FMDB *Transactions on Sustainable Energy Sequence*

Quantifying the Influence of Population on Carbon Emissions: A Comparative Study of Developed versus Developing Countries Using Machine Learning

Ify Nwabuokei[1,*], Mohammed Ghouse[2], Wenjun Chen[3], Elvis Tony Raphael Kollannur[4], Hai Li[5], Rejwan Bin Sulaiman[6]

[1,2,3,4,5,6]Department of Computer and Information Science, Northumbria University, London, United Kingdom.
inwabuokei8@gmail.com[1], mdghouse1293@gmail.com[2],wenjun.chenp@gmail.com[3], mail@elvistony.dev[4], haitin.lee@gmail.com[5], rejwan.sulaiman@northumbria.ac.uk[6].

Abstract: This research paper delves into the intricate relationship between population size and carbon emissions, focusing on a comparative analysis between developed and developing nations. The significance of this topic lies in its critical role in environmental studies and climate change mitigation. The study's purpose is to quantitatively analyze the impact of population on carbon emissions and provide insights for targeted climate change mitigation strategies, particularly contrasting the dynamics in developed versus developing countries. Employing advanced machine learning techniques, such as random forest, linear regression, and Xgboost, the research aims to quantify the influence of population on carbon emissions. The analysis reveals distinct emission trajectories for developed and developing nations, with the United States and China serving as primary examples. The findings underscore the significance of population size in shaping carbon emissions and emphasize the need for tailored policy interventions that consider each nation's demographic and industrial landscape. The research offers a comprehensive understanding of the correlation between population dynamics and carbon emissions, providing a foundation for future policy-making and interventions.

Keywords: Carbon Emissions; Population Size; Developed Countries; Developing Countries; Machine Learning; Comparative Analysis; Climate Change Mitigation; Policy Interventions; Global Warming; Quantum Dots (QDs).

Received on: 15/05/2023, **Revised on:** 29/07/2023, **Accepted on:** 17/09/2023, **Published on:** 23/12/2023

Cite as: I. Nwabuokei, M. Ghouse, W. Chen, E. T. Raphael Kollannur, H. Li, and R. Bin Sulaiman, "Quantifying the Influence of Population on Carbon Emissions: A Comparative Study of Developed versus Developing Countries Using Machine Learning," *FMDB Transactions on Sustainable Energy Sequence*, vol. 1, no. 2, pp. 71–82, 2023.

1. Introduction

The global energy landscape is undergoing a monumental shift, driven by an urgent need to address mounting environmental concerns. Carbon emissions and their relationship to climate change have attracted a lot of attention recently. The ongoing increase in carbon dioxide emissions mostly causes climate change. Of the environmental, social, and economic risks the world has encountered in the last century, global warming is one of the largest and most likely challenging. Sea level rise, floods, droughts, and other unfavourable environmental effects are caused by increasing atmospheric carbon dioxide emissions, which is one of the main causes of global warming [19]. There are two possible effects of rising CO2 emissions on human health: direct and indirect. When inhaled in large quantities, it directly impacts and can lead to significant illnesses such as dyspnea, delirium, blindness, and dizziness [3]. The indirect manifestation of global issues, including acid rain, climate change, and global warming, is high CO2 emissions. These pollutants come in a variety of highly dangerous forms for both the environment

*Corresponding author.

and people. According to The Lancet [24], serious illnesses brought on by air pollution claim the lives of almost 6.5 million people globally each year. Moreover, the number of deaths from AIDS/HIV, TB, and auto accidents put together is less than this one. Approximately 75 percent of Indians are exposed to air pollution, which is significantly more than the government-mandated minimal limit [27].

Recent studies have raised significant questions about the effects of population growth, industrialization, energy consumption, and renewable energy on the environment [5]; [17]; [18]; [28]. CO2 emissions are increased when fossil fuels are burned for transportation, industry, energy, and commercial buildings [25]; [30]; [4]. According to estimates, coal, gas, oil, and other fossil fuels accounted for more than 84 percent of the energy produced in 2019; coal and oil were the primary fuels used to create electricity [31]. Environmental degradation has become a global concern due to the expansion of the industrial sector in recent decades [6]. Environmental degradation is also a result of human activity, including burning fossil fuels and deforestation [21]. As a result of human activity, the atmospheric concentrations of CO2 have increased to levels not seen in previous centuries. However, renewable energy comes from natural resources that can help fight global warming. One of the most dependable energy sources is renewable energy, which can reduce pollution. This study explores the relationship between energy consumption, population, industrialization, GDP, FD, and renewable energy and the environmental quality of growing economies, taking into account the features mentioned earlier [29]. The relationship between population size and carbon emissions has received substantial attention within environmental studies, particularly in the context of climate change and population geography. Understanding how the population affects carbon emissions is critical for the effective development of solutions meant to mitigate the negative effects of climate change. Even though this relationship has been partly studied in the past, a comparative study that focuses primarily on industrialized and emerging nations is still necessary [22].

By better comprehending the varying impacts that population size has on carbon emissions, a deeper understanding can be gained regarding potential areas where targeted interventions may be implemented. Analyzing population size's impact on carbon emissions in developed and developing countries quantitatively is the research challenge this study attempts to solve. The goal is to find any notable variations between each group and to quantify the degree to which population size influences carbon emissions in each. The urgency with which climate change and its implications must be addressed is what spurred this study's motivation. We can find efficient tactics and policy interventions to lessen the effects of human activity on the environment by looking at the relationship between population size and carbon emissions in developed and developing nations. This study's main objective is to measure, using machine learning techniques, how much population has an impact on carbon emissions in developed and emerging nations. To ad- address the research problem and achieve the objectives outlined above, the following research questions will guide this study:

- How does population size influence carbon emissions in developed countries compared to developing countries?
- Is there a significant correlation between population growth rate and carbon emissions in developed and developing countries?

Through a comparison of how the population affects carbon emissions in developed and developing countries, this study adds to the collection of existing knowledge. By utilizing machine learning techniques, the study will generate quantitative insights and identify significant correlations between population size, growth rate, and carbon emissions. These findings have the potential to inform policymakers and researchers in formulating effective strategies to mitigate the adverse effects of carbon emissions.

2. Review of Literature

The concept of utilizing quantum dots (QDs) in solar cells has been a subject of intense research and exploration for nearly two decades, with significant strides made in understanding and harnessing their potential [8]. It was in 2004 that Schaller and Klimov laid the foundation for this promising technology by demonstrating the phenomenon of multiple exciton generation in quantum dots [1]. This ground-breaking discovery hinted at a theoretical efficiency limit that could surpass the capabilities of conventional solar cells, opening up exciting possibilities for renewable energy. Previous research on the relationship between population and carbon emissions has yielded mixed results. Some studies have found a positive correlation between population size and carbon emissions, suggesting that larger populations lead to increased emissions. For instance, the authors in [34] utilized the input-output model and combined regression analysis, input-output technology, and structural decomposition analysis to analyze the impact of various factors on indirect carbon emissions from households in China. The findings indicated that urban employment, rapid urbanization, and the two-child policy amplified the influence of the population variable on increasing indirect carbon emissions. The study also highlighted that growing household incomes and improved living standards pose a significant threat to climate change responses in developing countries [35].

In this study [12], researchers investigated the relationship between population and economic growth and carbon dioxide (CO2) emissions in light of the increasing urgency of climate change. The dataset utilized in the analysis covered a 30-year timeframe.

It is comprised of nine countries: developed (Japan, USA, UK), fast-developing (India, China, Brazil), and slow-developing (Congo, Bangladesh, Malaysia). Employing a stochastic model, the investigation employed CO2 emissions as the dependent variable and took into account independent variables, including population, affluence (measured by GDP per capita), and technology (evaluated by resources used per unit of production). The findings brought to light the significant impact of population and economic growth on the notable upswing in global CO2 emissions observed during the preceding two decades. Using multiple linear regression techniques, with population growth and electricity, [36] was able to forecast and analyze carbon emissions. Linear regression was used to find the relationship between Co2, which was categorized as a dependent variable and other factors, which were categorized as independent variables. To further understand their correlation, the dependent variable was provided as a label and the independent variable as a feature. The adopted strategy resulted in 98 percent accuracy in estimating carbon emissions based on population increase and electricity absorption from a 70:30 split [13].

This study [8] examined the influence of income, energy consumption, and population growth on CO2 emissions in India, Indonesia, China, and Brazil for the years 1970 to 2012. The study employed the Autoregressive Distributed Lag (ARDL) bounds test approach, considering both linear and non-linear relationships in the short and long run. Results indicated a significant positive correlation between CO2 emissions and income and energy consumption across all four countries. However, the relationship between CO2 emissions and population growth was statistically significant for India and Brazil but insignificant for China and Indonesia in both short and long-term analyses. The environmental Kuznets Curve (EKC) hypothesis was also tested, which suggested that in Brazil, China, and Indonesia, CO2 emissions decrease over time with increasing income, implying these countries need not adopt policies that negatively impact income to reduce emissions. GDP and CO2 emissions in Algeria were examined between 1970 and 2010 [11]. The findings indicated that there is a significant correlation between the two variables.

In Pakistan, the association between CO2 emissions, energy consumption, and economic advancement was investigated in both bivariate and multivariate models for the years 1971–2019 using the autoregressive distributed lag model (ARDL) and Johansen cointegration [2]; the result showed that rising economic growth and energy consumption have a positive and considerable effect on CO2 emissions. To effectively address carbon emissions control, examined the influence of population and economic growth on carbon dioxide (CO2) emissions in developing countries. In addition to total population, factors such as urbanization percentage and non-dependent population were also significant determinants. Furthermore, the total energy-driven gross domestic product (GDP) and the proportion of GDP generated by manufacturing industries were evaluated to ascertain their respective impacts on carbon emissions. National data from Taiwan spanning the years 1990 to 2014 were utilized to investigate the relationships between these determinants. The Stochastic Impacts by Regression on Population, Affluence, and Technology (STIRPAT) analytical tool was employed to assess the potential impact of these factors on global warming, as indicated by total atmospheric carbon emissions. Seven STIRPAT models are proposed and statistically tested for significance, resulting in the identification of two models that predicted the impact of population and economic growth on carbon emissions in Taiwan by the year 2025.

In this study, [9] analysed the influence of urbanization, economic growth, and population size on residential carbon emissions in the South Asian Association for Regional Cooperation (SAARC) member nations from 1994 to 2013. The empirical findings revealed a U-shaped relationship between urbanisation and residential carbon emissions, indicating population size and per capita GDP play significant roles as drivers of high carbon emissions in the SAARC countries. This contributes to the significance of population size as a key determinant of high carbon emissions in the SAARC member nations when examining carbon emissions trends. However, this literature does not directly address the influence of population size on carbon emissions in developed and developing countries, nor does it investigate the correlation between population growth rate and carbon emissions. By considering the connections and areas of agreement between the existing findings, along with the potential trends identified, this paper aims to extend the analysis to developed and developing countries, exploring the influence of population size on carbon emissions to address the research gaps by conducting a comprehensive understanding of the correlation between population dynamics and carbon emissions.

To address the detrimental effects of CO2 emissions on ecosystems in India, [23] integrated multiple models that facilitate this prediction task. The authors proposed three statistical models, including autoregressive-integrated moving average (ARIMA), seasonal autoregressive-integrated moving average with exogenous factors (SARIMAX) and the Holt-Winters model, two machine learning models, namely linear regression(LR) and random forest(RF), as well as a deep learning-based long short-term memory (LSTM) model. The objective was to identify the impact of CO2 emissions and predict the trend for the next decade using univariate time-series data from 1980 to 2019. Through a comparative analysis of six models evaluated in nine metrics: mean squared error (MSE), root mean squared error (RMSE), mean squared log error (MSLE), mean absolute error (MAE), mean absolute percentage error (MAPE), median absolute error (MedAE), R2 score error, maximum error, and explained variance score error (EVS). The results demonstrate that the deep learning-based LSTM model is the best suitable model for predicting CO2 emissions, with a MAPE value of 3.101 percent, RMSE value of 60.635, and MedAE value of 28.898.

This is in line with the study done by [23];[7], which aimed to identify the factors that influence carbon emissions and the patterns in China and India by applying the same deep learning method (i.e. LSTM). These two nations are the foremost users of coal and have the highest population figures globally. The authors sought to evaluate the influence of financial development (FD), energy consumption, population, gross domestic product (GDP), and renewable energy on CO2 emissions, considering the significant industrial and demographic expansion in both countries. The LSTM approach was employed to analyse the relative influence of several drivers on CO2 emissions in a dataset spanning from 1990 to 2014. The findings concluded that China would see a deceleration in CO2 emissions by 2022, whilst India will not achieve this until 2023.

Moreover, this also revealed that energy consumption exerted the most significant influence, while renewable energy had the least substantial effect on CO2 emissions on both sides. These two papers specifically examined the prediction of CO2 emissions and the relevant variables in emerging nations, which are closely connected to the topic in this study that compares developed and developing countries. The methodology and conclusions presented may be utilised to compare or contrast with analogous models in this field, facilitating comprehension of how various elements, such as technical progress or economic policies, impact the precision of emission forecasts in diverse circumstances, yielding a more nuanced comprehension of the correlation between population dynamics and carbon emissions in diverse economic environments.

With the goal of predicting the increasing levels of carbon dioxide (CO2) emissions regularly in China, [26] chose the most accurate model to forecast future CO2 emissions based on daily time-series data from January 1st, 2020, to September 30th, 2022. There were six proposed models, consisting of three statistical models: seasonal autoregressive integrated moving average with exogenous factors (SARIMAX), autoregressive integrated moving average (ARIMA), and grey prediction (GM(1,1)); three machine learning models: artificial neural network (ANN), RF, and LSTM. The performance of these techniques was assessed based on five criteria: MSE, RMSE, MAPE, Mean Absolute Error (MAE), and R2 score. The result indicated that the three machine learning models surpassed the performance of the three statistics models in all five categories. In particular, LSTM stood out for its outstanding ability to capture and predict intricate emission patterns, making it the optimal choice for predicting CO2 emission on a daily basis. The remarkable accuracy was achieved by an MSE value of 3.5179, a RMSE result of 0.0187, a MAE score of 0.0140, a value of MAPE at 0.1482, and a high coefficient of determination (R2) value of 0.9844. The methodologies used in this paper to forecast CO2 emissions displayed a diverse array of sophisticated approaches that can be applied, providing an in-depth analysis of emission patterns in connection to population dynamics. Additionally, China, as a prominent emerging nation with substantial CO2 emissions and swift demographic shifts, served as a pertinent case study for this similar scenario. They provided valuable insights into emission trends in China, which can be used to compare with data from developed countries to analyse the influence on emission patterns.

In parallel to the work conducted by [26], [1] introduced three models to forecast the carbon emissions pattern in a dataset obtained from nine South Asian nations from the years 2001 to 2020 due to the exceptional rate of industrialization and urbanisation in China, this has become the main driver of climate change. These three models were the Bidirectional Long Short-Term Memory Neural Network (BiLSTM), LSTM, and Gated Recurrent Unit (GRU). These models were trained using data collected from nine South Asian countries, including Sri Lanka, Afghanistan, Nepal, India, Bhutan, Pakistan, China, Bangladesh, and Maldives. Moreover, an analysis of the collective impact of scientific and technical advancements, industrial structures, and energy structure elements on CO2 emissions from 2022 to 2030 was also conducted. Upon evaluating the model performance of these three techniques, the Bi-LSTM model exhibited superior performance with lower values for MAE, MSE, and MAPE in comparison to the other models. The emphasis on South Asian countries was directly related to the interest of this study in emerging economies. These areas are undergoing substantial population expansion, which is strongly linked to heightened carbon emissions as a result of fast urbanisation. The utilisation of modern predictive modelling approaches, along with a thorough investigation of several influential elements, provides important insights for the upcoming study.

In a similar study conducted in mainland China, [20] utilised the Autoregressive Integrated Moving Average (ARIMA) statistical model, as well as machine learning forecasting techniques, including Recurrent Neural Networks (RNNs), LSTM, and LSTM with Clustering and Multivariate Time- Series, to estimate the level of carbon emissions from 2030 to 2060. In order to examine the similarities in trends among various regions in this developing country, the authors focused on the analysis in ten specific administrative regions: Jiangsu Province, Shandong Province, Hebei Province, Henan Province, the Inner Mongolia Autonomous Region, Guangdong Province, Chongqing City, Shanghai City, Beijing City and Tianjin City, to examine the corresponding GDP and population data. During this procedure, the mean average percentage error was computed, yielding a minimum value of

2.56 percent for the RNN model. This suggests that the RNN model is more dependable than the other methods. To conclude, the methodology, regional focus, and insights presented in this work made a substantial contribution to the study investigating the correlation between population density, economic growth, and carbon emissions. The utilisation of sophisticated forecasting methods and meticulous assessment of specific geographical areas provided a complete structure for analysing these connections in both developed and emerging nations. In this paper, [15] conducted two tasks to analyse the rising circumstances of carbon dioxide emissions from the energy sector in Bangladesh. This research focused on the period from 1972 to 2019. The authors aim to determine the correlations between CO2 emissions, gross domestic product (GDP) and electrical energy consumption by utilising the fully modified ordinary least squares (FMOLS) technique in the first part. The Philips-Perron (PP) and augmented Dickey-Fuller (ADF) tests are employed to perform a unit root test in order to assess the stationarity of the variables. The estimated results indicate the presence of a long-term and cointegration relationship between energy consumption GDP and CO emissions in this developing nation. Specifically, energy consumption has a positive effect on CO emissions, whereas GDP has an adverse impact. In the second phase, [16] devised five machine learning algorithms, namely convolution neural network (CNN), CNN long short-term memory (CNN–LSTM), LSTM, dense neural network (DNN), and multivariate polynomial regression to predict CO2 emissions. The evaluation metrics employed for analysing and comparing the performances of the prediction models encompass RMSE, MAE, and MAPE. The results indicate that the DNN strategy yields the most favourable outcomes, with the lowest MAPE of 3.678, in comparison to the alternative methods. The initial segment of this paper, which examines the correlations among CO2 emissions, GDP and electrical energy consumption, corresponds to the emphasis of this study on carbon emissions in both industrialised and developing countries. An analysis of the relationship between these factors and CO2 emissions in Bangladesh might offer valuable insights into comparable trends in other countries, regardless of the level of development.

In order to analyse the factors that affect CO2 emissions in five African countries (Algeria, Egypt, Libya, Nigeria, and South Africa), [14] put forward two machine learning models: EMD-GCN (Empirical Mode Decomposition Graph Convolutional Network) and Empirical Mode Decomposition (EMD). These algorithms were used to forecast CO2 emissions for the next decade, using time series data from 1980 to 2021. The results of a comparison study of these techniques in five countries show that EMD-GCN performs better than the other model, according to the four performance metrics: MAE, MASE, MAPE, and R2 score. The EMD-GCN model leverages the advantages of both EMD and GCN to effectively capture the fundamental patterns and fluctuations in the data, which can be advantageous for modelling intricate and non-linear connections. Furthermore, the research results also open up new research projects that can improve the accuracy of CO2 emission predictions by adding extra outside factors like technological progress, population growth, and different policy measures that could have either positive or negative effects on CO2 emissions. The novel methodology used in this paper provided a powerful means of capturing complicated and non-linear connections within environmental data. This tool can be particularly useful for comprehending the complex dynamics between population increase and CO2 emissions. The emphasis on African nations also presents a viewpoint from an alternative group of emerging countries, in contrast to this study. Gaining comprehension of the functioning of these models within the framework of African states might provide valuable insights into the applicability of such methodologies across diverse emerging areas.

In order to forecast the CO2 emissions of vehicles in Canada, [10] proposed eight machine learning methods to predict the ranking of CO2, utilising the fuel consumption rating dataset of Canada from the past five-year period (2017-2021). The models include Decision Tree, Naive Bayes, Random Forest, Logistic Regression, Support Vector Machine (Linear), Support Vector Machine (Polynomial), K-Nearest Neighbour, and Stochastic Gradient Descent. The performance of these techniques was assessed based on metrics such as accuracy, recall, precision, and F1 score. The findings indicate that the Random Forest method achieved the maximum accuracy of 96 percent among all the models, while the Naive Bayes classifier had the lowest precision of 73 percent. In contrast to the specific emphasis on automobile emissions in this paper, which delivers a comprehensive analysis of a substantial source of CO2 emissions, this study takes a broader approach by investigating the relationship between population density and carbon emissions across different nations. Nevertheless, the methodology and conclusions presented can contribute to a specific aspect of this project, which examines the influence of transport on total carbon emissions.

In [32], the authors suggested that population density has a moderate impact on the air quality index (AQI), while CO2 emissions exhibit a slight positive correlation. This study encompassed 32 countries and aimed to investigate the influence of population density, CO2 emissions, and tree coverage on air quality. The correlation analysis indicated that population density has a moderate effect on AQI, with a correlation coefficient of 0.43. Conversely, tree coverage displayed a negative correlation of -0.43 with AQI, suggesting that areas with a higher tree density tend to have less polluted air [33]. Additionally, CO2 emissions demonstrated a slight positive correlation of 0.21 with AQI. Although this study establishes the connection between population density and air quality, further research is needed to explore the relationship between population growth rate and carbon emission between developed and developing countries. In this research, the dataset will be classified into two main categories to estimate and identify potential factors associated with CO2 emissions and population growth.

3. Methodology

Preprocessing: Data preprocessing is a stage that is a part of data preparation where the raw data is processed to be more suitable for machine learning or other data-science applications.

Opening Dataset inappropriate Read mode: Upon opening the dataset in the default reading mode utf−8, the function read_CSV () threw an error specifying the presence of unknown characters. This is a characteristic of CSV files with a different encoding. The error was resolved using the ISO−8859−1 encoding.

Fixing Datatype mismatches: Pandas is a popular Python library that deals with tabular data. Pandas have an innate capability of assigning appropriate datatypes based on the data in the column. However, in some instances, this assignment may not be accurate, and hence, the developer must assign them manually to ensure no data is lost.

Handle missing values: Handling missing values is a crucial step in the preprocessing stage. Missing data can be dealt with differently based on the patterns they exhibit.

Random missing values can occur when the sensor fails to produce viable data and is set as null or zero. These missing values can often be filled by imputation or statistical values. It was seen that null values were present in the given dataset. Upon initial analysis, the missing values seem to follow a pattern where the absence of values in one column for a country increases the likeliness of other columns to also have null values. Patterns were found in the dataset where, for a specific country, columns have null values of a maximum of 271 rows. This also relates to the number of rows for each country, which is also 271 rows.

4. Exploratory Data Analysis

Exploratory data analysis can be defined as the process of investigating datasets and discovering characteristics that are present in the data. Performing EDA helps to explore the nature of the data and identify patterns within the dataset. Figure 1 visualizes the top 20 countries with the sum of their carbon emissions over the years from 2010 to 2020. The United States appears to have led with the highest CO_2-releasing nation consistently for the last decade. A contributing factor could include the fact that the United States has the third highest population at 338M. Figure 2 shows the bottom 20 countries that have contributed the least over the years from 2010 to 2020. Antarctica appears to have the least contribution to the carbon emission footprint. This can be assumed to be correlated to the human population present in the region. However, the data regarding the population is absent in the given dataset, and hence, this assumption cannot be confirmed.

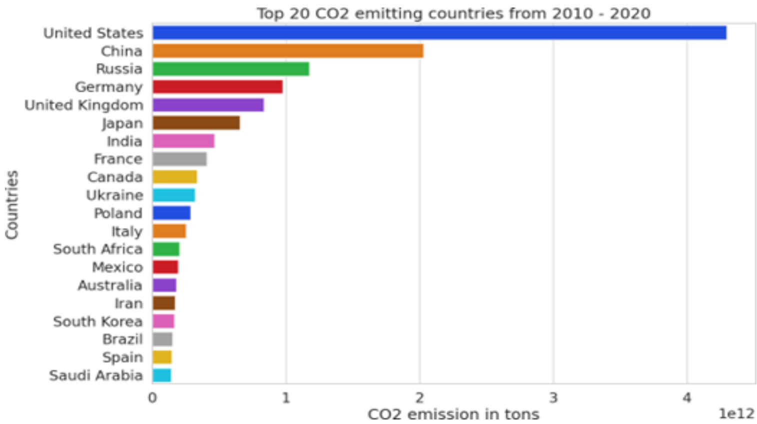

Figure 1: Top 20 Countries - 2010 – 2020

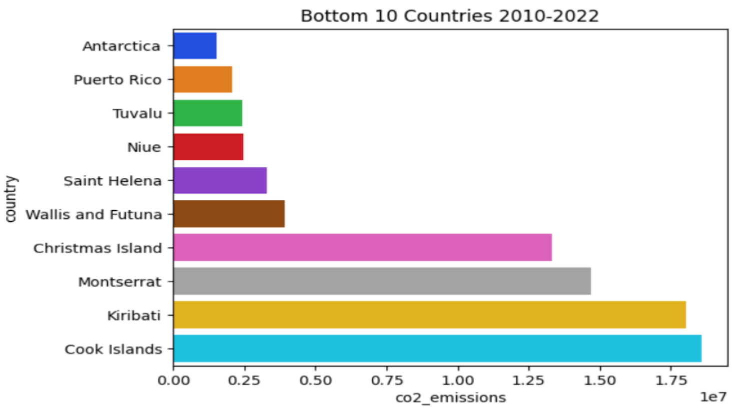

Figure 2: Bottom 20 Countries - 2010 - 2020

Figure 3 shows that CO2 significantly increased over the years. This steady increase in CO_2 emission has raised concerns about its impact on global climate change, contributing to the greenhouse effect and subsequent alterations in Earth's climate patterns. Comparing the op 5 CO_2 emitting countries in the years 1750, 1800, 1850, 1900, 1950, 2000, 2010, 2020. As shown in Figures 4 to 11 , the United Kingdom's carbon emissions have declined over time. The Climate Change Act, which the UK government enacted, appears to be the reason for this decrease in carbon emissions. In other countries, CO2 emissions steadily increase over time.

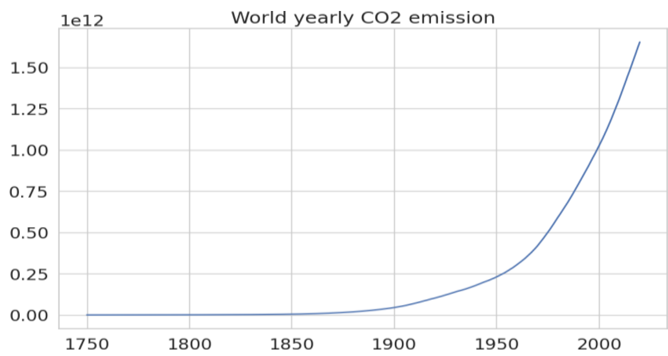

Figure 3: World yearly CO_2 emission

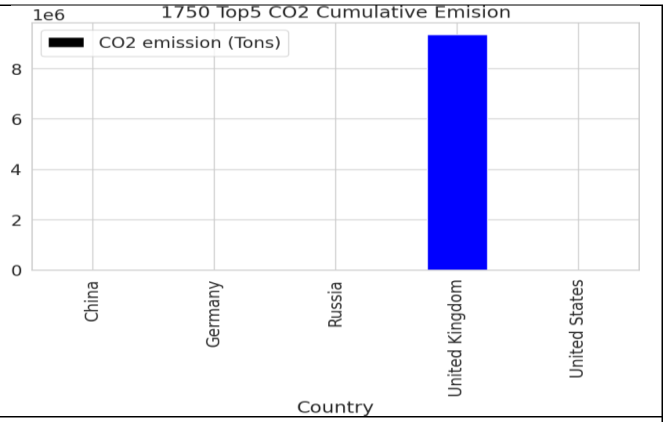

Figure 4: 1750 Top 5 Countries with cumulative CO_2 Emission

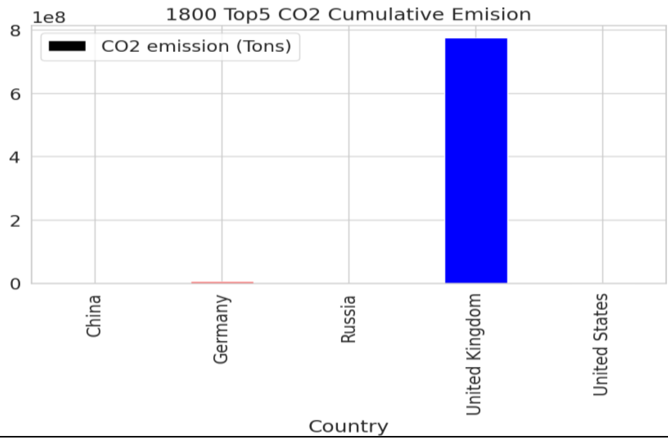

Figure 5: 1800 Top 5 Countries with cumulative CO_2 Emission

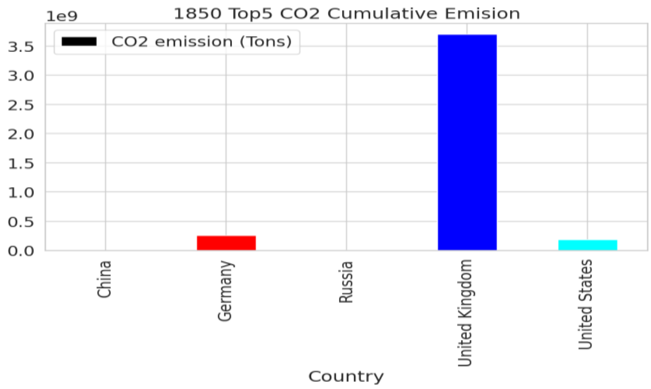

Figure 6: 1850 Top 5 Countries with cumulative CO2 Emission

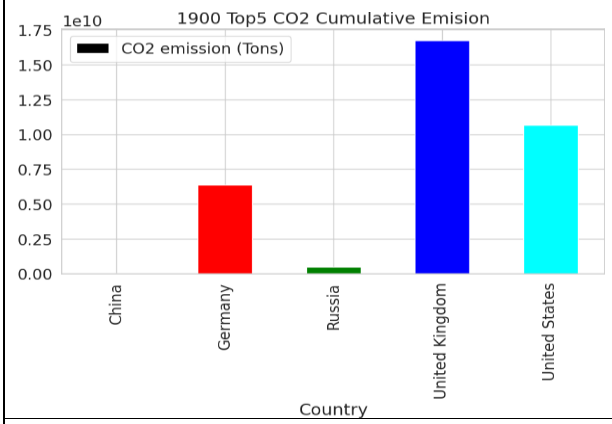

Figure 7: 1900 Top 5 Countries with cumulative CO_2 Emission

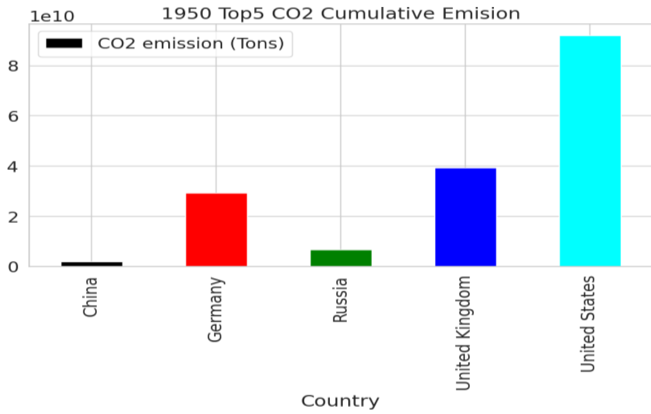

Figure 8: 1950 Top 5 Countries with Cumulative CO_2 Emission

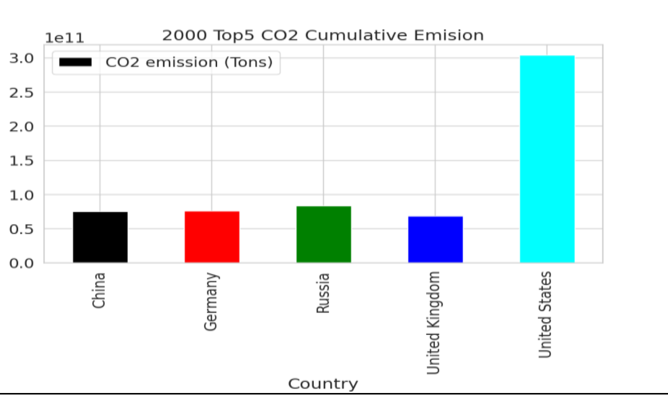

Figure 9: 2000 Top 5 Countries with cumulative CO_2 Emission

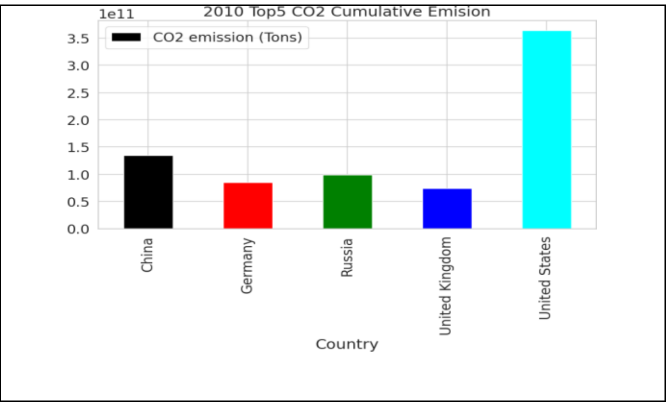

Figure 10: 2010 Top 5 Countries with cumulative CO_2 Emission

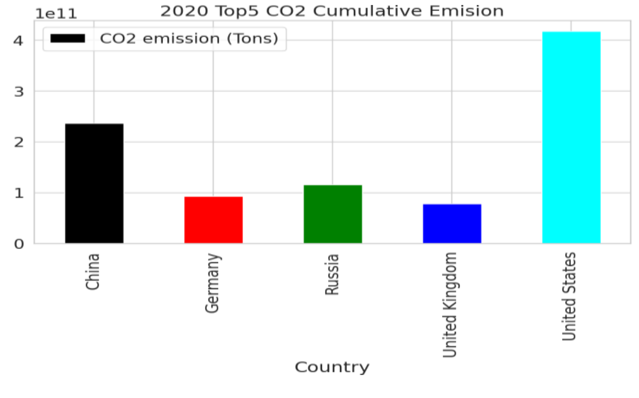

Figure 11: 2020 Top 5 Countries with cumulative CO_2 Emission

4.1. Algorithm Selection

Model Evaluation: Figure 4 highlights that we have used three different models: a) random forest, b) linear regression, and c) Xgboost to compare the performance, and random forest has performed well compared to the other two models.

Significant Correlation of Population: Is there a significant correlation between population growth rate and carbon emissions in developed and developing countries?

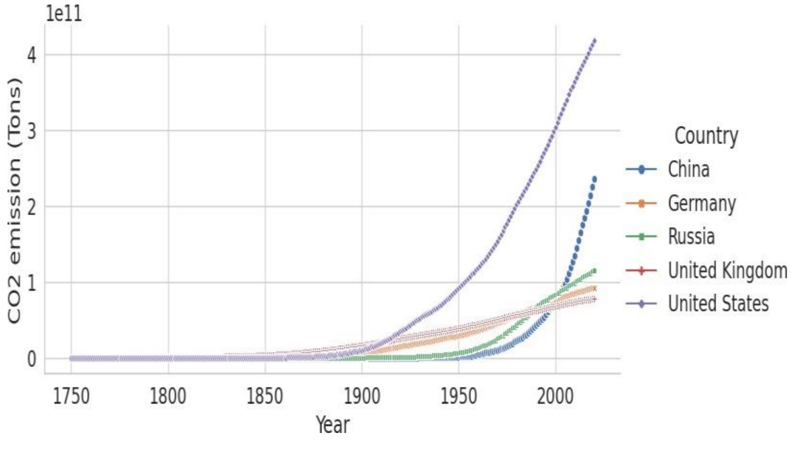

Figure 12: CO_2 Emissions of Top 5 countries over the years

Verdict: Population Size Influence: How does population size influence carbon emissions in developed countries compared to developing countries?

Considering that the United States is one of the developed countries, it can be seen that the carbon emissions from this nation are higher than those from China, and the rate of ascent is much lower than the stronger rate of China.

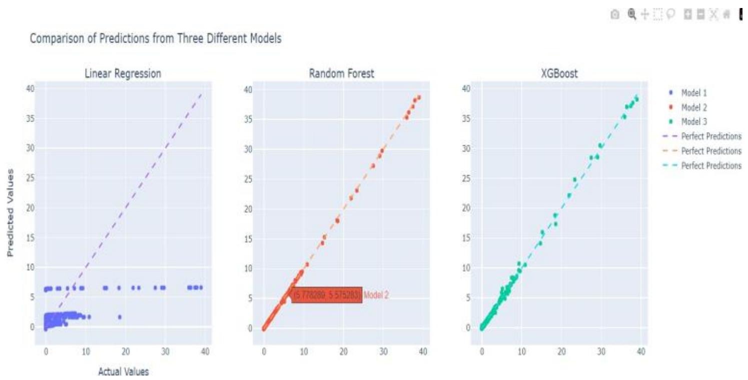

Figure 13. Comparison of three different model performance

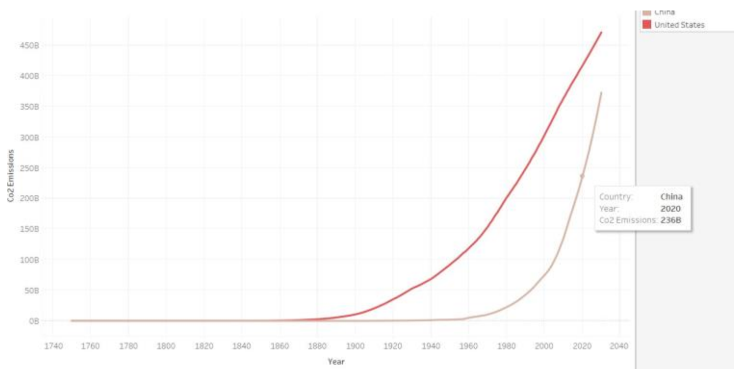

Figure 14: CO2 Emissions vs Year for China

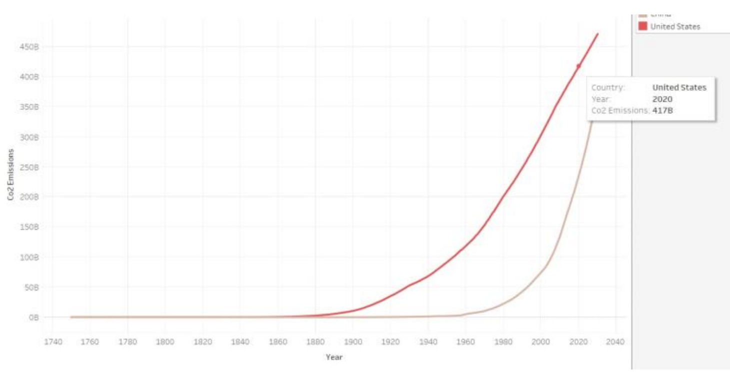

Figure 15: CO2 Emissions vs Year for the USA

It was noted that the population of developing countries contributed significantly to the nation's carbon emissions, as seen in Figure (). The United States, having a mere population of 92B and China, boasting of a population of 386B, can be correlated to their carbon footprint counterparts (Figures 12 to 15). It can also be found that developing countries may not be focusing their efforts on reducing the carbon emissions in their nation. With the increase in population and the industrial revolution in the previous century, these nations were affected by an inflated carbon footprint.

5. Conclusion

In this study, an exploration of the relationship between population size and carbon emissions was undertaken, with the aim of illuminating its implications for climate change mitigation. The study focused on conducting a comparative analysis between developed and developing nations, utilizing machine learning techniques to quantify the influence of population size on carbon emissions. This research was driven by the need for a comprehensive understanding of how the population impacts carbon emissions within the context of climate change. By examining both developed and developing countries, the objective was to quantify the extent of this influence and provide insights for targeted interventions. The analysis revealed that the United States, a developed nation, consistently maintained a high level of carbon emissions over the past decade, although with a slower rate of increase compared to China, a developing country. This highlights a notable difference in emission trajectories between developed and developing nations.

The top 20 emitting countries, led by the United States, showcased distinct patterns in carbon emissions. Conversely, the bottom 20 countries, exemplified by Antarctica, exhibited significantly lower emissions, likely due to their sparse populations. Machine learning models, particularly random forests, proved effective in discerning these patterns. This study emphasizes the role of population size in shaping carbon emissions, especially in the context of differentiating between developed and developing countries. The findings provide a foundation for targeted policy interventions to mitigate environmental impacts, emphasizing the need for strategic approaches tailored to each nation's demographic and industrial landscape. In conclusion, this study offers valuable insights into the relationship between population size and carbon emissions, highlighting disparities between developed and developing nations. By understanding and quantifying these dynamics, policymakers can tailor interventions to effectively mitigate the environmental consequences of population growth. Addressing these challenges is imperative for a sustainable and environmentally conscious future.

Acknowledgement: We extend our sincere gratitude to Dr. Rejwan Bin Sulaiman for his exceptional mentorship and support throughout this research work. The invaluable resources provided in the course module by Northumbria University played a pivotal role in the successful completion of this endeavour. We also wish to express our appreciation to our families for their unwavering support during the challenging phases of this research journey.

Data Availability Statement: The datasets utilized and analyzed during this study are openly available on Kaggle (https://www.kaggle.com/). We embrace the principles of transparency and open science, inviting fellow researchers to access, utilize, and build upon these datasets for further exploration.

Funding Statement: This research was self-funded, as all authors independently pursued this project out of personal interest following the completion of a module in our course of study facilitated by Dr Rejwan Bin Sulaiman at Northumbria University, London.

Conflicts of Interest Statement: All authors declare that there are no conflicts of interest regarding the publication of this paper. This project was conducted with integrity, ensuring independence from any external influences that could compromise the objectivity of the study or the reporting of its findings. Furthermore, all references from which the data were obtained are duly acknowledged within this publication.

Ethics and Consent Statement: This study adhered to the ethical standards and guidelines. All research protocols were thoroughly reviewed and approved by our supervisor, Dr. Rejwan Bin Sulaiman.

References

1. M. Aamir et al., "Predicting the environmental change of carbon emission patterns in south Asia: A deep learning approach using BiLSTM," Atmosphere (Basel), vol. 13, no. 12, p. 2011, 2022.
2. S. Aftab, A. Ahmed, A. A. Chandio, B. A. Korankye, A. Ali, and W. Fang, "Modeling the nexus between carbon emissions, energy consumption, and economic progress in Pakistan: Evidence from cointegration and causality analysis," Energy Rep., vol. 7, pp. 4642–4658, 2021.
3. Ü. Ağbulut, "Forecasting of transportation-related energy demand and CO2 emissions in Turkey with different machine learning algorithms," Sustain. Prod. Consum., vol. 29, pp. 141–157, 2022.
4. Ü. Ağbulut, İ. Ceylan, A. E. Gürel, and A. Ergün, "The history of greenhouse gas emissions and relation with the nuclear energy policy for Turkey," Int. J. Ambient Energy, vol. 42, no. 12, pp. 1447–1455, 2021.

5. F. Ahmad, M. U. Draz, A. A. Chandio, L. Su, M. Ahmad, and M. Irfan, “Investigating the myth of smokeless industry: environ- mental sustainability in the ASEAN countries and the role of service sector and renewable energy,” Environmental Science and Pollution Research, vol. 28, pp. 55344–55361, 2021.
6. M. Ahmad, Z. Khan, M. K. Anser, and G. Jabeen, “Do rural-urban migration and industrial agglomeration mitigate the environmental degradation across China’s regional development levels?,” Sustainable Production and Consumption, vol. 27, no. 7, pp. 679–697, 2021, doi: https://doi.org/10.1016/j.spc.2021.01.038.
7. M. Ahmed, C. Shuai, and M. Ahmed, “Influencing factors of carbon emissions and their trends in China and India: a machine learning method,” Environmental Science and Pollution Research, vol.29, no.32, pp. 48424–48437, 2022, doi: https://doi.org/10.1007/s11356-022-18711-3.
8. Md. M. Alam, Md. W. Murad, A. H. Md. Noman, and I. Ozturk, “Relationships among carbon emissions, economic growth, energy consumption and population growth: Testing Environmental Kuznets Curve hypothesis for Brazil, China, India and Indonesia,” Ecological Indicators, vol. 70, no.11, pp. 466–479, 2016, doi: https://doi.org/10.1016/j.ecolind.2016.06.043.
9. M. K. Anser, M. Alharthi, B. Aziz, and S. Wasim, “Impact of urbanization, economic growth, and population size on residential carbon emissions in the SAARC countries,” Clean Technologies and Environmental Policy, vol. 22, no.3, pp. 923-936, 2020, doi: https://doi.org/10.1007/s10098-020-01833-y.
10. S. D. Bappon, A. Dey, S. M. Sabuj, and A. Das, “Toward a Machine Learning Approach to Predict the CO2 Rating of Fuel-Consuming Vehicles in Canada,” In 2022 25th International Conference on Computer and Information Technology (ICCIT) (2022), IEEE, pp. 384–389, 2022.
11. M. Bouznit and M. del P. Pablo-Romero, “CO2 emission and economic growth in Algeria,” Energy Policy, vol. 96, no.9, pp. 93–104, 2016, doi: https://doi.org/10.1016/j.enpol.2016.05.036.
12. V. Chotia and P. Pankaj, “Impact of population and economic growth on carbon emissions of developed and developing countries,” Int. J. Green Econ., vol. 13, no. 3/4, p. 276, 2019.
13. R. Dong, L. J. Ratliff, Á. A. Cárdenas, H. Ohlsson, and S. S. Sastry, “Quantifying the Utility--Privacy tradeoff in the internet of things,” ACM Transactions on Cyber-physical Systems, vol. 2, no. 2, pp. 1–28, Apr. 2018, doi: 10.1145/3185511.
14. A. Fahim, H. Fadil and U. A. Bhatti, “Forecasting the Carbon Emissions in African Countries Using EMD-GCN,” 2023 IEEE 6th International Conference on Pattern Recognition and Artificial Intelligence (PRAI), Haikou, China, pp. 1248-1253, 2023, doi: 10.1109/PRAI59366.2023.10332019.
15. Md. O. Faruque, Md. A. J. Rabby, H. Ma, Md. R. Islam, M. M. Rashid, and S. M. Muyeen, “A comparative analysis to forecast carbon dioxide emissions,” Energy Reports, vol. 8, no.11, pp. 8046–8060, 2022, doi: 10.1016/j.egyr.2022.06.025.
16. F. Firouzi et al., “Fusion of IoT, AI, Edge–Fog–Cloud, and Blockchain: Challenges, solutions, and a case study in healthcare and medicine,” IEEE Internet of Things Journal, vol. 10, no. 5, pp. 3686–3705, 2023, doi: 10.1109/jiot.2022.3191881.
17. T. Gan, H. Yang, W. Liang, and X. Liao, “Do economic development and population agglomeration inevitably aggravate haze pollution in China? New evidence from spatial econometric analysis,” Environmental Science and Pollution Research International, vol. 28, no. 5, pp. 5063–5079, 2020, doi: 10.1007/s11356-020-10847-4.
18. S. Graham et al., “Artificial Intelligence for Mental Health and Mental Illnesses: an Overview,” Current Psycchiatry Reports/Current Psychiatry Reports, vol. 21, no. 11, p.116, 2019, doi: 10.1007/s11920-019-1094-0.
19. S. M. Hosseini, A. Saifoddin, R. Shirmohammadi, and A. Aslani, “Forecasting of CO2 emissions in Iran based on time series and regression analysis,” Energy Reports, vol. 5, no. 4, pp. 619–631, 2019, doi: 10.1016/j.egyr.2019.05.004.
20. L. Huang, “A Study on the CO2 Emissions of Mainland China Using Deep Learning Models,” 2023 4th International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), Nanjing, China, pp. 256-265, 2023, doi: 10.1109/ICBASE59196.2023.10303182.
21. Md. R. Islam, A. B. A. Ghani, and E. Mahyudin, “Carbon Dioxide Emission, Energy Consumption, Economic Growth, Population, Poverty and Forest Area: Evidence from Panel Data Analysis,” IJEEP, vol. 7, no. 4, pp. 99–106, 2017.
22. S. K. Jagatheesaperumal, M. Rahouti, K. Ahmad, A. Al-Fuqaha, and M. Guizani, “The duo of artificial intelligence and big data for industry 4.0: applications, techniques, challenges, and future research directions,” IEEE Internet of Things Journal, vol. 9, no. 15, pp. 12861–12885, 2022, doi: 10.1109/jiot.2021.3139827.
23. S. Kumari and S. K. Singh, “Machine learning-based time series models for effective CO2 emission prediction in India,” Environmental Science and Pollution Research International, vol. 30, no. 55, pp. 116601–116616, 2022, doi: 10.1007/s11356-022-21723-8.
24. X. Li, K. Chalvatzis, and D. Pappas, “Life cycle greenhouse gas emissions from power generation in China’s provinces in 2020,” Applied Energy, vol. 223, pp. 93–102, 2018, doi: 10.1016/j.apenergy.2018.04.040.
25. X. B. Li and X. Zhang, “A comparative study of statistical and machine learning models on carbon dioxide emissions prediction of China,” Environmental Science and Pollution Research International, vol. 30, no. 55, pp. 117485–

117502, 2023, doi: 10.1007/s11356-023-30428-5.
26. C. Magazzino, M. Mele, G. Morelli, and N. Schneider, “The nexus between information technology and environmental pollution: Application of a new machine learning algorithm to OECD countries,” Utilities Policy, vol. 72, no.10, p. 101256, 2021, doi: 10.1016/j.jup.2021.101256.
27. G. Meng, Z. Guo, and J. Li, “The dynamic linkage among urbanisation, industrialisation and carbon emissions in China: Insights from spatiotemporal effect,” Science of the Total Environment, vol. 760, no. 3, p. 144042, 2021, doi: 10.1016/j.scitotenv.2020.144042.
28. N. Nesa and I. Banerjee, “IoT-Based sensor Data Fusion for occupancy sensing using Dempster–Shafer Evidence Theory for smart buildings,” IEEE Internet of Things Journal, vol. 4, no. 5, pp. 1563–1570, Oct. 2017, doi: 10.1109/jiot.2017.2723424.
29. R. Parker et al., “Fuel use and greenhouse gas emissions of world fisheries,” Nature Climate Change, vol. 8, no. 4, pp. 333–337, 2018, doi: 10.1038/s41558-018-0117-x.
30. H. Ritchie, P. Rosado, and M. Roser, “Fossil fuels,” Our World in Data, January 05th, 2023. https://ourworldindata.org/fossil-fuels
31. Z. Sadriddin, R. R. Mekuria and R. Isaev, “A Comparative Study of the Analysis of PM2.5 Sources in Kyrgyzstan with 31 Selected Countries,” 2023 17th International Conference on Electronics Computer and Computation (ICECCO), Kaskelen, Kazakhstan, pp. 1-5, 2023, doi: 10.1109/ICECCO58239.2023.10147148.
32. A. Salam, “Internet of things for environmental sustainability and climate change,” in Internet of things, pp. 33–69, 2019. doi: 10.1007/978-3-030-35291-2_2.
33. C. Sikdar and K. Mukhopadhyay, “Impact of population on carbon emission: lessons from India,” Asia-Pacific Development Journal, vol. 23, no. 1, pp. 105–132, 2017, doi: 10.18356/b4c7cba0-en.
34. J. C. Yeh and C.-C. Liao, “Impact of population and economic growth on carbon emissions in Taiwan using an analytic tool STIRPAT,” Sustainable Environment Research, vol. 27, no. 1, pp. 41–48, 2017, doi: 10.1016/j.serj.2016.10.001.
35. V. Tanania, S. Shukla and S. Singh, “Time Series Data Analysis And Prediction Of CO2 Emissions,” 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, pp. 665-669, 2020, doi: 10.1109/Confluence47617.2020.9058001.
36. Y. Xia, H. Wang, and W. Liu, “The indirect carbon emission from household consumption in China between 1995–2009 and 2010–2030: A decomposition and prediction analysis,” Computers & Industrial Engineering, vol. 128, pp. 264–276, 2019, doi: 10.1016/j.cie.2018.12.031.