y., Shen, L. (2018), Identifying the key impact factors of carbon emission in China: Results from a largely expanded pool of potential impact factors. Journal of Cleaner Production, 175, 612-623.

Stojić, A., Stanić, N., Vuković, G., Stanišić, S., Perišić, M., Šoštarić, A., Lazić, L. (2019), Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition. Science of The Total Environment, 653, 140-147.

Sulaiman, C., Abdul-Rahim, A.S. (2018), Population growth and CO_2 emission in Nigeria: A recursive ARDL approach. Sage Open, 8(2), 1-14.

Tenorio, J., Pérez, W. (2023), GDP Nowcasting with Machine Learning and Unstructured data to Peru, Working Papers No 197.

Teoh, E.Z., Yau, W.C., Ong, T.S., Connie, T. (2023), Explainable housing price prediction with determinant analysis. International Journal of Housing Markets and Analysis, 16(5), 1021-1045.

Tiffin, M.A. (2016), Seeing in the Dark: A Machine-Learning Approach to Nowcasting in Lebanon. United States: International Monetary Fund.

Tsimisaraka, R.S.M., Xiang, L., Andrianarivo, A.R.N.A., Josoa, E.Z., Khan, N., Hanif, M.S., Khurshid, A., Limongi, R. (2023), Impact of financial inclusion, globalization, renewable energy, ICT, and economic growth on CO_2 emission in OBOR Countries. Sustainability, 15(8), 6534.

United Nations Environment Programme (UNEP). (2022), "Emissions Gap Report 2022: The Closing Window - Climate Crisis Calls for Rapid Transformation of Societies. Available from: https://www.unep/org/resources/emissions-gap-report-2022

United Nations Environment Programme. (2023), Emissions Gap Report 2023: Broken Record - Temperatures Hit New Highs, Yet World Fails to Cut Emissions. Available form: https://www.unep.org/resources/emissions-gap-report-2023v

Wang, G., Chen, X., Zhang, Z., Niu, C. (2015), Influencing factors of energy-related CO_2 emissions in China: A decomposition analysis. Sustainability, 7(10), 14408-14426.

Wang, S., Li, Q., Fang, C., Zhou, C. (2016), The relationship between economic growth, energy consumption, and CO_2 emissions: Empirical evidence from China. Science of the Total Environment, 542, 360-371.

Wang, S., Liu, X., Zhou, C., Hu, J., Ou, J. (2017), Examining the impacts of socioeconomic factors, Urban form, and transportation networks on CO_2 emissions in China's megacities. Applied Energy, 185, 189-200.

Wang, Y., Kang, Y., Wang, J., Xu, L. (2017), Panel estimation for the impacts of population-related factors on CO_2 emissions: A regional analysis in China. Ecological Indicators, 78, 322-330.

World Population Review. (2024), CO_2 emissions by country. Available from: https://ww.worldpopulationreview.com

Wu, R., Zhang, J., Bao, Y., Lai, Q., Tong, S., and Song, Y. (2016), Decomposing the influencing factors of industrial sector carbon dioxide emissions in Inner Mongolia based on the LMDI method. Sustainability, 8(7), 661.

Yoon, J. (2021), Forecasting of real GDP growth using machine learning models: Gradient boosting and random forest approach. Computational Economics, 57(1), 247-265.

Yu, W., Xia, L., Cao, Q. (2024), A machine learning algorithm to explore the drivers of carbon emissions in Chinese cities. Scientific Reports, 14(1), 23609.

Yuaningsih, L., Febrianti, R.A.M., Kamran, H.W. (2021), Climate change and energy consumption patterns in Thailand: Time trends during 1988-2013. International Journal of Energy Economics and Policy, 11(1), 571-576.

Zhang, J., Zhang, Y. (2021), Tourism, economic growth, energy consumption, and CO_2 emissions in China. Tourism Economics, 27(5), 1060-1080.

Zhong, W., Zhai, D., Xu, W., Gong, W., Yan, C., Zhang, Y., Qi, L. (2024), Accurate and efficient daily carbon emission forecasting based on improved ARIMA. Applied Energy, 376, 124232.