



**DE MONTFORT  
UNIVERSITY  
LEICESTER**

# **PREDICTIVE MODELLING FOR EMISSIONS PER COUNTRY USING MACHINE LEARNING**

Giwa-Daramola Inioluwa

P2712256

This project is submitted as part of the requirements for the degree of MSc in Data Analytics

at the Faculty of Computing, Engineering & Media,

De Montfort University- Leicester, UK..

May, 2023

Supervisor: Sean Xavier Laurence

## Abstract

The research aims to investigate the relationships and forecast the future trends of these greenhouse gas emissions. This study utilizes VARMAX modeling to predict CO<sub>2</sub>, N<sub>2</sub>O, and CH<sub>4</sub> emissions in seven selected countries: Spain, Japan, United States, Belarus, Ukraine, Nigeria, and Colombia. Statistical tests confirmed non-stationarity, leading to data differencing for improved stationarity. The Granger Causality test revealed a causal effect of N<sub>2</sub>O emissions on CH<sub>4</sub> and CO<sub>2</sub> emissions. The dataset was split into training and test sets, with predictions indicating varying emission trends across countries. In Spain, CO<sub>2</sub> emissions peaked in 2007, followed by a downward trend, while CH<sub>4</sub> and N<sub>2</sub>O emissions decreased over time but started to rise again in 2012. Predictions indicate a steady increase in CH<sub>4</sub> and CO<sub>2</sub> emissions over the forecasted period, with stable N<sub>2</sub>O emissions. Similar patterns were observed in Japan and the United States, with slight variations in the emission trends. Belarus exhibited increasing levels of CH<sub>4</sub>, CO<sub>2</sub>, and N<sub>2</sub>O emissions since 2010, and predictions suggest a gradual decrease in CH<sub>4</sub> and CO<sub>2</sub> emissions, while N<sub>2</sub>O emissions are also expected to decline. Ukraine displayed a downward trend in CH<sub>4</sub> and CO<sub>2</sub> emissions, while N<sub>2</sub>O emissions consistently increased. Nigeria's emissions remained relatively stable, with a slight increase over time, while Colombia experienced an overall increasing trend in CH<sub>4</sub>, CO<sub>2</sub>, and N<sub>2</sub>O emissions.

## Acknowledgments

I would like to express sincere gratitude to everyone who contributed to the success of this project.

First and foremost, I would like to thank my project supervisor Sean Xavier Laurence for his unwavering guidance, support, and expertise throughout the project. His valuable insights and feedback have been instrumental in shaping the direction of this project.

Finally, I'm grateful to my families, friends, and loved ones, whose unwavering support, encouragement, and understanding have been invaluable throughout this project.

## Declaration

I hereby declare that all the information present in this project was researched and compiled solely by me, without any external aid, and without any previous use case. All sources used throughout this study have been appropriately cited

Signed ..... Giwa-Daramola Inioluwa.....

Date .....May, 2023.....

## Table of Contents

<b>Abstract.....</b>	<b>i</b>
<b>Acknowledgments .....</b>	<b>ii</b>
<b>Declaration.....</b>	<b>iii</b>
<b>List of Tables .....</b>	<b>vii</b>
<b>List of Figures.....</b>	<b>viii</b>
<b>Acronyms .....</b>	<b>x</b>
<b>1.0 INTRODUCTION.....</b>	<b>1</b>
1.1 Background of the study .....	1
1.2 Statement of the problem .....	2
1.3 Aims and Objectives .....	3
1.4 Research Questions .....	4
1.5 Project Structure.....	4
1.6 Conclusion .....	4
<b>2.0 Literature Review .....</b>	<b>5</b>
2.1 Introduction.....	5
2.2 Climate Change.....	6
2.3 Machine Learning .....	9
2.4 Machine Learning for Predictive Analysis .....	10
2.5 Machine Learning Models .....	11
2.5.1 Autoregressive-integrated moving average (ARIMA) model .....	11
2.5.2 Seasonal autoregressive-integrated moving average with exogenous factors (SARIMAX) model .....	12
2.5.3 Holt-Winters model .....	13
2.5.4 Long Short-Term Memory Neural Network (LSTM) .....	15
2.5.5 Bidirectional Long Short-Term Memory Neural Network (BiLSTM).....	17
2.5.6 Artificial Neural Network (ANN).....	18
<b>3.0 Methodology .....</b>	<b>20</b>
3.1 Introduction.....	20
3.2 Feature Engineering .....	20

3.3 Machine Learning Model.....	20
3.4 Performance Evaluation Metrics.....	21
<b>Chapter 4 .....</b>	<b>23</b>
4.1 Data Analysis .....	23
4.1.1 Data Availability .....	23
4.1.2 Dataset Description.....	23
4.1.3 Data Processing.....	24
4.1.4 Date Restructuring .....	24
4.1.5 Statistical Tests .....	25
4.1.6 Train-Test Split .....	25
4.1.7 Model Building and Evaluation .....	25
4.2 Emission patterns in Countries .....	26
4.2.1 Emission patterns in Spain.....	26
4.2.2 Emission patterns in Japan.....	27
4.2.3 Emission patterns in United States .....	28
4.2.4 Emission patterns in Belarus.....	29
4.2.5 Emission patterns in Ukraine .....	30
4.2.6 Emission patterns in Nigeria.....	31
4.2.7 Emission patterns in Colombia .....	32
4.3 Prediction of Emission in Countries .....	33
4.3.1 Prediction of Emission in Spain.....	33
4.3.2 Prediction of Emission in Japan.....	35
4.3.3 Prediction of Emission in United States .....	37
4.3.4 Prediction of Emission in Belarus.....	39
4.3.5 Prediction of Emission in Ukraine .....	41
4.3.6 Prediction of Emission in Nigeria.....	43
4.3.7 Prediction of Emission in Colombia .....	45
4.4 Statistical test .....	46
4.4.1 Statistical test for Spain .....	46
4.4.2 Statistical test for Japan .....	48
4.4.3 Statistical test for United States .....	50
4.4.4 Statistical test for Belarus .....	52
4.4.5 Statistical test for Ukraine.....	54
4.4.6 Statistical test for Nigeria.....	56

4.4.7 Statistical test for Colombia.....	58
<b>Chapter 5 .....</b>	<b>61</b>
5.0 Conclusion .....	61
5.1 Findings and Limitation.....	61
5.2 Future Work .....	63
<b>References.....</b>	<b>64</b>
<b>Appendix.....</b>	<b>70</b>
<b>Research Proposal.....</b>	<b>70</b>
<b>Ethics Application Form .....</b>	<b>75</b>

## List of Tables

Table 1. Augmented Dickey-Fuller (ADF) test for Spain

Table 2. Granger Causality test for Spain

Table 3. Augmented Dickey-Fuller (ADF) test for Japan

Table 4. Granger Causality test for Japan

Table 5. Augmented Dickey-Fuller (ADF) test for United States

Table 6. Granger Causality test for United States

Table 7. Augmented Dickey-Fuller (ADF) test for Belarus

Table 8. Granger Causality test for Belarus

Table 9. Augmented Dickey-Fuller (ADF) test for Ukraine

Table 10. Granger Causality test for Ukraine

Table 11. Augmented Dickey-Fuller (ADF) test for Nigeria

Table 12. Granger Causality test for Nigeria

Table 13. Augmented Dickey-Fuller (ADF) test for Columbia

Table 14. Granger Causality test for Columbia



## List of Figures

Fig 2.1 Change in global temperatures .....	7
Fig. 2.2 Change in global surface temperature showing year 1940, 2000 and 2022 .....	8
Fig 2.3 Comparison between RNN and LSTM .....	15
Fig 2.4 LSTM Architecture.....	16
Fig 2.5 BiLSTM Basic Architecture .....	17
Fig 2.6 Basic Architectire of Artificial Neural Network .....	18
Fig 3.1 RMSE formular .....	21
Fig. 4.1 CH <sub>4</sub> , CO <sub>2</sub> and N <sub>2</sub> O Emissions for Spain .....	26
Fig. 4.2 CH <sub>4</sub> , CO <sub>2</sub> and N <sub>2</sub> O Emissions for Japan .....	27
Fig. 4.3 CH <sub>4</sub> , CO <sub>2</sub> and N <sub>2</sub> O Emissions for United States .....	28
Fig. 4.4 CH <sub>4</sub> , CO <sub>2</sub> and N <sub>2</sub> O Emissions for Belarus .....	29
Fig. 4.5 CH <sub>4</sub> , CO <sub>2</sub> and N <sub>2</sub> O Emissions for Ukraine.....	30
Fig. 4.6 CH <sub>4</sub> , CO <sub>2</sub> and N <sub>2</sub> O Emissions for Nigeria.....	31
Fig. 4.7 CH <sub>4</sub> , CO <sub>2</sub> and N <sub>2</sub> O Emissions for Colombia.....	32
Fig. 4.8 Predicted Emissions for CO <sub>2</sub> in Spain.....	33
Fig. 4.9 Predicted Emissions for CH <sub>4</sub> in Spain.....	33
Fig. 4.10 Predicted Emissions for N <sub>2</sub> O in Spain.....	34
Fig. 4.11 Predicted Emissions for CO <sub>2</sub> in Japan.....	35
Fig. 4.12 Predicted Emissions for CH <sub>4</sub> in Japan.....	35
Fig. 4.13 Predicted Emissions for N <sub>2</sub> O in Japan.....	36
Fig. 4.14 Predicted Emissions for CO <sub>2</sub> in United States.....	37
Fig. 4.15 Predicted Emissions for CH <sub>4</sub> in United States.....	37
Fig. 4.16 Predicted Emissions for N <sub>2</sub> O in United States .....	38
Fig. 4.17 Predicted Emissions for CO <sub>2</sub> in Belarus.....	39
Fig. 4.18 Predicted Emissions for CH <sub>4</sub> in Belarus.....	39
Fig. 4.19 Predicted Emissions for N <sub>2</sub> O in Belarus.....	40
Fig. 4.20 Predicted Emissions for CO <sub>2</sub> in Ukraine .....	41
Fig. 4.21 Predicted Emissions for CH <sub>4</sub> in Ukraine .....	41
Fig. 4.22 Predicted Emissions for N <sub>2</sub> O in Ukraine.....	42
Fig. 4.23 Predicted Emissions for CO <sub>2</sub> in Nigeria .....	43

Fig. 4.24 Predicted Emissions for CH <sub>4</sub> in Nigeria .....	43
Fig. 4.25 Predicted Emissions for N <sub>2</sub> O in Nigeria.....	44
Fig. 4.26 Predicted Emissions for CO <sub>2</sub> in Colombia .....	45
Fig. 4.27 Predicted Emissions for CH <sub>4</sub> in Colombia .....	45
Fig. 4.28 Predicted Emissions for N <sub>2</sub> O in Colombia.....	46

## Acronyms

ADF Augmented Dickey-Fuller

AI Artificial Intelligence

AIC Akaike Information Criterion

ANN Artificial Neural Networks

AR Autoregression

ARIMA Autoregressive-integrated moving average

BIC (Bayesian Information Criterion)

BiLSTM Bidirectional Long short-term memory model

CH<sub>4</sub> Methane

CO<sub>2</sub> Carbon dioxide

DL Deep Learning

IPCC Intergovernmental Panel on Climate Change

LSTM Long short-term memory model.

MA Moving Average

ML Machine Learning

N<sub>2</sub>O Nitrous Oxide

NASA National Aeronautics and Space Administration

NDI National Development Index

RMSE Root mean squared error

SARIMAX Seasonal autoregressive-integrated moving average with exogenous factors

SVM Support Vector Machine

VARMAX Vector Autoregressive Moving Average model with exogenous variables

WMO World Meteorological Organization

## 1.0 INTRODUCTION

Climate change is one of the greatest challenges facing humanity in the 21st century. According to the National Aeronautics and Space Administration (NASA), the earth's global surface temperature in 2020 was the second-highest on record since 1880 (NASA, 2020). The Earth's climate has been changing throughout its history, but the current warming trend is of particular concern due to its rapid rate and human-induced causes. From rising sea levels to more frequent and severe extreme weather events, the impacts of climate change are already being felt around the world (Shi, 2018). As such, monitoring climate change is crucial in order to understand its impacts on ecosystems, economies, and societies, and to develop effective strategies for mitigating and adapting to these impacts. The need to identify the sources of greenhouse gas emissions and predict future emissions has become increasingly important in order to mitigate the impacts of climate change (Deetman et al., 2020). One approach to this problem is to use predictive modelling and machine learning techniques

Predictive modeling, involves the use of machine learning algorithms to make predictions hinged on historical data. This technology presents a workable solution to this problem (Lakshay and Pratika, 2017). The analysis of historical data on emission changes, patterns, trends and other relevant factors can aid the creation of predictive models that can generate accurate predictions of future emissions (Fatimetou, 2017). These predictions can then be used by policymakers to design effective policies and interventions to reduce emissions at a global scale.

This study aims to adopt a comprehensive approach that traverses various regions and continents to improve our understanding of emission patterns. By using machine learning techniques, researchers can identify complex patterns and relationships between different variables that may not be immediately apparent using traditional statistical methods.

### 1.1 Background of the study

Climate change is a global challenge that has become increasingly urgent to address. Greenhouse gas emissions, primarily carbon dioxide (CO<sub>2</sub>), are the main contributors to global warming and climate change (Shi, 2018). According to the Intergovernmental Panel on Climate Change (IPCC), the global average temperature has increased by 1.1°C since pre-industrial times, and it is projected to continue to increase in the coming decades if emissions continue to rise at the current rate (IPCC, 2021).

One way to address this challenge is to reduce greenhouse gas emissions. To achieve this, it is important to understand the sources and drivers of emissions, as well as the factors that influence them. Countries play a crucial role in this regard, as they are responsible for a significant proportion of global emissions. In 2019, global CO<sub>2</sub> emissions from fossil fuels and industry were about 36.4 billion tonnes, and the top five emitters (China, United States, India, Russia, and Japan) accounted for more than half of these emissions (Friedlingstein et al., 2020).

Some projects such as the Forecasting of transportation-related energy demand and CO<sub>2</sub> emissions in Turkey with different machine learning algorithms by Umit Agbulut, and Machine learning-based time series models for effective CO<sub>2</sub> emission prediction in India by Surbhi Kumanr and Suni Kumar Singh have done project related to the prediction of CO<sub>2</sub> emissions in India and Turkey. Kumari and Singh used three statistical models which were the autoregressive-integrated moving average (ARIMA) model, the seasonal autoregressive-integrated moving average with exogenous factors (SARIMAX) model, and the Holt-Winters model, and two machine learning models, i.e., linear regression and random forest model and a deep learning based long short-term memory (LSTM) model. Agbulut used Deep Learning (DL), Support Vector Machine (SVM) and Artificial Neural Networks (ANN). An exploration of clustering and classification machine learning algorithms will inform new ideas and aid the creation of predictive models.

climate change is a pressing global issue with significant implications for ecosystems, economies, and societies around the world. Monitoring climate change is crucial to understanding its impacts and developing effective strategies for mitigating and adapting to these impacts (Haines et al., 2006). Through continued research and data collection, we can work towards a more sustainable future for ourselves and for future generations.

## 1.2 Statement of the problem

The most recent past decade, which includes year 2011 through year 2020 was observed to be the hottest decade. With current trends, global temperatures are expected to rise by another 1.5°C between 2030 and 2052, which would lead to more frequent and severe climate-related disasters (IPCC, 2021). The release of emissions such as CO<sub>2</sub>, CH<sub>4</sub>, or N<sub>2</sub>O etc., into the atmosphere as a result of human activities has significantly contributed to current global temperatures, negatively affecting the health of individuals directly and indirectly (Patz et al., 2005). This creates the need

for a level of urgency when dealing with this global challenge. Governments, organizations and global stakeholders have outlined some best practices which can contribute to the reduction of gas emissions and its effects, however, accurate and reliable information on the country-by-country emissions is needed for setting and achieving targets which reduce the emission of greenhouse gases; which are the main drivers of climate change.

With climate change comes negative occurrences which are already being observed in different countries, all around the globe. Rising sea levels, due to melting ice sheets and glaciers, are causing more frequent and severe flooding and storm surges, which threaten coastal communities and infrastructure. Changes in temperature and rainfall patterns are also affecting food security, with some estimates suggesting that global crop yields could decline by up to 25% by 2050 (Pathak et al., 2010). Additionally, climate change is exacerbating other global challenges, such as biodiversity loss, water scarcity, and public health risks.

Given the significant impacts of climate change, it is essential to monitor its effects in order to understand the scope of the problem and develop effective solutions. This requires collecting and analyzing data on a range of indicators, such as temperature, precipitation, sea level, and greenhouse gas concentrations. By monitoring these indicators over time, researchers can track changes in the climate system and identify areas where intervention is needed to reduce greenhouse gas emissions and promote adaptation to climate change impacts.

### 1.3 Aims and Objectives

The main aim of this project is to identify the causes of the country level emission patterns, and build solutions which can better forecast expected emission trends. This will aid with the policy making and decision taking.

1. Build predictive models to forecast future emissions trends by location.
2. Analyze the changes in emissions across different countries.
3. Present an analysis of emission trends by types (e.g., CO<sub>2</sub>, CH<sub>4</sub>, or N<sub>2</sub>O etc).
4. Developing an understanding of the detrimental effect of the different types of emissions based on predictions for the next 10 years.

## 1.4 Research Questions

1. How can machine learning models be effectively used to forecast future emissions trends?
2. What emission disparity can be observed amongst different countries?
3. How do different emission types (e.g., CO<sub>2</sub>, CH<sub>4</sub>, or N<sub>2</sub>O etc) vary across countries, and what factors contribute most to these differences?
4. How can we explain the detrimental effect of the different types of emissions based on predictions for the next 10 years?

## 1.5 Project Structure

This introduction section of the project highlights the goal of this project. It identifies the aim, objectives, as well as the research questions which will be answered during the course of this research. An in-depth review of literature which focuses on climate change, green gas emissions and machine learning algorithms that will aid predictive modelling will be discussed in the Literature Review. The Methodology section will highlight all methods used to achieve my goal, while the Result section will describe the outcomes observed during this study.

Finally, key findings will be concluded and summarized in the Conclusion section of the project

## 1.6 Conclusion

The scope and impact of climate change on the globe is vast, creating the need for diverse solutions which tackle this problem at a global scale.

## 2.0 Literature Review

### 2.1 Introduction

Environmental challenges are among the most urgent problems that modern society is facing. Present-day developed and emerging economies' top environmental and political concerns are climate change and global warming (Aftab et al., 2021, Rehman et al., 2021). Its detrimental effects on humans are getting worse, which is difficult. The primary factor for climate change is the emission of greenhouse gases, the majority of which are carbon dioxide (CO<sub>2</sub>) (Mitić et al., 2017, Faruque et al., 2022).

The Worldwide Climate Report from the World Meteorological Organization (WMO) estimates that the average global temperature in 2020 was around 1.2 °C higher than preindustrial levels (Li et al., 2021). The industrialized nations work to cut back on the consumption of these fossil fuels, but doing so comes at a price that many nations cannot bear (Wani et al., 2022). Over the decade from 2000 to 2018, developed nations and economies in transition's global greenhouse gas emissions decreased by 6.5 percent. Between 2000 and 2013, the emissions of the developing nations increased by 43.2%. The rise is mostly due to higher industrialization and improved GDP-measured economic production (United Nations, 2023, Ritchie et al., 2020).

Energy scientists and policy makers are concerned about the adverse environmental effects of consumption of energy and related social benefits due to the growing decline in the environment observed at all levels, including the national and global (Kone and Buke, 2010).

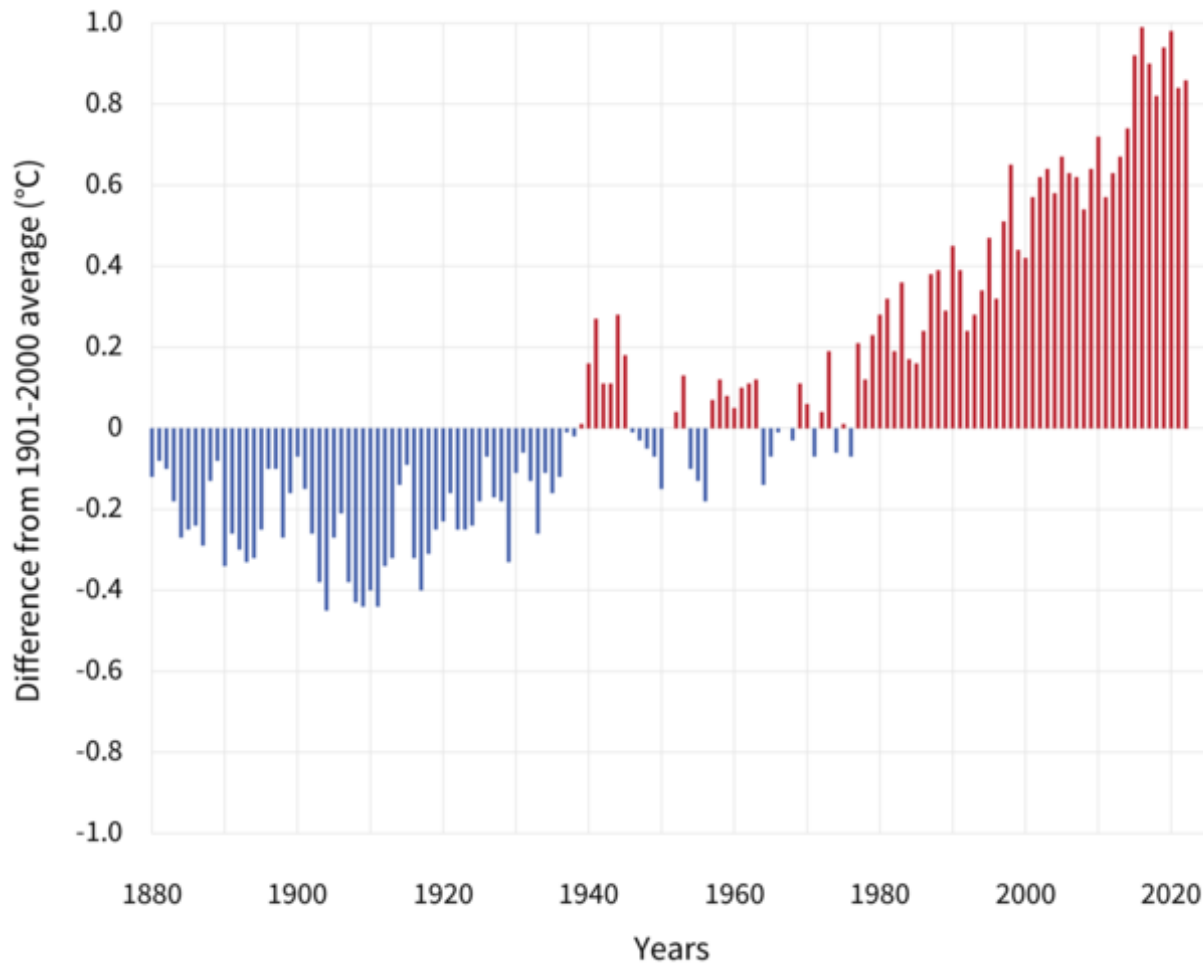
Global worry over CO<sub>2</sub> emissions forecasts has grown as a result of research showing that this greenhouse gas (GHG) has the greatest effects on environmental issues. A crucial component of raising public awareness of environmental issues is the forecasting of CO<sub>2</sub> emissions (Abdullah and Pauzi, 2015). The fundamental elements of a clean energy economy and a market with high growth are analyses and projections of carbon emissions, energy consumption, and real outputs (Xu et al., 2021). Reducing CO<sub>2</sub> emissions is a global issue that has to be tackled in order to create a sustainable society (Faruque et al., 2022).



## 2.2 Climate Change

Greenhouse gases are those that contribute to the greenhouse effect, a natural phenomenon that has been present on Earth since its creation. The basic greenhouse gasses are Carbon Dioxide, Methane( $\text{CH}_4$ ), Nitrous Oxide( $\text{N}_2\text{O}$ ), and Industrial Gases, such as sulfur hexafluoride hydrofluorocarbons, and perfluorocarbons (Doll and Baransk, 2011). While some of these gases are emitted by natural processes, others are released as a result of human activities such as combustion of fossil fuels, oil, gas, coal, and deforestation. The accumulation of greenhouse gases in the Earth's atmosphere leads to an increase in the Earth's surface temperature, which is commonly referred to as global warming (Darkwah et al., 2018). This phenomenon occurs as greenhouse gases trap and absorb solar radiation, thereby increasing the amount of heat in the atmosphere. As a result, the temperature of the Earth's surface gradually rises, leading to a range of environmental impacts, such as sea level rise, changes in precipitation patterns, and ecosystem disruptions (Juniarko et al., 2015, Kweku et al., 2017).

One of the most embarrassing challenges in the world today, particularly in third-world nations, is global warming and climate change. Adequate  $\text{CO}_2$  is required for plants, yet commercial chimneys and land, space, and maritime vehicles emit tons of extra  $\text{CO}_2$ , contributing significantly to the impact of greenhouse gases, global warming, and climate change (Meng and Noman, 2022). In the early 20th century, a number of scientists opposed the theory that rising  $\text{CO}_2$  levels are responsible for global warming. One such scientist was Angström, who argued that the overlap between the  $\text{CO}_2$  and water vapor spectral bands and the saturation of consumption near the focal point of the 15 m band would leave little room for additional effects. However, by the 1970s, most scientists agreed that rising  $\text{CO}_2$  concentrations would lead to higher global surface temperatures (Zhong and Haigh, 2013).



*Fig 2.1 Change in global temperatures*

The rate of industrialization and urbanization is accelerating quickly, which has increased the amount of greenhouse gas emissions (Mohammed Redha Qader et al., 2022). The primary greenhouse gas that contributes to global warming is carbon dioxide, and burning fossil fuels like coal and petroleum would significantly increase carbon dioxide emissions.(Xu et al., 2021) Across industries and geographical areas, human-induced climate change is having broad negative consequences and inflicting harm to the environment and human lives (Intergovernmental Panel on Climate Change, 2022). According to estimates, atmospheric levels of carbon dioxide rose by 40%, methane by 150%, and nitrous oxide concentrations increased by 20% between 1750 and 2011 (Denchak, 2019). Burning fossil fuels contributed to higher atmospheric carbon dioxide over the past century as a consequence of commercial activity. Given that CO<sub>2</sub> emissions account for

nearly half of the Earth's net solar retention, it is commonly accepted that they are the primary driver of global warming.

Economic growth has a profound impact on a country's energy consumption patterns, which, in turn, can significantly increase the rate of carbon dioxide (CO<sub>2</sub>) emissions into the atmosphere. This effect is primarily due to the expansion of industries and the resulting rise in energy demand (Malik and Lan, 2016). As industries grow and diversify, they require more and more energy to sustain their operations. Consequently, the increase in energy consumption leads to a surge in CO<sub>2</sub> emissions, which can have adverse effects on the environment and contribute to global warming (Magazzino and Mele, 2022).

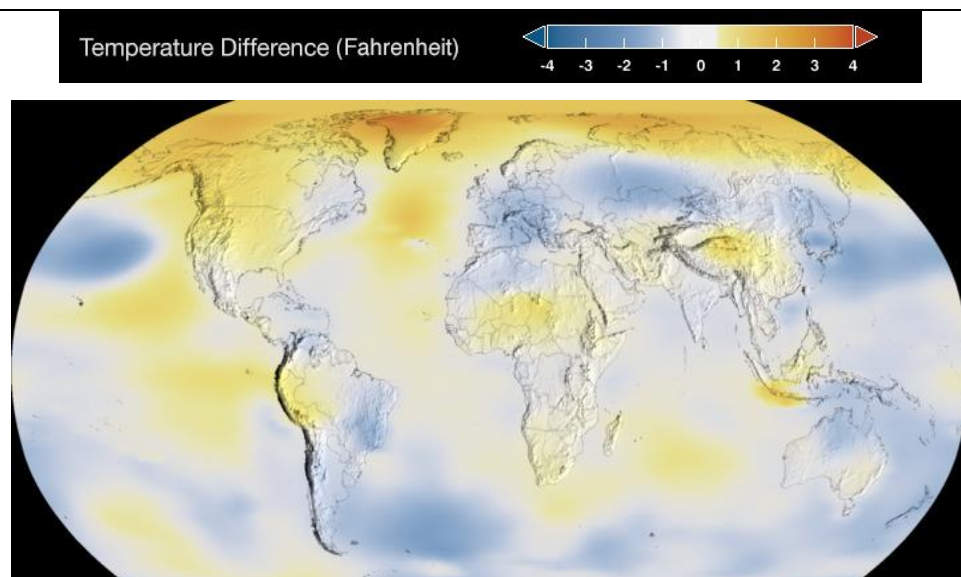
Due to higher temperatures and more frequent catastrophic weather events, global warming is predicted to bring widespread extinction. It is undeniably true that efforts to cut greenhouse gas emissions helped to accelerate international cooperation and the approval of international agreements. Dozens of international environmental accords already exist, even if there are just a few participating nations, and they have helped to shape the global environmental system (Yu et al., 2022).

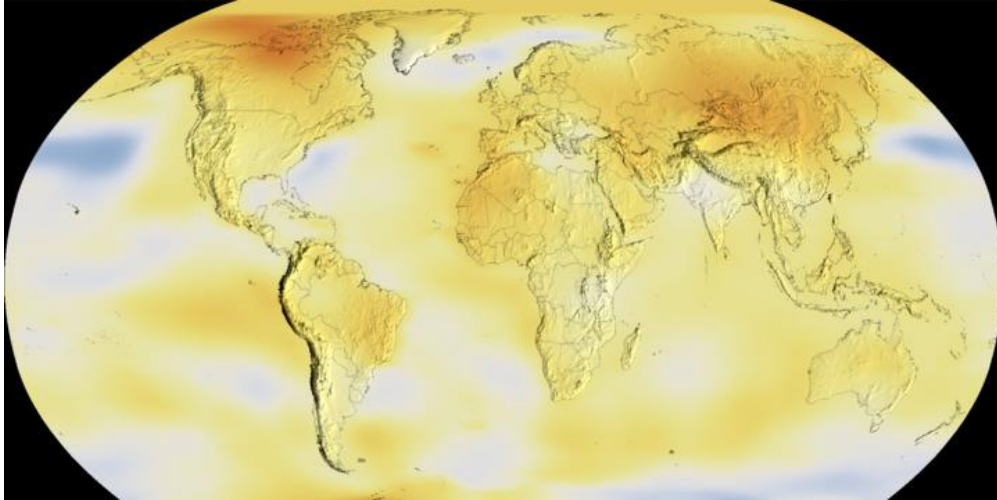
---

*Fig. 2.2 Change in global surface temperature showing year 1940, 2000 and 2022*

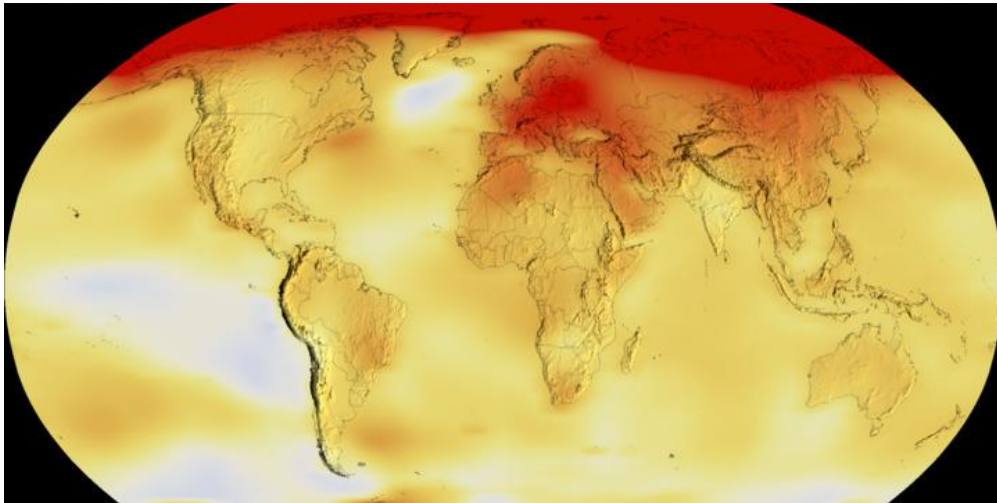
Source: National Aeronautics and Space Administration (2023)

---





Year:2000



Year:2022

---

### 2.3 Machine Learning

Artificial intelligence (AI) has grown in popularity over the past ten years, both inside and outside of the scientific community. In 1956, a team of computer scientists put forward the idea that machines could be programmed to simulate human thinking and reasoning. They believed that every aspect of learning and intelligence could be accurately described and that machines could be designed to imitate these processes. They termed this concept "artificial intelligence" (AI)(Moor, 2006). Essentially, AI is a field that aims to automate cognitive tasks that are typically carried out

by humans. Machine learning (ML) and deep learning (DL) are specific techniques used to achieve this goal(Choi et al., 2020).

Machine learning (ML) is a computational technique that involves using algorithms to automatically learn from data and discover hidden patterns or insights. For example, in a medical context, an ML model could be trained on a dataset of patients with various health conditions, along with data on their symptoms, lifestyle factors, and medical history. The model would learn to associate patterns in this data with different health conditions, and could then be used to predict the likelihood of a new patient having a particular condition based on their symptoms and other data (Jovel and Greiner, 2021, Alloghani et al., 2019, Choi et al., 2020).

Similarly, in a chemistry context, an ML model could be trained on data on the chemical properties of molecules and their interactions with other molecules. The model would learn to recognize patterns in this data that indicate whether two molecules are likely to interact or not, and could then be used to predict the behavior of new molecules (Liu et al., 2020).

ML can be broadly categorized into two types: supervised learning and unsupervised learning. In supervised learning, data is labeled and the algorithm makes predictions based on input-output pairs. The output is known before creating the model, and then the model is used to make predictions on new input data. Unsupervised learning, on the other hand, does not require labeled data and output datasets to create a model. It is based solely on the input dataset and is typically used for tasks like data clustering(Alloghani et al., 2019, Choi et al., 2020). There are many clustering algorithms available in popular ML libraries for Python or R (Zareba et al., 2022).

## 2.4 Machine Learning for Predictive Analysis

Forecasting is a prevalent aspect of various AI applications. For instance, predicting the likelihood of developing a severe illness, identifying voters who are most inclined to support a specific candidate, anticipating driver conduct, and suggesting videos or advertisements that may appeal to a particular individual. Predictive analysis is becoming more widespread in use (Bokonda et al., 2020).

Predictive analytics involves analyzing past trends to create forecasts for the future in various fields. In recent years, machine and deep learning have led to the development of sophisticated predictive models, which have helped to identify complex hidden patterns in data and significantly

improve prediction accuracy (Sghir et al., 2022). To conduct predictive analysis using machine learning, Brooks and Thompson (2017) suggest a few key steps. Firstly, the problem must be identified. Then, the necessary data must be collected and the predicted outcome must be defined. Next, appropriate predictor variables that have a strong correlation with the desired output must be selected. Finally, a predictive model can be built using one or more algorithms.

## 2.5 Machine Learning Models

There are various models that are commonly used for time series prediction, and the choice of model depends on the specific function being predicted and the desired level of accuracy. In many studies focused on predicting carbon emissions, popular models include Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive-Integrated Moving Average with Exogenous Factors (SARIMAX), Holt-Winters model, and Long Short-Term Memory Neural Network (LSTM).

### 2.5.1 Autoregressive-integrated moving average (ARIMA) model

A set of time-domain models called the ARIMA model, created by Box and Jenkins in 1970 and often referred to as the Box-Jenkins technique, is frequently used to fit and predict time series that demonstrate temporal correlation (Lai and Dzombak, 2020, Jennings et al., 2016). Many researchers believe that ARIMA models provide projections that are more accurate than those produced by econometric methods (Nyoni and Wellington G., 2019). The ARIMA model has been used in several research for a variety of purposes, including predicting time series associated with climate or climate-related variables.

There are three factors that define the ARIMA model and they are p, d and q where p is the number of previous observations that would be considered for the autoregression (lag order), d is the frequency of difference of the raw observations (degree of differencing), and q is the moving average window's size (moving average order) (Hayes, 2022, Projectpro, 2022). According to Projectpro (2022), ARIMA can be defined by the Moving Average model as:

$$y_t = c + \epsilon_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}$$

Where  $y_t$  is the data that the model would be applied on

Using ARIMA is advantageous because it takes into account the trend, seasonality, and cyclical behavior of a time series, making it more robust than simpler models such as the exponential smoothing model. Additionally, ARIMA models can be easily extended to include external factors or predictors, allowing for more accurate forecasting in complex scenarios. The model is also relatively easy to interpret and can provide insights into the underlying factors driving the time series behavior. However, ARIMA models can be computationally intensive and require a significant amount of historical data to fit the model accurately (Bora, 2021).

Juniarko et al. (2015) used the ARIMA model to predict the emission of Carbon Dioxide in Surabaya Municipality in Indonesia, Ning et al. (2021) used ARIMA model to forecast Carbon emissions in China, Nyoni and Wellington G. (2019) used ARIMA model to predict Carbon emissions in India, Rahman and Hassan (2017) used ARIMA model to forecast Carbon Dioxide emissions in Bangladesh, Lotfalipour et al. (2013) combined ARIMA and Gray method in Iran, Chigora et al. (2019) used the Box-Jenkins ARIMA model to forecast Carbon emission for Zimbabwe with focus on Tourism.

### **2.5.2 Seasonal autoregressive-integrated moving average with exogenous factors (SARIMAX) model**

The SARIMAX model refers to a seasonal autoregressive-integrated moving average framework that allows for the inclusion of exogenous factors in addition to trends and seasonality. Despite being a form of ARIMA, SARIMAX can effectively model time series data that display both seasonality and trend (Kumari and Singh, 2022).

The SARIMAX model according to Meng and Noman (2022), (Verma, 2021) can be defined by

$$\varphi_p(B)\Phi_p(B)B^S(1-B)^d(1-B^S)^DY_t = \beta_k x_k + \theta_q(B)\Theta_Q(B^S)\varepsilon_t$$

Where:

$$\varphi_p(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$$

$$\Phi_p(B)B^S = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots - \Phi_p B^{pS}$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_p B^q$$

$$\Theta_Q(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots - \Phi_Q B^{QS}$$

$\varphi_p B$  refers to the autoregressive non-seasonal lag

$\Phi_p(B)B^S$  refers to the autoregressive seasonal lag

$\beta_k x_k$  is the trend which includes the intercept

$(1 - B)^d(1 - B^S)^D Y_t$  refers to the time series that is differenced ‘d’ times and differenced seasonally ‘D’ times

$\theta_q(B)$  refers to the moving average non-seasonal lag

$\Theta_Q(B^S)\varepsilon_t$  refers to the moving average seasonal lag

(Verma, 2021)

The advantage of SARIMAX over ARIMA in relation to modeling CO<sub>2</sub> emissions of a country over time is that SARIMAX allows for the inclusion of exogenous variables, while ARIMA does not. This means that SARIMAX can take into account external factors that may affect CO<sub>2</sub> emissions, such as economic growth, population growth, and technological advancements. By incorporating these variables into the model, SARIMAX can potentially provide more accurate and reliable predictions of CO<sub>2</sub> emissions compared to ARIMA. Additionally, SARIMAX can handle seasonal variations and trends in the data, which can be important in modeling CO<sub>2</sub> emissions that may exhibit seasonal patterns or long-term trends.

Singh et al. (2021) utilized SARIMAX as a model to predict carbon emissions in India caused by the Paddy crop, while Kumari and Singh (2022) also employed SARIMAX in their prediction of CO<sub>2</sub> in India. Similarly, Meng and Noman (2022) also used SARIMAX model to predict Carbon emission footprint by comparing Pre, start, Trans and Post COVID-19.

### 2.5.3 Holt-Winters model

Holt-Winters forecasting approach is a straightforward, commonly used projection technique that can handle trend and seasonal variation. Experimental studies, however, tend to demonstrate that the approach is generally less accurate than the more challenging Box-Jenkins strategy (Chatfield, 1987).



Holt-Winters prediction is given as

$$\ell_t = a \frac{y_t}{s_{t-m}} + (1 - \alpha)(\ell_{t-1} + b_{t-1})$$

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$$

$$s_t = y \frac{y_t}{(\ell_t)} + (1 - y)s_{t-m}$$

$$F_{t+m} = (L_t + b_{t+m})S_{t-s+m}$$

Additive Holt-Winters

$$\ell_t = a(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$$

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$$

$$s_t = y(y_t - \ell_{t-1} - b_{t-1}) + (1 - y)s_{t-m}$$

$$F_{t+m} = (L_t + b_{t+m})S_{t-s+m}$$

Where

$y_t$  the latest observation

$\ell_t$  is the series' most recent level estimate

$\ell_{t-1}$  is the level that has already been smoothed

$a$  is the level's smoothing constant

$\beta$  is the trend estimate's smoothing constant

$b_t$  is the estimated current trend

$b_{t-1}$  is the already smoothed trend

$y$  is the smoothing constant for the assessment of seasonality

$s_t$  is the estimated seasonal component

$s_{t-m}$  is the previous seasonal component

$m$  is the number of seasons in a year

$s$  is the length of the seasons

$t$  is the time preiod

The Holt-Winters model has been widely used in carbon emission prediction research. Alam and Alarjani (2021b) used Holt-Winters Exponential Smoothing among others to forecast the CO<sub>2</sub> emissions in Saudi Arabia, while Alam and Alarjani (2021a) used Holt-Winters model to forecast carbon emission in the Gulf countries. Overall, the Holt-Winters model has proven to be a useful tool for predicting carbon emissions, especially when seasonal patterns are present in the data.

#### 2.5.4 Long Short-Term Memory Neural Network (LSTM)

Recurrent neural networks (RNN) have a particular subset known as long short term memory networks(LSTM). LSTMs has the capacity to pick up knowledge from enduring dependence while RNNs train via truncated backpropagation across time (Li et al., 2020). However, when the number of time steps is enormous, RNNs have a vanishing gradient issue. LSTMS are made to get around the vanishing gradient issue. The cell state and numerous gates are updated by LSTMs (Saa and Ranathunga, 2020). This neural network's strong long-term memory function allows it to thoroughly examine the long-term linkages and patterns of small data sets(Alamgir et al., 2020).

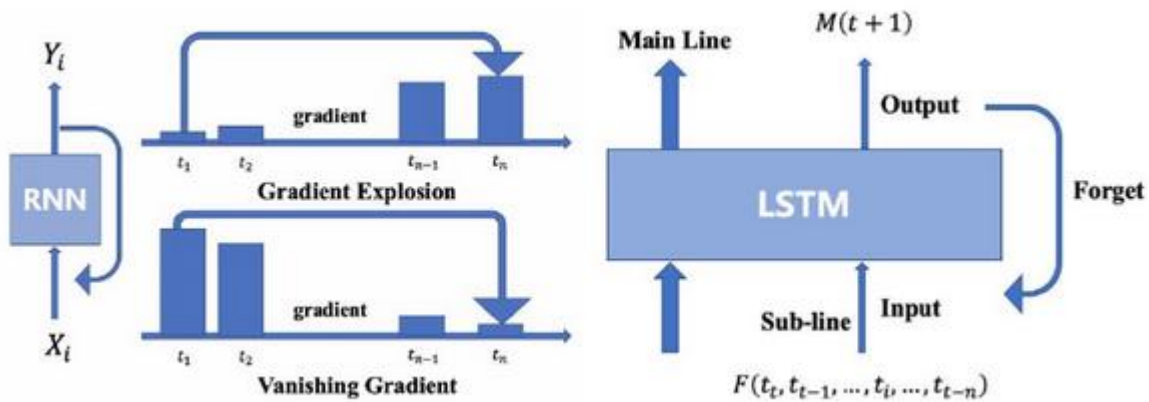


Fig 2.3 Comparison between RNN and LSTM (Hu, 2023)

The input gate, forget gate, and output gate are the three gates used by an LSTM model to regulate its characteristics. The input gate regulates how new information enters the cell state. The forget gate purges the cell state of old, irrelevant data. The output gate controls information that has been taken from the cell state and then chooses the following concealed state. These gates allow an LSTM model to automatically save or delete memory from storage (Hung, 2023). Due to these reasons, one of the models that is recommended as being the best suitable for predicting CO<sub>2</sub> emissions is the LSTM model (Kumari and Singh, 2022).

According to Faruque et al. (2022), LSTM can be computed using the following formula

$$f_t = \sigma g(W_f \cdot x_t + U_f \cdot h_t - 1 + b_f)$$

$$i_t = \sigma g(W_i \cdot x_t + U_i \cdot h_t - 1 + b_i)$$

$$o_t = \sigma g(W_o \cdot x_t + U_o \cdot h_t - 1 + b_o)$$

$$\tilde{c}_t = \tanh(W_c \cdot x_t + U_c \cdot h_t - 1 + b_c)$$

Where  $b_f$ ,  $b_i$ ,  $b_o$ , and  $b_c$  are the bias vectors and  $U_f$ ,  $U_i$ ,  $U_o$  and  $U_c$  are the weight matrices that connect the former output to the three gates and memory cells.

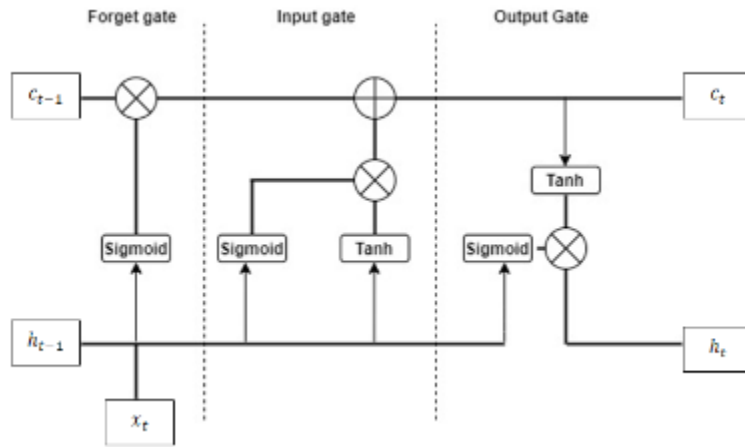


Fig 2.4 LSTM Architecture (Saa and Ranathunga, 2020)

Kumari and Singh (2022) applied the LSTM model to forecast CO<sub>2</sub> emissions in India, while Saa and Ranathunga (2020) use LSTM for weather forecasting. Similarly, Faruque et al. (2022) conducted a comparative analysis using LSTM.

### 2.5.5 Bidirectional Long Short-Term Memory Neural Network (BiLSTM)

A Bidirectional LSTM, commonly referred to as a BiLSTM, is a type of model for sequence processing. It comprises of two separate LSTMs that process the input sequence in two directions: forward and backward. This method enhances the amount of information accessible to the model, thus improving the context that the algorithm has access to. The BiLSTM is often used to gain additional insight into the relationship between different data points, such as the immediate previous and next words in a sentence, thus improving the overall performance of the model (Cornegruta et al., 2016).

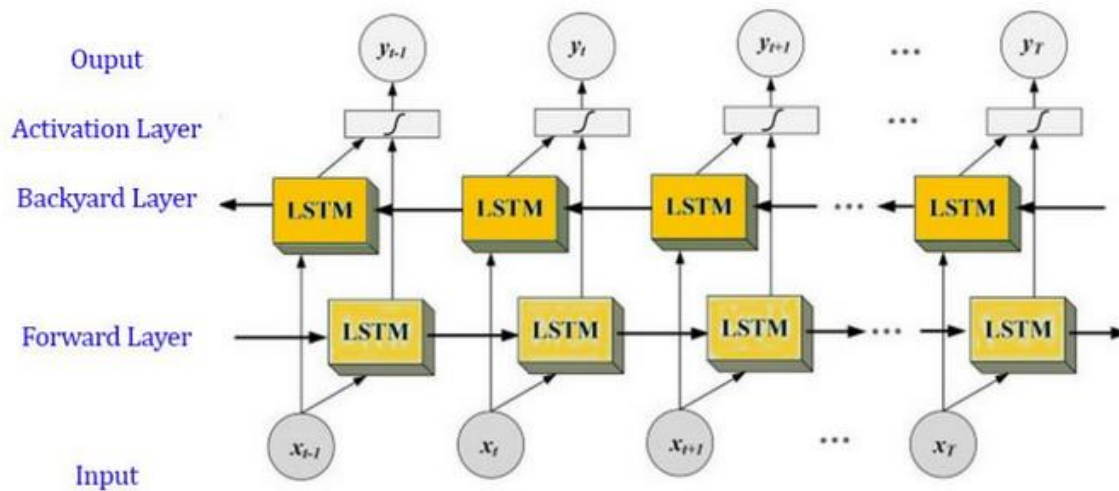


Fig 2.5 BiLSTM Basic Architecture (Bhatti et al., 2022, Aamir et al., 2022)

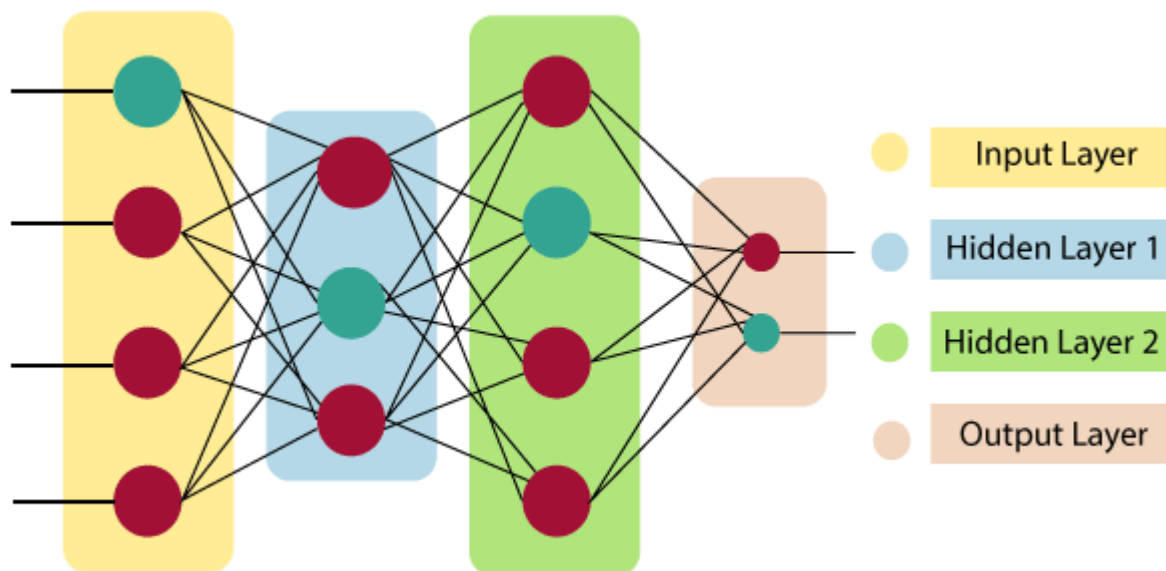
BiLSTM and LSTM differ mainly in the way they process the input sequence. While LSTM processes the sequence only in the forward direction, BiLSTM is bidirectional and processes it in both forward and backward directions. This allows BiLSTM to capture not only past but also future information, making it particularly useful for tasks that require context from both directions. BiLSTM consists of two LSTM layers, one for processing the input sequence in the forward direction and the other for processing it in the backward direction, and their outputs are concatenated at each time step. However, BiLSTM is more complex and computationally expensive than LSTM. The choice between BiLSTM and LSTM depends on the specific task and the input data type, where BiLSTM is more effective for tasks requiring context from both

directions and LSTM is preferred for simpler tasks or when processing speed is a concern(Siami-Namin et al., 2019).

While BiSTM is not a commonly used model in emissions prediction research, Aamir et al. (2022) used it to forecast changes in emission patterns in South Asia.

### 2.5.6 Artificial Neural Network (ANN)

Artificial neural networks are computer systems that take inspiration from biological neural networks in the brain and nervous system. Although they do not replicate the full complexity of these biological systems, they use similar concepts to process information. These networks are composed of processing elements, also known as neurons or perceptrons, which are interconnected to perform computations. The electrical activity of the brain and nervous system is simulated in these models to enable the processing of information (Park and Lek, 2016).



*Fig 2.6 Basic Architectire of Artificial Neural Network (Javatpoint, 2022)*

ANN can be a useful tool for predicting emissions. Researchers have used ANN models to predict carbon dioxide emissions from various sources, including power plants and transportation. ANN models can be trained on historical data and used to make predictions based on current and future input data. Additionally, Artificial neural networks (ANN) have the ability to handle noisy data and can accommodate multiple variables with non-linear, linear, and unknown interactions. This makes them a suitable tool for making generalizations in complex systems (Safa et al., 2016).

Acheampong and Boateng (2019) employed ANN as a forecasting tool to predict carbon emissions in five large countries: Brazil, USA, India, Australia, and China. Similarly, (Thanh et al., 2022) utilized ANN to predict the carbon storage capacity of residual oil zones. ) applied ANN to predict carbon emissions in the United States, while Saleh et al. (2015) used ANN to predict carbon emissions in a sugar industry.

## 3.0 Methodology

### 3.1 Introduction

This chapter outlines the procedure that will be used to achieve the objectives of the study. It will describe the methods used during the experiment and the approach to conducting the literature review. Additionally, the chapter will explain why specific techniques were chosen for the study. The purpose of this section is to provide a clear and concise overview of the steps that will be taken to ensure that the research is conducted in a systematic and rigorous manner. The process will be documented to ensure transparency and replicability of the study. By detailing the methodology, readers will gain an understanding of how the data was collected, analyzed, and interpreted. This will provide the necessary foundation for readers to evaluate the validity and reliability of the findings.

### 3.2 Feature Engineering

Feature engineering is the process of selecting and transforming raw data into features that can be used to train machine learning models. This involves identifying important features, removing irrelevant ones, and creating new features from existing data to improve model accuracy and performance (Hyndman and Athanasopoulos, 2018).

The “elements” attributes will be used as the feature for this study case. Sub categorical data will be filtered out and CO<sub>2</sub> emissions and its equivalents will be merged.

As time series data, differencing will be applied to make it stationary. Differencing entails computing the difference between consecutive values in the time series data. By taking the difference between the current and the previous observation, we can remove the trend and seasonality components from the data, making it stationary (De Brabandere et al., 2021). Stationary data has constant statistical properties such as constant mean and variance over time, which makes it easier to model and analyze.

### 3.3 Machine Learning Model

The Autoregressive-integrated moving average (ARIMA) model will be implemented in this study as it is highly capable of predicting future values, based off of historical behavior. Three components which include the Autoregression (AR) component that refers to predicting future

values using past values of the time series, the Integrated (I) component that refers to the application of differencing to make the data stationary and constant over time and the Moving Average (MA) component refers to the use of past forecast errors to predict future values are key to the success of this model (Fattah et al., 2018).

The appropriate value for each component will be selected based on selection criteria such as AIC (Akaike Information Criterion) or BIC (Bayesian Information Criterion). Next, the model will be built and used to forecast the future emissions of the top 3 developed countries and top 3 developing countries as judged by the national development index (NDI) and world bank development reports

### 3.4 Performance Evaluation Metrics

The performance of the Autoregressive-integrated moving average (ARIMA) model will be evaluated using the root mean square error (RMSE).

The RMSE measures the differences between the predicted values and the real, existing values that are present in the dataset. It is calculated by taking the square root of the average of the squared differences between the predicted values and the actual values (Hodson, 2022). The formular is shown below.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Predicted_i - Actual_i)^2}{N}}$$

*Fig 3.1 RMSE formular (Hodson, 2022)*

This shows how close the predicted values are to the actual values. The RMSE is a great evaluation method for the performance of an ARIMA model as it closely assesses the fit of the model to the data.



A lower RMSE indicates a better fit between predicted values and actual values, while a high RMSE indicates a poor fit between predicted values and actual values (Karno, 2020).

The RMSE of the ARIMA model will accurately determine its overall performance in predicting future emissions of the selected countries.

## Chapter 4

This chapter comprehensively outlines the entire experimental process undertaken to transform the stated aims and objectives of this study into tangible results. It encompasses various stages, starting from data preprocessing, followed by model development, optimization, and evaluation. Additionally, the performance of the models was assessed and compared based on predefined evaluation metrics.

### 4.1 Data Analysis

This part of the text outlines the different activities carried out during the exploratory data analysis phase, which include tasks such as data gathering, description, preparation as well as preparing and examining it.

#### 4.1.1 Data Availability

A dataset was identified and downloaded from the site “Kaggle.com”. This dataset tracks emission source, emission type, and total emissions (e.g., CO<sub>2</sub>, CH<sub>4</sub>, or N<sub>2</sub>O etc.) of greenhouse gasses. The main source of information relied upon was the website of the Food and Agriculture Organization (FAO) of the United Nations. This website presents tables of data relating to environmental issues on a global scale. Given the careful recording of official statistics by this legally established intergovernmental organization, there is no other entity that is better positioned to offer valuable information on emission levels, about over 200 countries from 2000 to 2020, over 7670 days.

#### 4.1.2 Dataset Description

The "Total Emissions Per Country (2000-2020)" dataset on Kaggle is a collection of data on the total amount of greenhouse gas emissions produced by each country between the years 2000 and 2020. The dataset includes information on carbon dioxide (CO<sub>2</sub>), methane (CH<sub>4</sub>), and nitrous oxide (N<sub>2</sub>O) emissions, as well as the total greenhouse gas emissions. It contains 25 attributes and 58765 records.

The data is sourced from the World Bank and is presented in metric tons of carbon dioxide equivalent (MtCO<sub>2</sub>e). The dataset includes information on over 200 countries and territories, making it a comprehensive resource for studying global emissions patterns over the past two decades. It can be used for a variety of research purposes, including climate change analysis, emissions tracking, and policy development.

### 4.1.3 Data Processing

The dataset in a CSV format was initially loaded, and the first five rows were previewed for an initial grasp of the data. A new copy of the dataset was created to preserve the original dataset, and pre-processing was carried out on the new copy. In particular, the proportion of missing values to valid data was evaluated for each column, and due to the small ratio of missing data points, these data points were dropped for simplicity. Similarly, any duplicate data points were removed.

A careful examination of the dataset description revealed that the years in the dataset serve as the time series index of the emissions for each category. Thus, the primary processing logic involved the effective parsing of the categorical features to extract the optimal level of analysis that would facilitate the data's comprehension and processing. To accomplish this, the categorical features were analyzed to determine the total number of unique values in each feature, as well as the correlation between the different categorical features and emissions for each period.

The dataset had four categorical features, namely 'Area', 'Item', 'Element', and 'Unit'. However, only the first three were found to be useful since the 'Unit' column had identical values throughout the dataset. 'Area', representing the country or region for which data was collected, had 268 unique values. 'Item', representing the activity that generated the emissions, had 42 unique values, while 'Element', representing the type of emission, had 9 unique values.

Further exploration revealed that it would be more meaningful to classify the emissions by 'Element' for each country, as this approach yields a manageable number of categories.

### 4.1.4 Date Restructuring

The data underwent a transformation process where all the years were concatenated into one column instead of having separate columns for each year. Visual inspection of the first 500 entries of the stacked data revealed a pattern, indicating that some emission categories were subcategories of others, consistently adding up to form the super-category across different values in the 'Item' column. Furthermore, CO<sub>2</sub> equivalents from different elements were summed up to obtain the CO<sub>2</sub> equivalent for each 'Item'. Consequently, the data was parsed to address these observations. Sub-categories were summed up, and duplicate data were removed, resulting in only four unique items in the 'Element' column instead of the initial nine: 'Emissions (N<sub>2</sub>O)', 'Emissions (CO<sub>2</sub>eq) (AR5)', 'Emissions (CH<sub>4</sub>)', and 'Emissions (CO<sub>2</sub>)'.

To reduce the computational complexity of further pre-processing, only data for one country was utilized. Using a pivot table, the data was restructured, such that the unique values in the 'Element' column were separated into distinct columns, with the year column serving as the index of this new DataFrame structure. This configuration allowed for the filtering and organization of emission data for all years in the dataset, for any selected country, in columns based on the type of element emitted.

#### **4.1.5 Statistical Tests**

A statistical test, namely the Augmented Dickey-Fuller test, was performed on each Element of the dataset to verify that the mean and standard deviation of the data remained constant over time. The results of the test indicated that the data was non-stationary over time. To address this issue, a differencing operation was applied to the dataset which resulted in a significant improvement in the data's stationarity.

Furthermore, the Granger Causality test was conducted on the dataset to investigate whether any causal relationships exist between different emissions. The findings revealed that there is a causal effect of N<sub>2</sub>O emissions on CH<sub>4</sub> and CO<sub>2</sub> emissions.

#### **4.1.6 Train-Test Split**

The data was split into train and test sets with the first 18 years of data for training and the last 3 years for training

#### **4.1.7 Model Building and Evaluation**

Given the observed relationship between the Elements of the dataset via the Granger Causality test, it became necessary to employ a predictive model capable of accounting for the interdependence among features when making predictions for a specific feature. Therefore, the VARMAX (Vector Autoregressive Moving Average model with eXogenous variables) was utilized, as it satisfies this criterion and also offers interpretability relative to deep learning models, which tend to adopt a black-box approach to model construction. The VARMAX model was trained on the training dataset and tested for accuracy on the test dataset. The model's accuracy was assessed by utilizing it to forecast the final three years of data in the test set and calculating the root mean squared error (RMSE) between the test set and the model's predictions. The model's accuracy was deemed higher when the RMSE was lower. The VARMAX model was trained using VAR and VMA order values of 'p' and 'd.' To identify the most optimal model, hyperparameter tuning was conducted by training

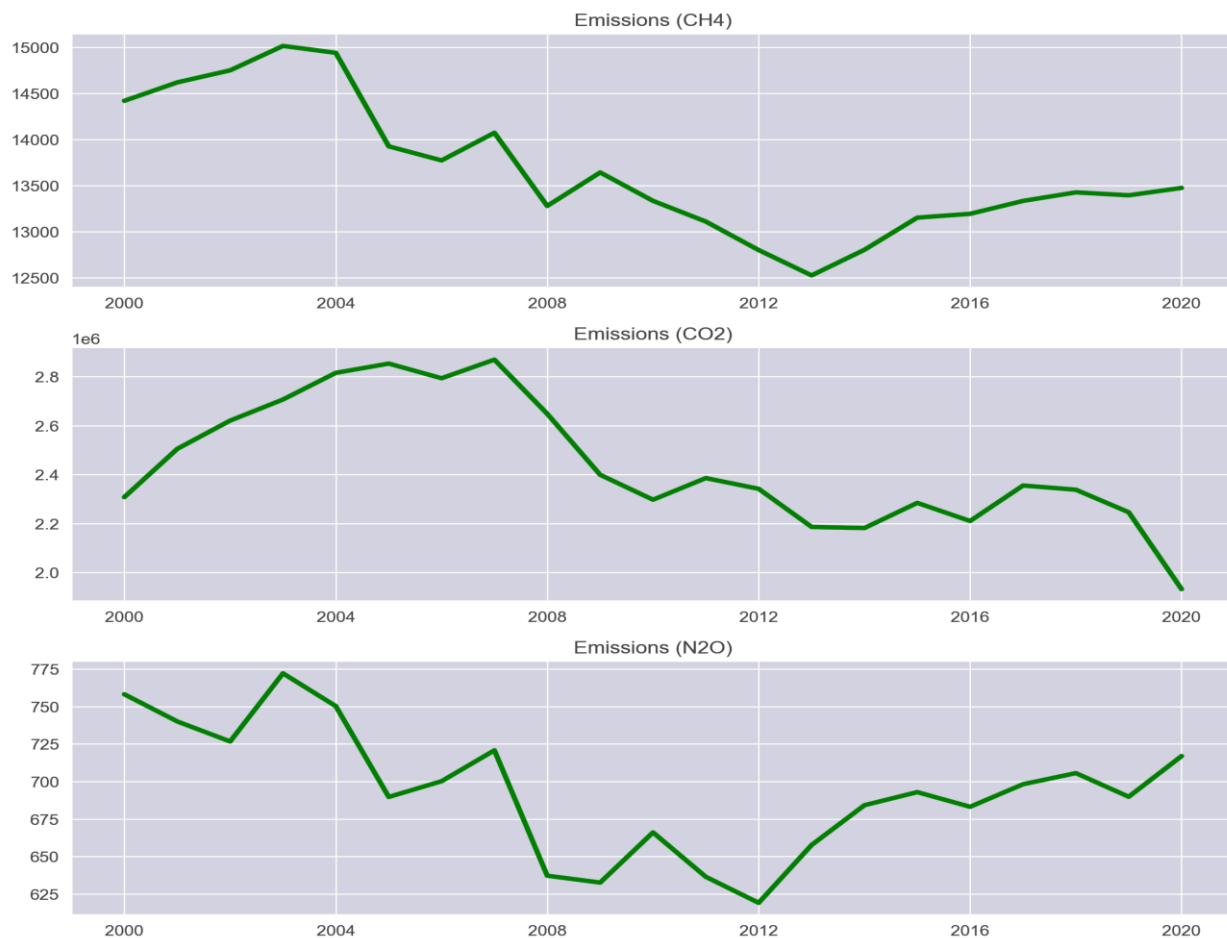
various VARMAX models with different 'p' and 'd' values, resulting in the model with the lowest RMSE score for all three Elements being predicted. It was concluded that if the model's RMSE is less than 10% of the mean emission value for that Element, then the model is deemed acceptable.

## 4.2 Emission patterns in Countries

The countries were picked based on the World Economic Situation and Prospects report by the United Nations. Three countries were picked from the developed economies, two were picked from the economies in transition and two were picked from developing economies.

Spain, Japan, United States, Belarus, Ukraine, Nigeria, Colombia

### 4.2.1 Emission patterns in Spain

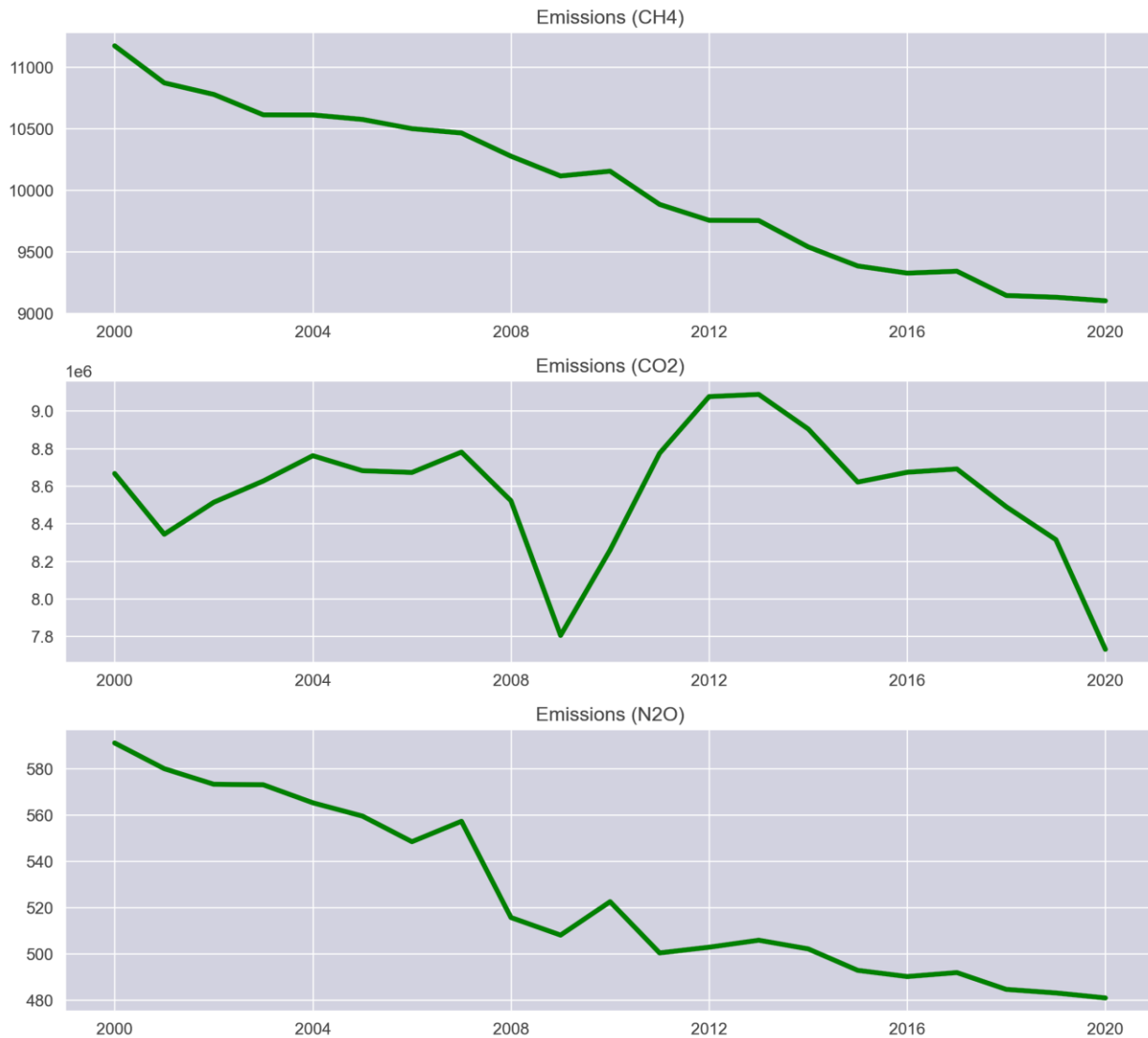


*Fig. 4.1 CH<sub>4</sub>, CO<sub>2</sub> and N<sub>2</sub>O Emissions for Spain*

Carbon dioxide (CO<sub>2</sub>) has been found to be the most prevalent greenhouse gas emission across many countries. As illustrated in Figure 4.1, the peak of carbon dioxide emissions occurred in 2007, and there has been a downward trend since then. Similarly, methane (CH<sub>4</sub>) emissions have

also been decreasing over time, but from 2012, there has been an observed increase. Nitrous oxide ( $\text{N}_2\text{O}$ ) emissions have also followed a comparable pattern with that of  $\text{CH}_4$ , but at a relatively lower magnitude.

#### 4.2.2 Emission patterns in Japan



*Fig. 4.2 CH<sub>4</sub>, CO<sub>2</sub> and N<sub>2</sub>O Emissions for Japan*

According to the information presented in Figure 4.2, there has been a declining trend in the emissions of methane ( $\text{CH}_4$ ) and nitrous oxide ( $\text{N}_2\text{O}$ ) in Japan. However, the emission of carbon dioxide ( $\text{CO}_2$ ) decreased in 2009, followed by an increase up to its peak in 2013, and has since been gradually decreasing.

### 4.2.3 Emission patterns in United States



*Fig. 4.3 CH<sub>4</sub>, CO<sub>2</sub> and N<sub>2</sub>O Emissions for United States*

The COVID-19 pandemic has had a significant impact on the emission levels of greenhouse gases in the United States. Specifically, methane (CH<sub>4</sub>) emissions, which had been consistently high since the early 2000s, experienced a marked decline in the 2010s. Carbon dioxide (CO<sub>2</sub>) emissions, which also started at high levels, showed a general trend of decline, but experienced a significant decrease in 2020 due to the pandemic. Nitrous oxide (N<sub>2</sub>O) emissions, on the other hand, exhibited a low and fluctuating pattern, but also experienced a reduction in 2020. These findings have important implications for understanding the dynamics of greenhouse gas emissions and the impact of global events on their levels.

#### 4.2.4 Emission patterns in Belarus



*Fig. 4.4 CH<sub>4</sub>, CO<sub>2</sub> and N<sub>2</sub>O Emissions for Belarus*

A notable increase in methane (CH<sub>4</sub>), carbon dioxide (CO<sub>2</sub>), and nitrous oxide (N<sub>2</sub>O) emissions has been observed from the year 2010, and these emissions levels have not shown a significant reduction even during the COVID-19 pandemic in 2020. This observed pattern of sustained emissions increase despite external factors such as the pandemic may be attributed to the country's status as an economy in transition.



### 4.2.5 Emission patterns in Ukraine



*Fig. 4.5 CH<sub>4</sub>, CO<sub>2</sub> and N<sub>2</sub>O Emissions for Ukraine*

The data indicates a downward trend in CH<sub>4</sub> and CO<sub>2</sub> emissions over time, while N<sub>2</sub>O emissions demonstrate a consistent upward trend. The decline in CH<sub>4</sub> and CO<sub>2</sub> emissions can be attributed to a range of factors including the disintegration of the Soviet Union, the global financial downturn of 2008, and domestic conflicts between 2014 and 2015 (Belousova, 2023).

#### 4.2.6 Emission patterns in Nigeria



*Fig. 4.6 CH<sub>4</sub>, CO<sub>2</sub> and N<sub>2</sub>O Emissions for Nigeria*

The trend analysis of the emission data indicates that while CH<sub>4</sub> emissions remained relatively stable over the observed period, both N<sub>2</sub>O and CO<sub>2</sub> emissions displayed an increasing trend. In the year 2009, all three emissions experienced a decline, but the reduction in N<sub>2</sub>O emission was comparatively lower than the other two. Subsequently, the emissions resumed an upward trend.

#### 4.2.7 Emission patterns in Colombia



*Fig. 4.7 CH<sub>4</sub>, CO<sub>2</sub> and N<sub>2</sub>O Emissions for Colombia*

The emissions of CH<sub>4</sub>, CO<sub>2</sub>, and N<sub>2</sub>O have exhibited an overall increasing trend. Nonetheless, between 2012 and 2016, all three emissions experienced a decline, with CO<sub>2</sub> emissions experiencing the most significant reduction. However, this trend was short-lived as all three emissions increased again after 2016, exceeding their previous levels.

## 4.3 Prediction of Emission in Countries

### 4.3.1 Prediction of Emission in Spain

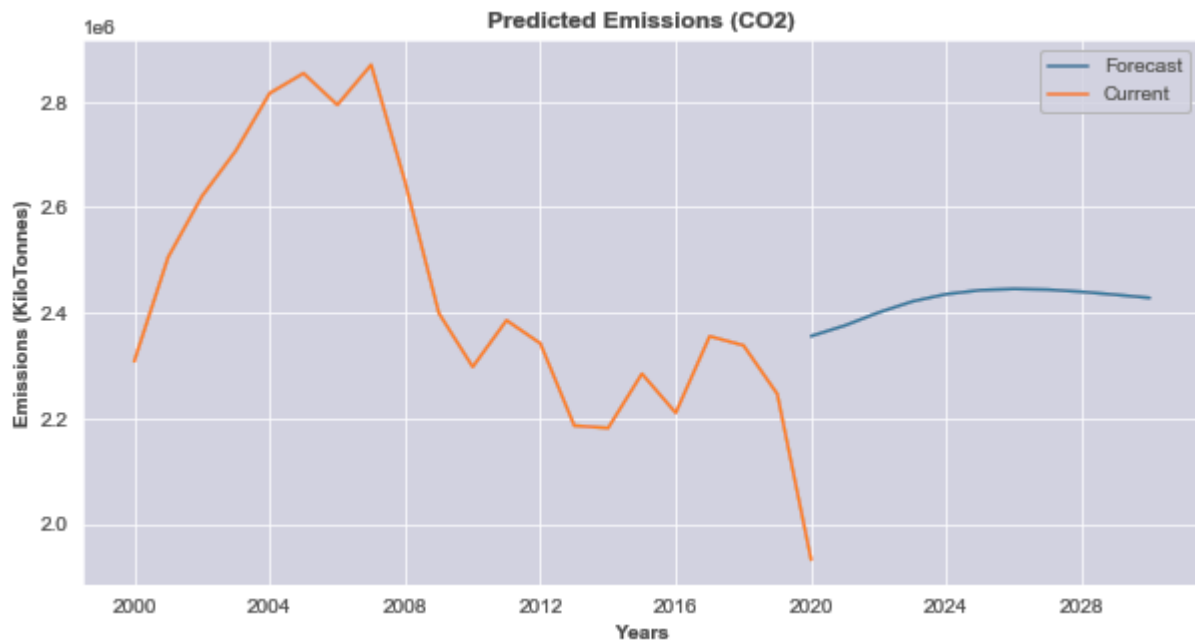


Fig. 4.8 Predicted Emissions for CO<sub>2</sub> in Spain

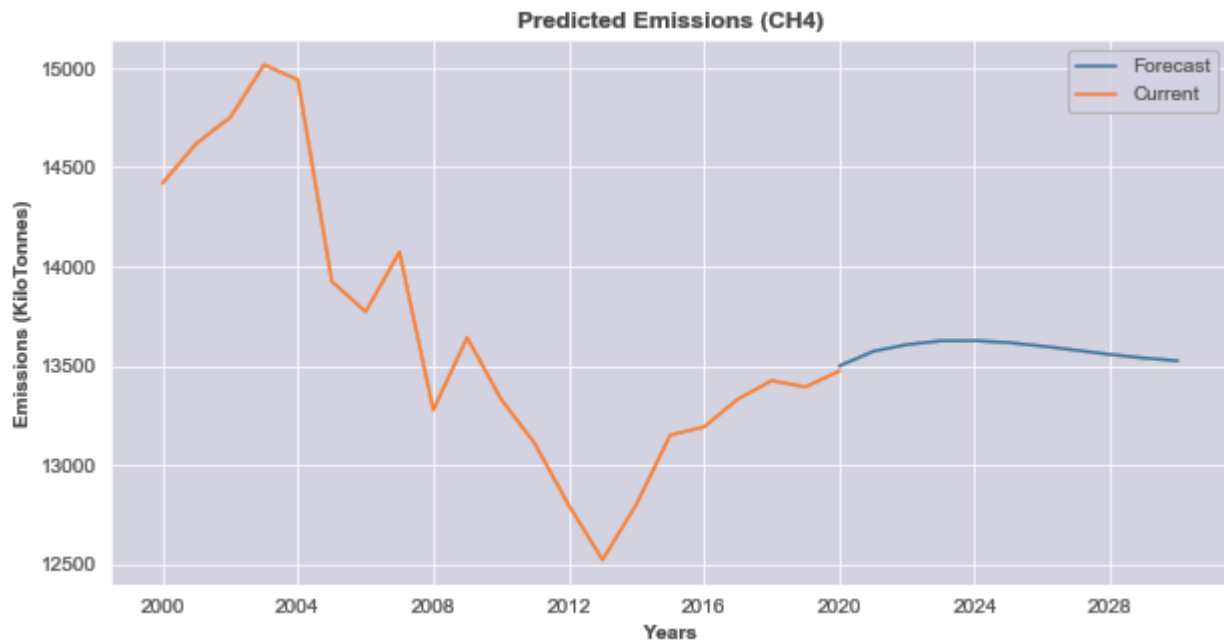


Fig. 4.9 Predicted Emissions for CH<sub>4</sub> in Spain

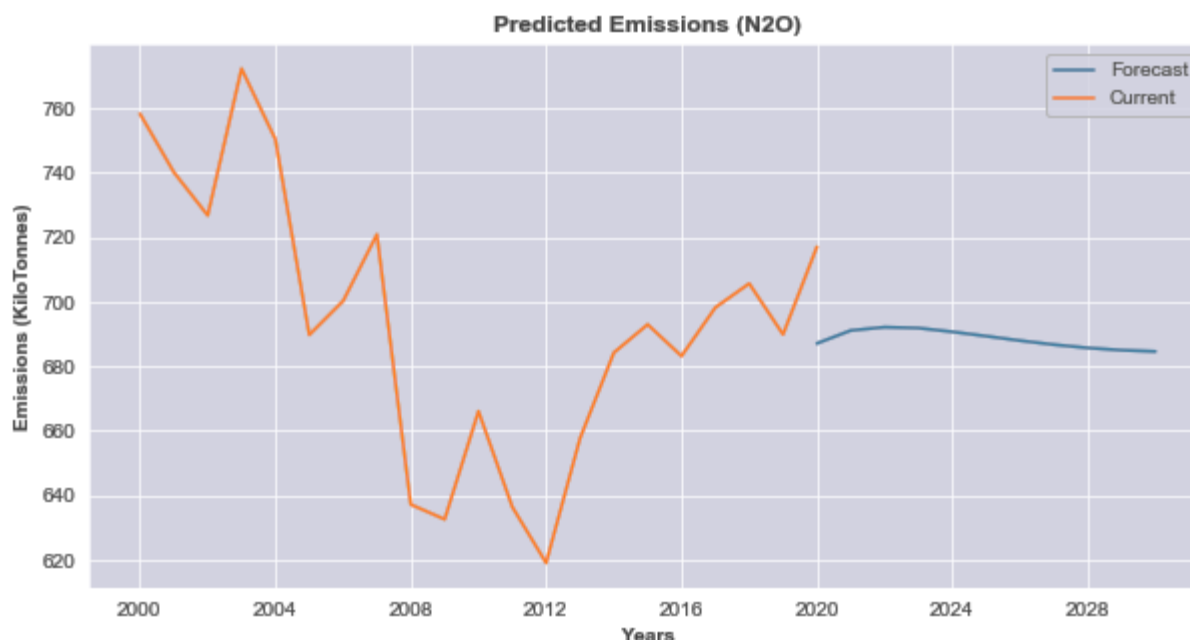


Fig. 4.10 Predicted Emissions for N<sub>2</sub>O in Spain

According to the predictions, there will be a steady increase in emissions of CH<sub>4</sub> and CO<sub>2</sub> over the coming years, while N<sub>2</sub>O emissions will remain relatively stable. Specifically, CH<sub>4</sub> emissions are predicted to increase from 13,498.45 Gg in 2020 to 13,524.75 Gg in 2030. Meanwhile, CO<sub>2</sub> emissions are expected to increase from 2,355,785 Gg in 2020 to 2,428,219 Gg in 2030. On the other hand, N<sub>2</sub>O emissions are predicted to only slightly increase from 687.04 Gg in 2020 to 684.56 Gg in 2030.

It is important to note that CO<sub>2</sub> emissions are the most prevalent greenhouse gas emissions across many countries and have been decreasing since their peak in 2007. Conversely, CH<sub>4</sub> emissions have been decreasing over time but have shown an increase since 2012, while N<sub>2</sub>O emissions have followed a comparable pattern with that of CH<sub>4</sub>, but at a relatively lower magnitude.

### 4.3.2 Prediction of Emission in Japan

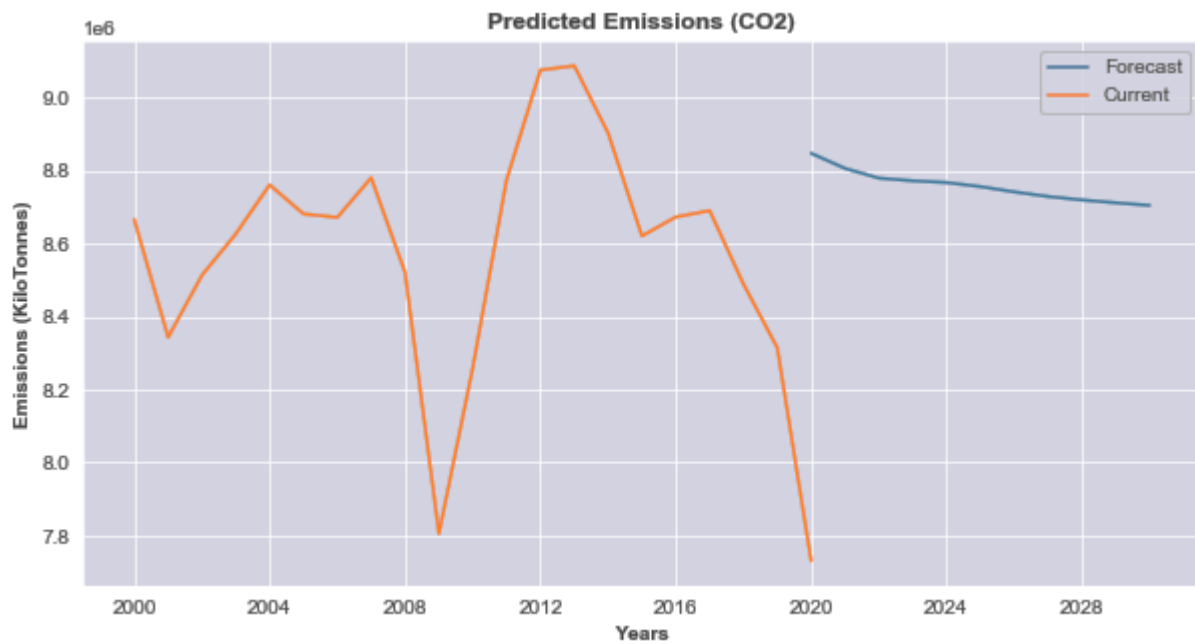


Fig. 4.11 Predicted Emissions for CO<sub>2</sub> in Japan

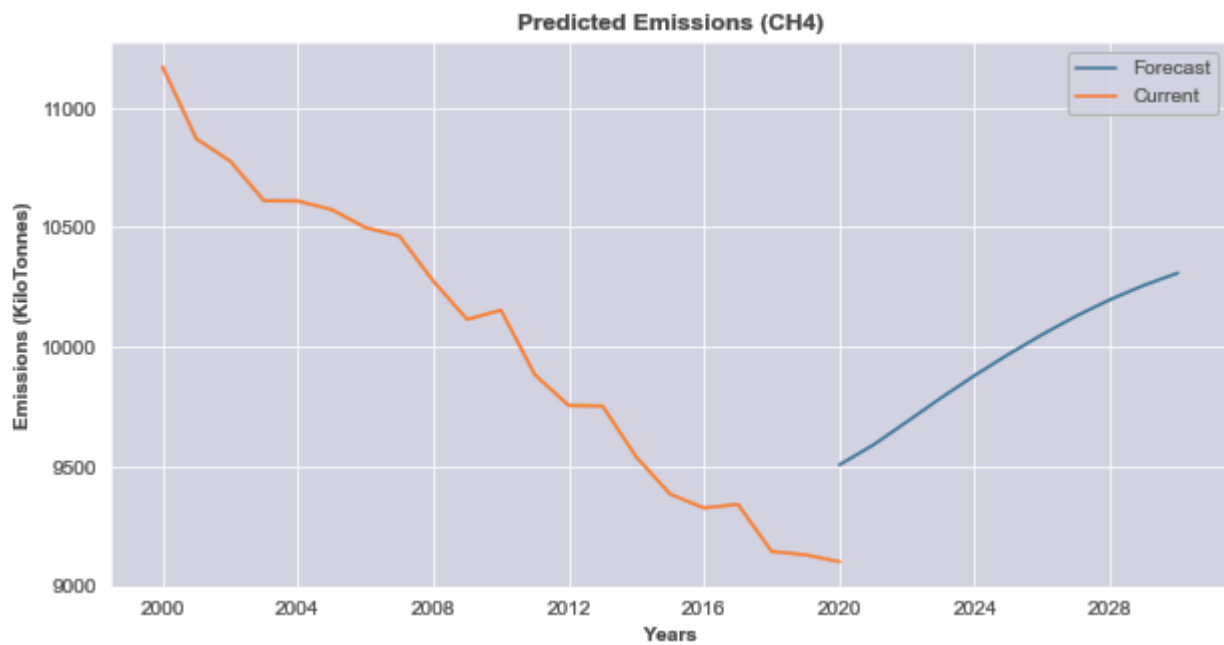


Fig. 4.12 Predicted Emissions for CH<sub>4</sub> in Japan

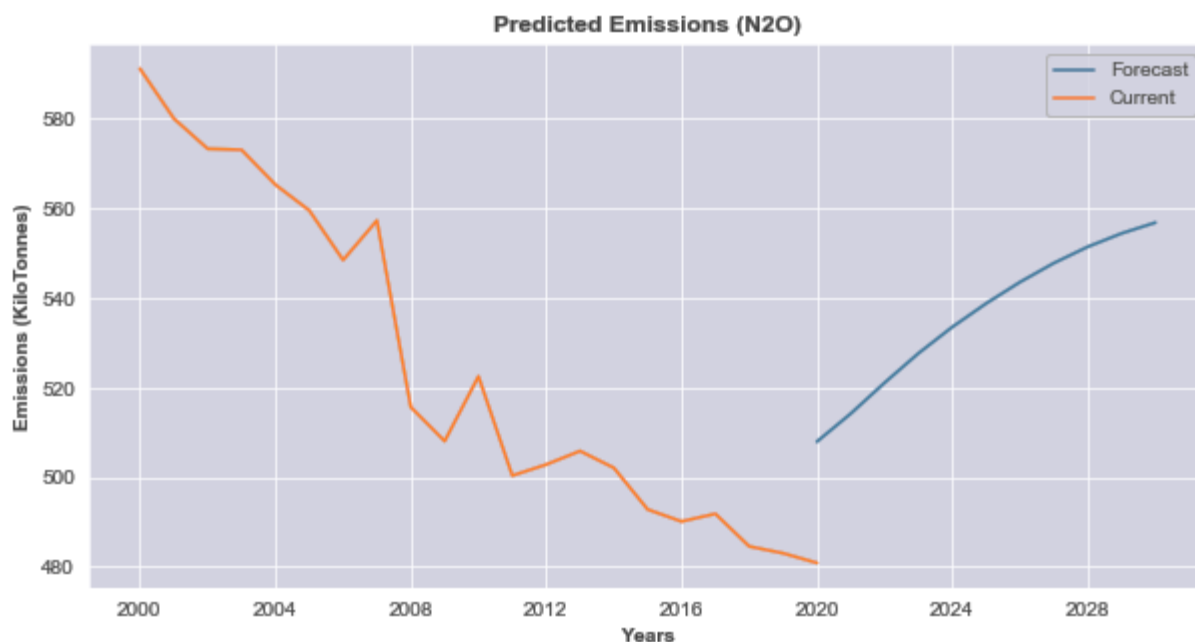


Fig. 4.13 Predicted Emissions for N<sub>2</sub>O in Japan

According to the predictions, there will be a steady increase in emissions of CH<sub>4</sub> and CO<sub>2</sub> over the coming years, while N<sub>2</sub>O emissions will remain relatively stable. Specifically, CH<sub>4</sub> emissions are predicted to increase from 9,507.02 Gg in 2020 to 10,308.95 Gg in 2030. Meanwhile, CO<sub>2</sub> emissions are expected to increase from 8,847,428 Gg in 2020 to 8,704,751 Gg in 2030. On the other hand, N<sub>2</sub>O emissions are predicted to only slightly increase from 507.93 Gg in 2020 to 556.84 Gg in 2030.

### 4.3.3 Prediction of Emission in United States

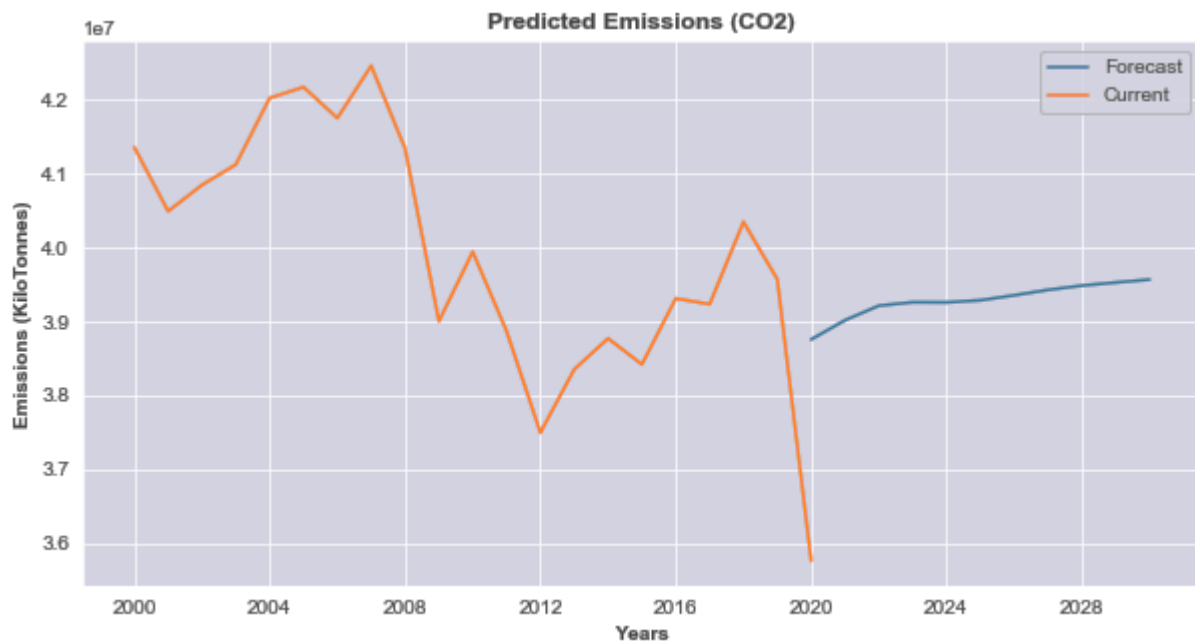


Fig. 4.14 Predicted Emissions for CO<sub>2</sub> in United States

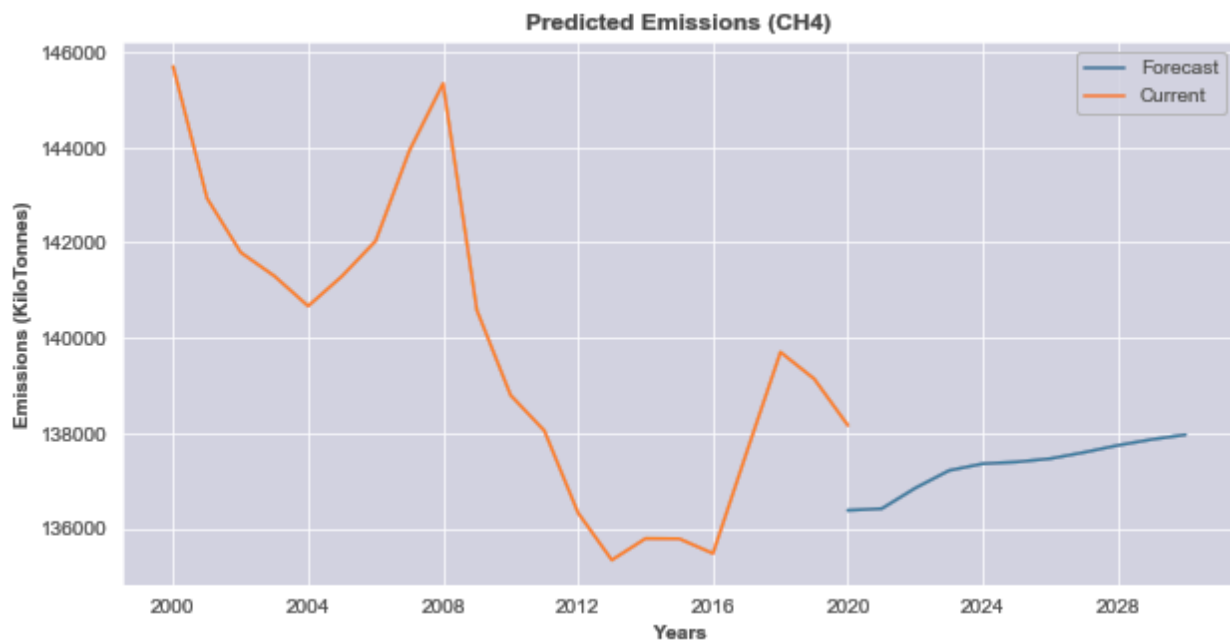
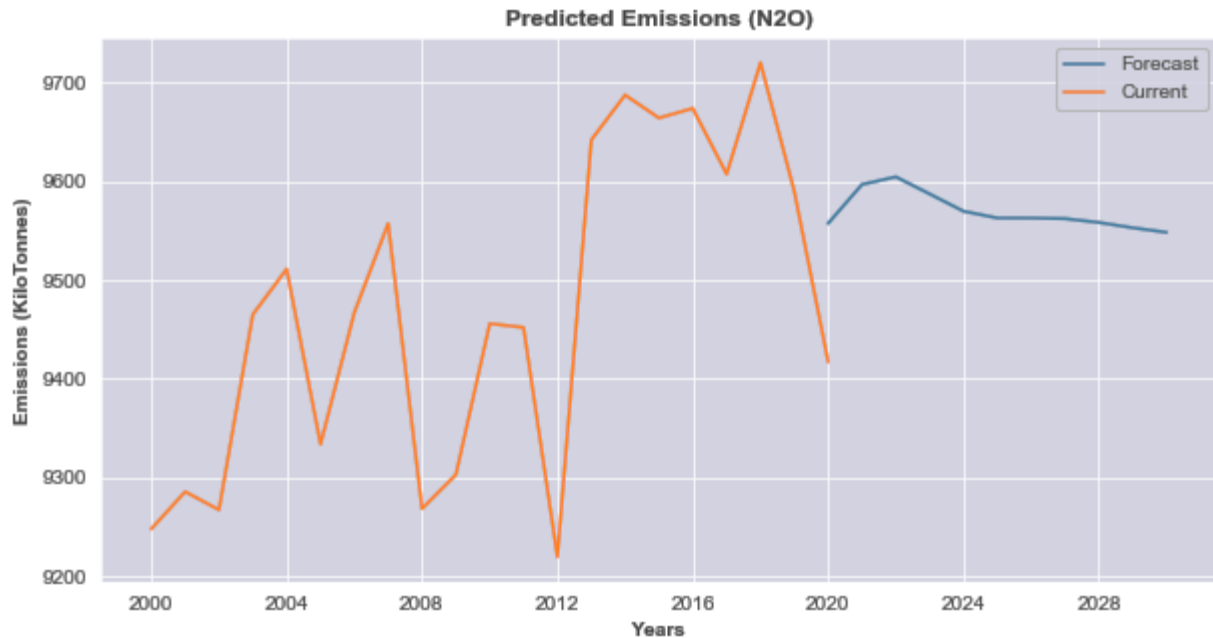


Fig. 4.15 Predicted Emissions for CH<sub>4</sub> in United States





*Fig. 4.16 Predicted Emissions for N<sub>2</sub>O in United States*

According to the given predictions, the emissions of CH<sub>4</sub> and CO<sub>2</sub> are expected to steadily increase in the upcoming years, while the N<sub>2</sub>O emissions will remain relatively stable. Specifically, the CH<sub>4</sub> emissions are predicted to rise from 136,376.54 Gg in 2020 to 137,962.22 Gg in 2030. Similarly, the CO<sub>2</sub> emissions are expected to increase from 38,751,303.65 Gg in 2020 to 39,561,435.44 Gg in 2030. However, the N<sub>2</sub>O emissions are only predicted to slightly reduce from 9,557.20 Gg in 2020 to 9,548.13 Gg in 2030.

#### 4.3.4 Prediction of Emission in Belarus

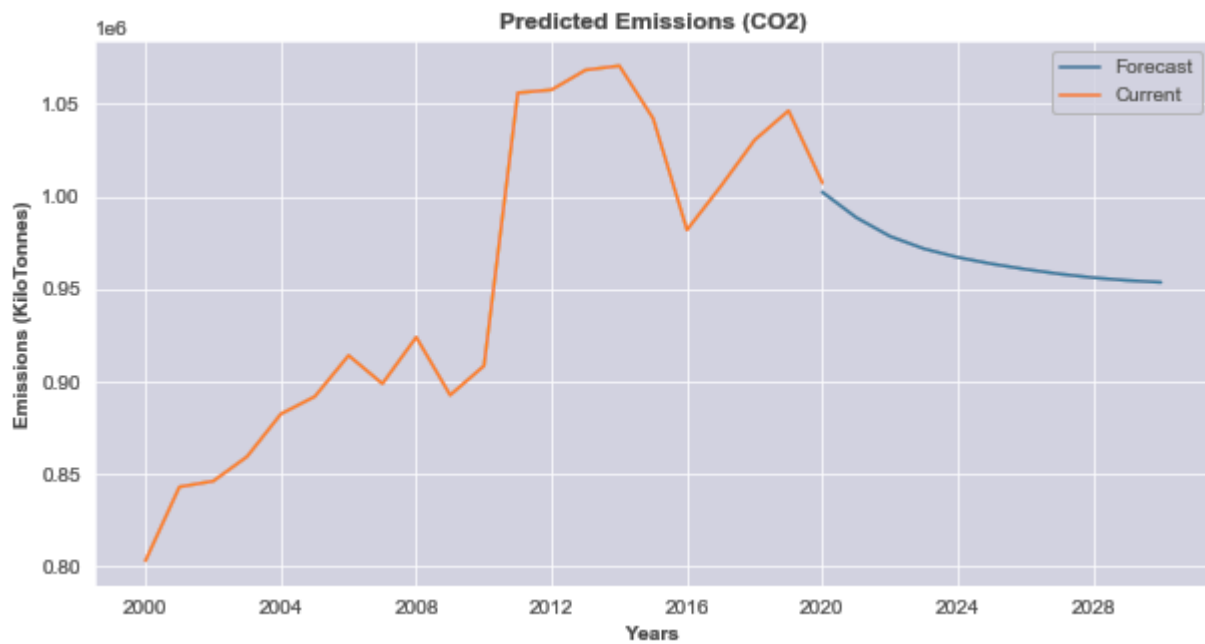


Fig. 4.17 Predicted Emissions for CO<sub>2</sub> in Belarus

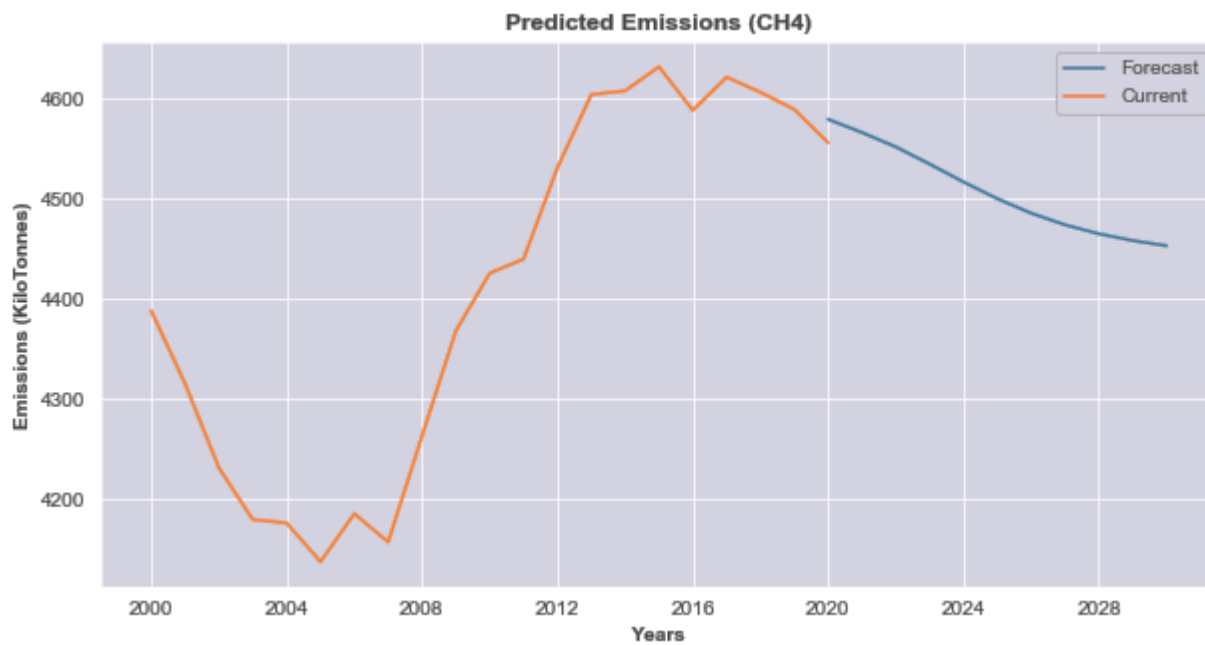
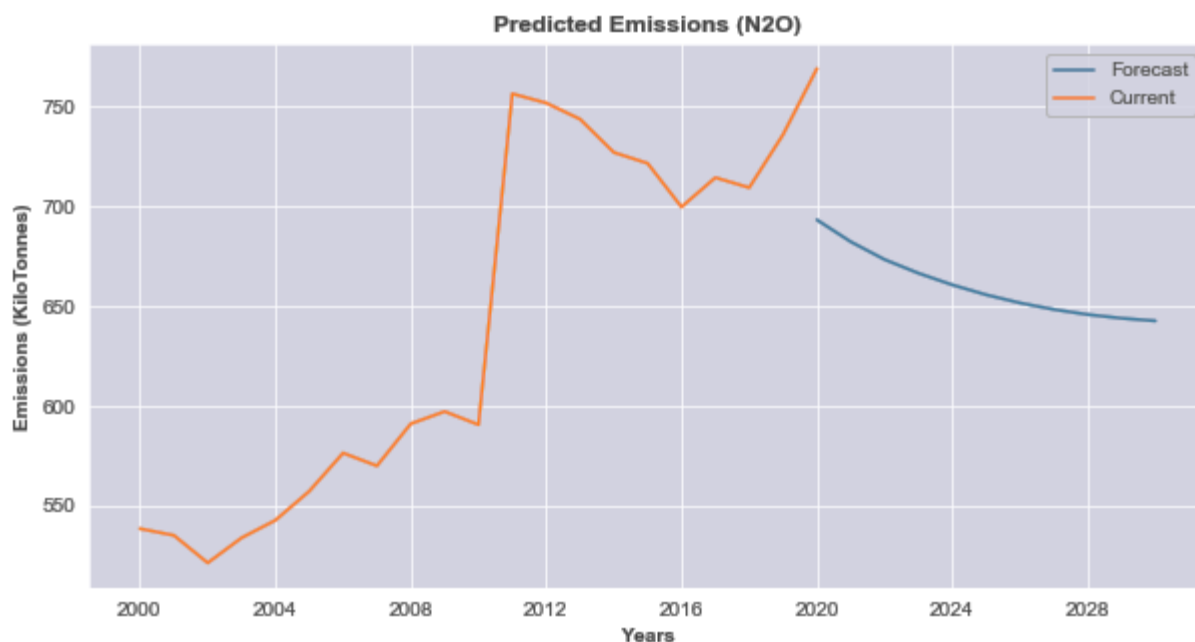


Fig. 4.18 Predicted Emissions for CH<sub>4</sub> in Belarus



*Fig. 4.19 Predicted Emissions for N<sub>2</sub>O in Belarus*

According to the given predictions, the emissions of CH<sub>4</sub> are expected to steadily decrease in the upcoming years, while the CO<sub>2</sub> emissions will remain relatively stable. Specifically, the CH<sub>4</sub> emissions are predicted to decrease from 4579.107946 Gg in 2020 to 4452.871106 Gg in 2030. Similarly, the CO<sub>2</sub> emissions are expected to decrease from 1.002316e+06 Gg in 2020 to 9.535096e+05 Gg in 2030. However, the N<sub>2</sub>O emissions are only predicted to slightly reduce from 693.219728 Gg in 2020 to 642.481489 Gg in 2030.

4.3.5 Prediction of Emission in Ukraine

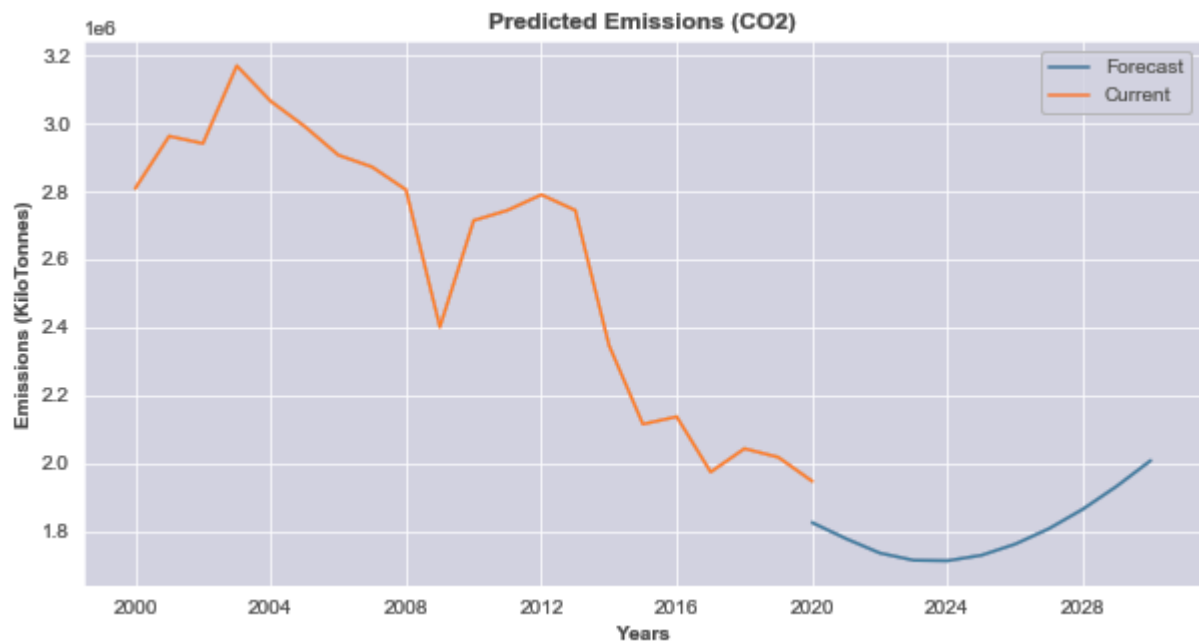


Fig. 4.20 Predicted Emissions for CO<sub>2</sub> in Ukraine

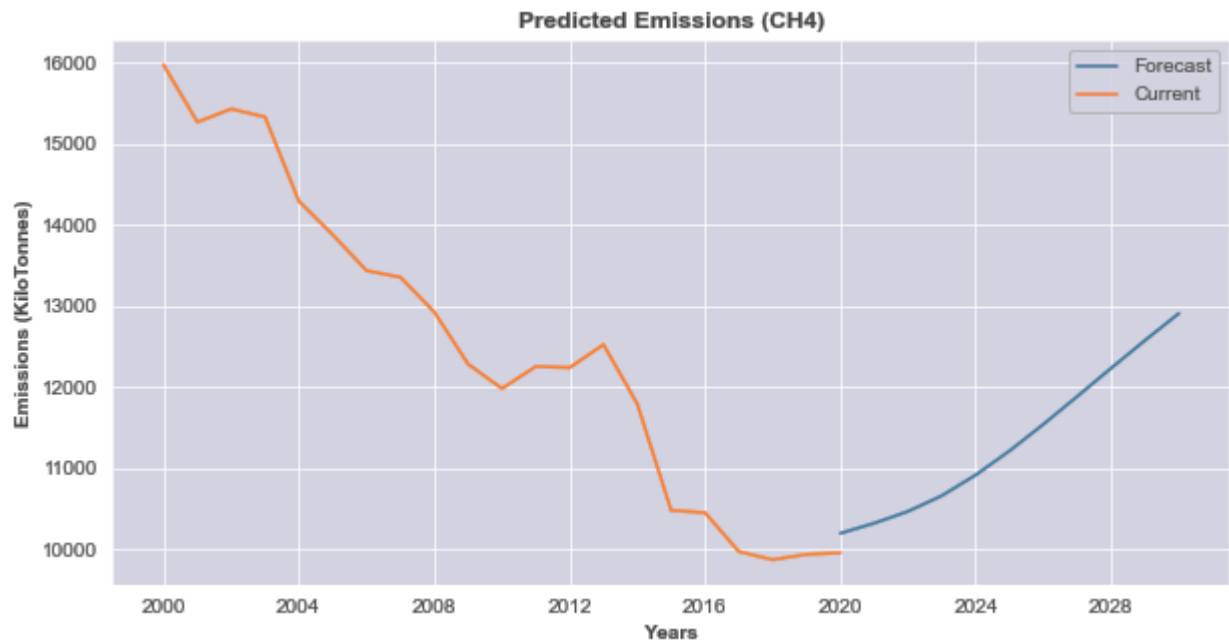


Fig. 4.21 Predicted Emissions for CH<sub>4</sub> in Ukraine

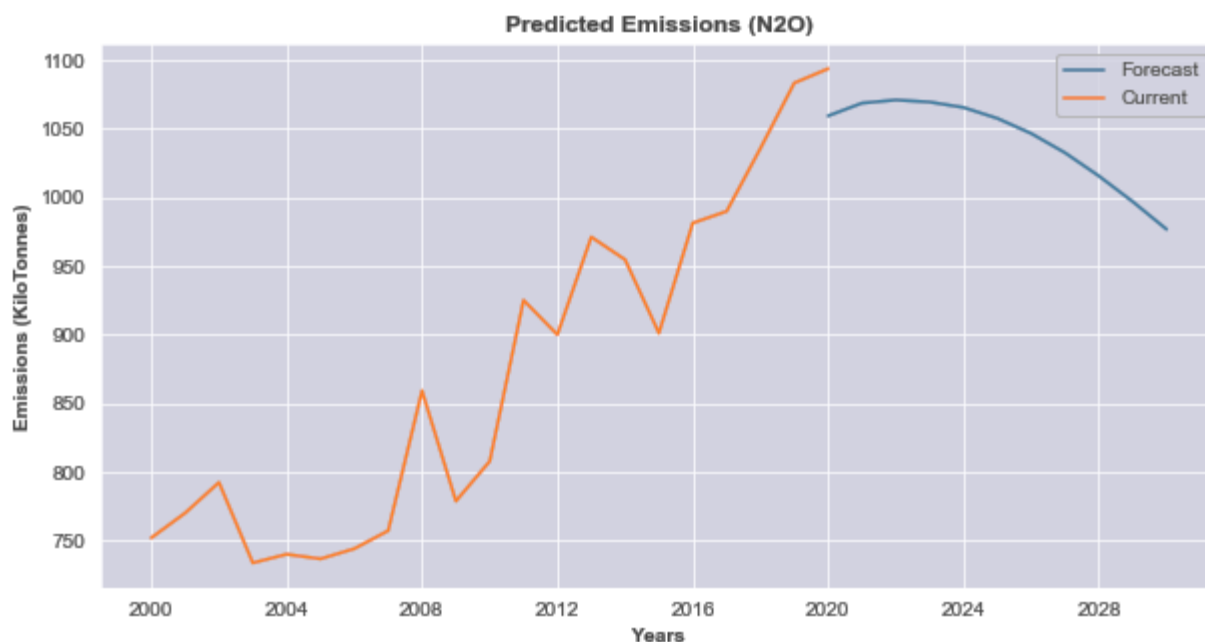


Fig. 4.22 Predicted Emissions for N<sub>2</sub>O in Ukraine

According to the given predictions, the emissions of CH<sub>4</sub> are expected to steadily increase in the upcoming years, while the CO<sub>2</sub> emissions will also increase. Specifically, the CH<sub>4</sub> emissions are predicted to increase from 10194.515239 Gg in 2020 to 12905.881501 Gg in 2030. Similarly, the CO<sub>2</sub> emissions are expected to increase from 1.824629e+06 Gg in 2020 to 2.006878e+06 Gg in 2030. However, the N<sub>2</sub>O emissions are only predicted to slightly reduce from 1059.149865 Gg in 2020 to 976.604660 Gg in 2030.

#### 4.3.6 Prediction of Emission in Nigeria

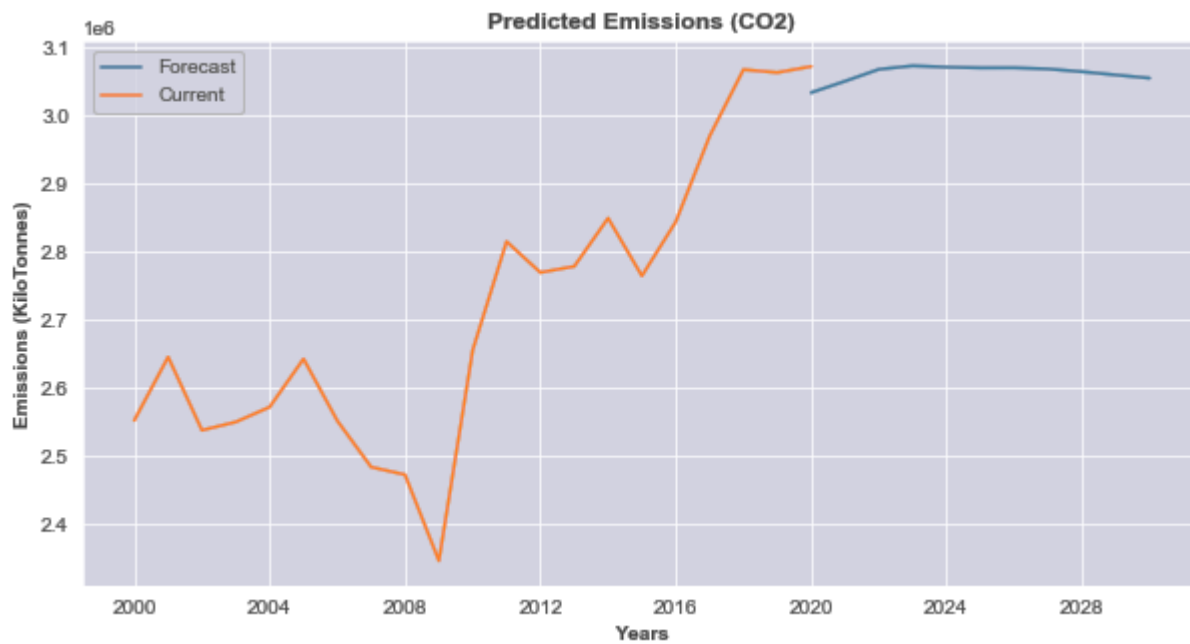


Fig. 4.23 Predicted Emissions for CO<sub>2</sub> in Nigeria

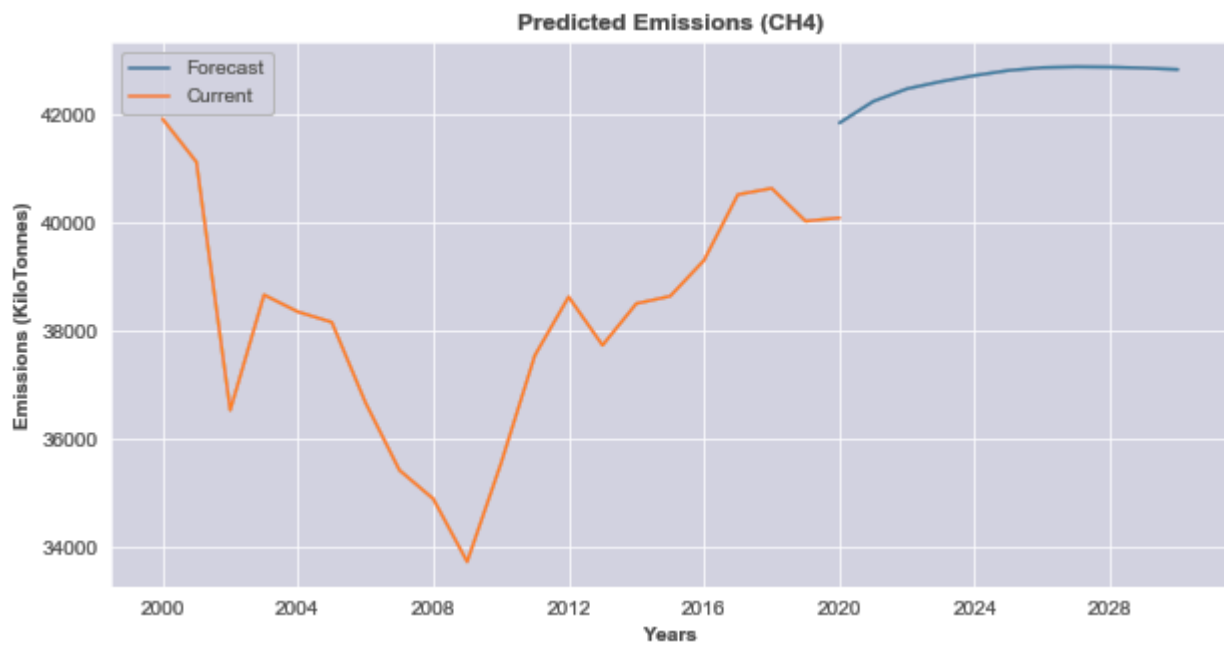


Fig. 4.24 Predicted Emissions for CH<sub>4</sub> in Nigeria

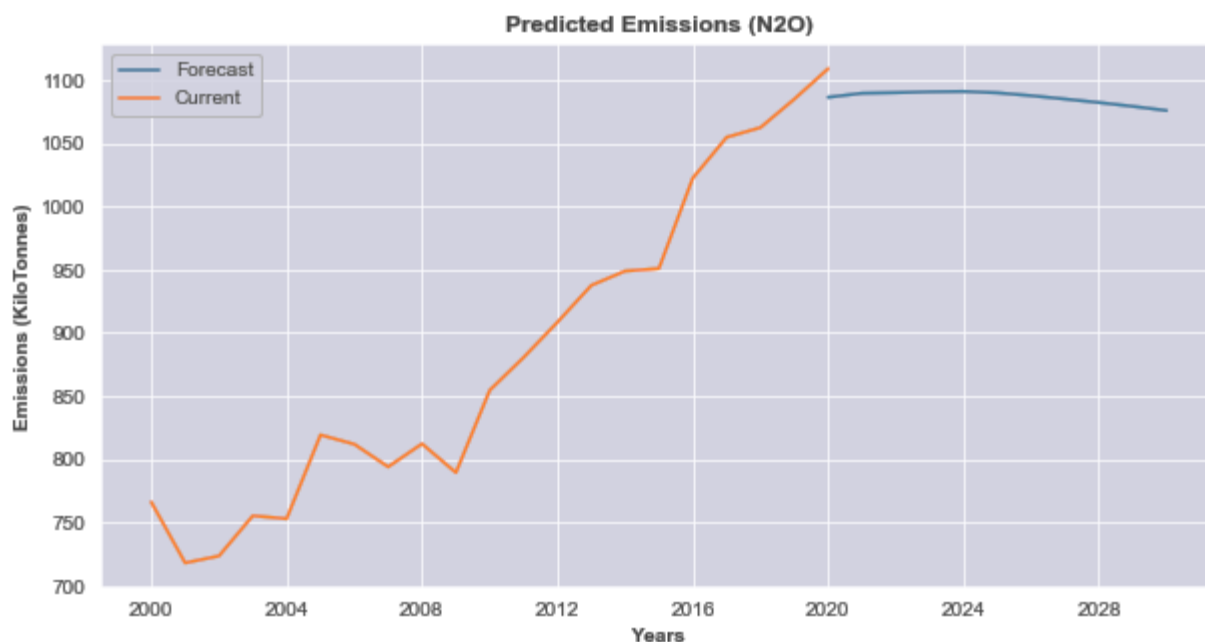


Fig. 4.25 Predicted Emissions for  $N_2O$  in Nigeria

According to the given predictions, the emissions of  $CH_4$  are expected to steadily increase in the upcoming years, while the  $CO_2$  emissions will remain relatively stable. Specifically, the  $CH_4$  emissions are predicted to increase from 41840.642523 Gg in 2020 to 42829.154240 Gg in 2030. Similarly, the  $CO_2$  emissions are expected to increase from 3.033347e+06 Gg in 2020 to 3.054608e+06 Gg in 2030. However, the  $N_2O$  emissions are only predicted to slightly reduce from 1086.221344 Gg in 2020 to 1075.731444 Gg in 2030.

4.3.7 Prediction of Emission in Colombia

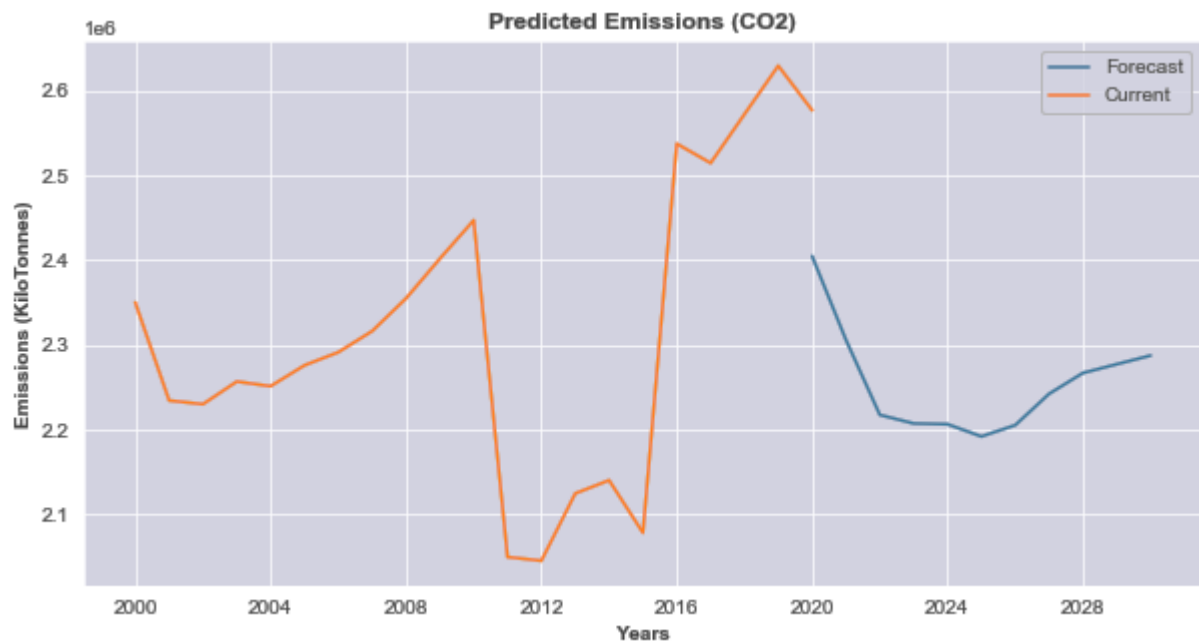


Fig. 4.26 Predicted Emissions for CO<sub>2</sub> in Colombia

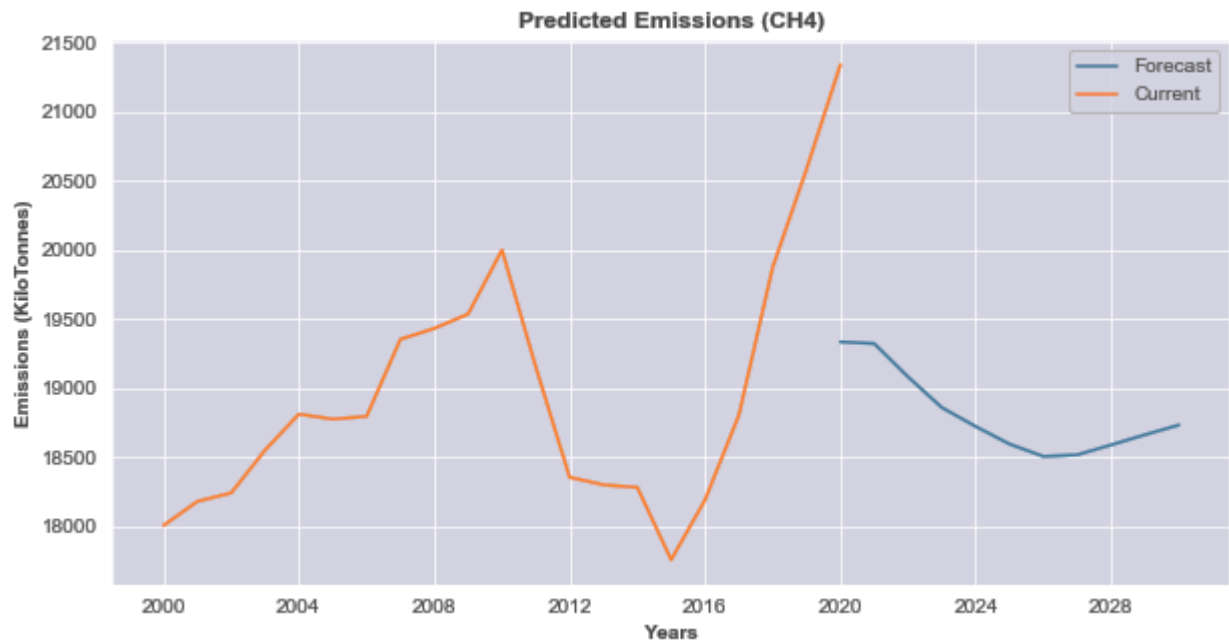


Fig. 4.27 Predicted Emissions for CH<sub>4</sub> in Colombia



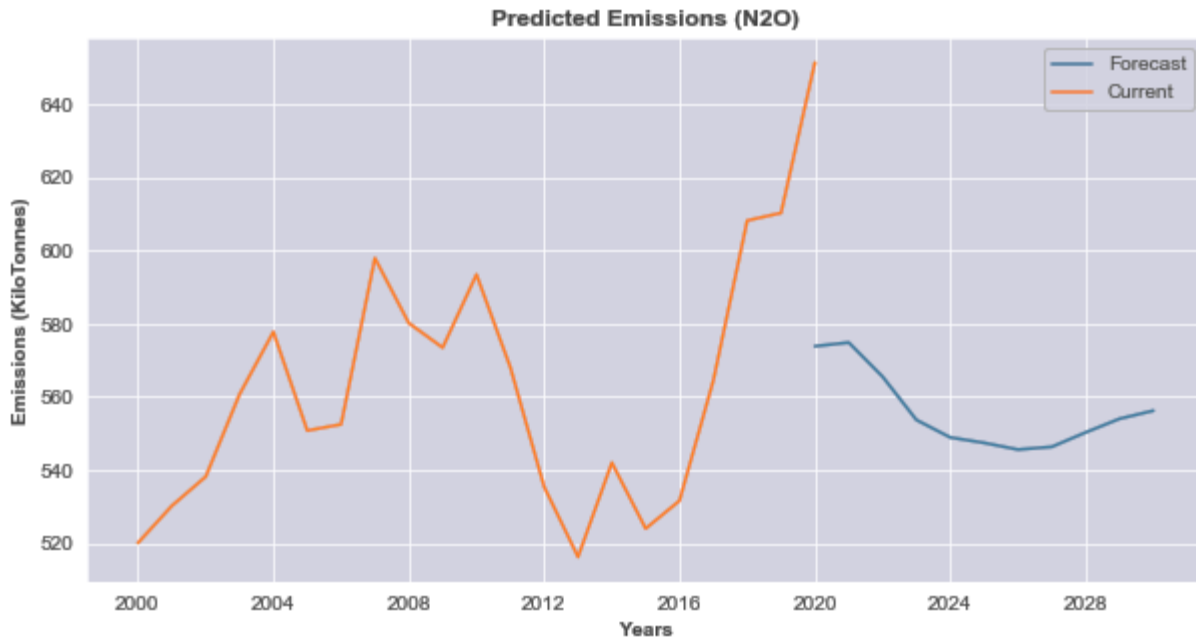


Fig. 4.28 Predicted Emissions for N<sub>2</sub>O in Colombia

According to the given predictions, there would be a drop in all emissions from 2020. The emissions of CH<sub>4</sub> are expected to steadily decrease in the upcoming years, while the CO<sub>2</sub> emissions will remain relatively stable. Specifically, the CH<sub>4</sub> emissions are predicted to decrease from 19334.172608 Gg in 2020 to 18732.354228 Gg in 2030. Similarly, the CO<sub>2</sub> emissions are expected to decrease from 2.404636e+06 Gg in 2020 to 2.287395e+06 Gg in 2030. However, the N<sub>2</sub>O emissions are only predicted to slightly reduce from 573.839110 Gg in 2020 to 556.221861 Gg in 2030.

## 4.4 Statistical test

### 4.4.1 Statistical test for Spain

#### 4.4.1.1 Table 1. Augmented Dickey-Fuller (ADF) test for Spain

Augmented Dickey-Fuller Test: Emissions (CH<sub>4</sub>)

Test Statistic	-4.510830
p-value	0.000188
No. of lags Used	0.000000
No. of obsv used	19.000000
Confidence Interval (1%)	-3.832603
Confidence Interval (5%)	-3.031227
Confidence Interval (10%)	-2.655520
dtype:	float64

-----

#### Augumented Dickey-Fuller Test: Emissions (CO<sub>2</sub>)

```
Test Statistic      -5.775266e+00
p-value             5.277450e-07
No. of lags Used    8.000000e+00
No. of obsv used    1.100000e+01
Confidence Interval (1%) -4.223238e+00
Confidence Interval (5%) -3.189369e+00
Confidence Interval (10%) -2.729839e+00
dtype: float64
```

#### Augumented Dickey-Fuller Test: Emissions (N<sub>2</sub>O)

```
Test Statistic      -5.987381e+00
p-value             1.778847e-07
No. of lags Used    8.000000e+00
No. of obsv used    1.100000e+01
Confidence Interval (1%) -4.223238e+00
Confidence Interval (5%) -3.189369e+00
Confidence Interval (10%) -2.729839e+00
dtype: float64
```

In the case of CH<sub>4</sub>, the Test Statistic is -4.510830 and the p-value is 0.000188, which indicates strong evidence against the null hypothesis of non-stationarity. Thus, we can conclude that the CH<sub>4</sub> emissions time series is stationary.

Similarly, for CO<sub>2</sub> and N<sub>2</sub>O, the Test Statistics are -5.775266e+00 and -5.987381e+00, respectively, with p-values of 5.277450e-07 and 1.778847e-07, respectively, indicating strong evidence against the null hypothesis of non-stationarity for both gases. Therefore, we can conclude that the CO<sub>2</sub> and N<sub>2</sub>O emissions time series are stationary.

#### 4.4.1.2 *Table 2. Granger Causality test for spain*

Causality Matrix (1 Time Lag):

	Emissions (CH <sub>4</sub> )	Emissions (CO <sub>2</sub> )	Emissions (N <sub>2</sub> O)
Emissions (CH <sub>4</sub> )	NaN	0.016327	0.928547
Emissions (CO <sub>2</sub> )	0.192337	NaN	0.25371
Emissions (N <sub>2</sub> O)	0.247709	0.01234	NaN

Causality Matrix (2 Time Lags):

	Emissions (CH <sub>4</sub> )	Emissions (CO <sub>2</sub> )	Emissions (N <sub>2</sub> O)
Emissions (CH <sub>4</sub> )	NaN	0.387927	0.968471
Emissions (CO <sub>2</sub> )	0.509064	NaN	0.361299
Emissions (N <sub>2</sub> O)	0.551419	0.291193	NaN

Based on Granger Causality tests conducted at lag order 1 and 2, it was found that Emissions of Methane (CH<sub>4</sub>) Granger cause Emissions of Carbon Dioxide (CO<sub>2</sub>) at a significance level of  $\alpha = 0.05$ , with p-values of 0.0101 and 0.0818, respectively. This implies that past values of CH<sub>4</sub> emissions have predictive power in explaining the current and future values of CO<sub>2</sub> emissions.

However, there is no evidence of Granger Causality between Emissions of Methane (CH<sub>4</sub>) and Emissions of Nitrous Oxide (N<sub>2</sub>O), as the p-values are greater than the significance level. This suggests that past values of CH<sub>4</sub> emissions do not have a statistically significant influence on N<sub>2</sub>O emissions.

Similarly, there is no evidence of Granger Causality between Emissions of Carbon Dioxide (CO<sub>2</sub>) and Emissions of Methane (CH<sub>4</sub>) or Emissions of Carbon Dioxide (CO<sub>2</sub>) and Emissions of Nitrous Oxide (N<sub>2</sub>O), as the p-values are also greater than the significance level. This indicates that past values of CO<sub>2</sub> emissions do not have a statistically significant influence on CH<sub>4</sub> or N<sub>2</sub>O emissions.

#### 4.4.1.3 Model Performance

```
Mean val Emissions (CH4): 13430.2807 -- RMSE: 86.5077 -- RMSE < 10% Mean: True
Mean val Emissions (CO2): 2172171.3978 -- RMSE: 253907.996 -- RMSE < 10% Mean: False
Mean val Emissions (N2O): 704.1294 -- RMSE: 22.1128 -- RMSE < 10% Mean: True
```

For Emissions (CH<sub>4</sub>) and Emissions (N<sub>2</sub>O), the model's predictions are accurate, as the RMSE is less than 10% of the mean value, indicated by the "True" value in the "RMSE < 10% Mean" column. For Emissions (CO<sub>2</sub>), however, the RMSE is greater than 10% of the mean value, indicated by the "False" value in the "RMSE < 10% Mean" column, suggesting that the model's predictions for Emissions (CO<sub>2</sub>) are not as accurate as for the other two variables. The mean value of Emissions (CH<sub>4</sub>) is also reported as 13,430.2807 and the mean value of Emissions (CO<sub>2</sub>) is reported as 2,172,171.3978, while the mean value of Emissions (N<sub>2</sub>O) is reported as 704.1294.

### 4.4.2 Statistical test for Japan

#### 4.4.2.1 *Table 3. Augmented Dickey-Fuller (ADF) test for Japan*

Augmented Dickey-Fuller Test: Emissions (CH<sub>4</sub>)

Test Statistic	-1.804023
p-value	0.378466
No. of lags Used	3.000000
No. of obsv used	16.000000
Confidence Interval (1%)	-3.924019
Confidence Interval (5%)	-3.068498
Confidence Interval (10%)	-2.673893
dtype: float64	

Augmented Dickey-Fuller Test: Emissions (CO<sub>2</sub>)

Test Statistic	-3.022048
p-value	0.032886
No. of lags Used	0.000000

```

No. of obsv used          19.000000
Confidence Interval (1%)  -3.832603
Confidence Interval (5%)  -3.031227
Confidence Interval (10%) -2.655520
dtype: float64
-----
Augmented Dickey-Fuller Test: Emissions (N2O)

Test Statistic          -2.225361
p-value                  0.197170
No. of lags Used         8.000000
No. of obsv used         11.000000
Confidence Interval (1%) -4.223238
Confidence Interval (5%) -3.189369
Confidence Interval (10%) -2.729839
dtype: float64
-----

```

For CH<sub>4</sub>, the Augmented Dickey-Fuller Test resulted in a Test Statistic of -1.804023 and a p-value of 0.378466. This suggests weak evidence against the null hypothesis of non-stationarity, and thus, we cannot conclude that the CH<sub>4</sub> emissions time series is stationary.

For CO<sub>2</sub>, the Test Statistic is -3.022048 with a p-value of 0.032886, indicating strong evidence against the null hypothesis of non-stationarity. Therefore, we can conclude that the CO<sub>2</sub> emissions time series is stationary.

For N<sub>2</sub>O, the Test Statistic is -2.225361 with a p-value of 0.197170. This suggests weak evidence against the null hypothesis of non-stationarity, and thus, we cannot conclude that the N<sub>2</sub>O emissions time series is stationary.

#### 4.4.2.2 *Table 4. Granger Causality test for Japan*

```

Causality Matrix (1 Time Lag):
      Emissions (CH4) Emissions (CO2) Emissions (N2O)
Emissions (CH4)      NaN          0.474473         0.332713
Emissions (CO2)    0.317621          NaN          0.389782
Emissions (N2O)    0.671623          0.574378          NaN
Causality Matrix (2 Time Lags):
      Emissions (CH4) Emissions (CO2) Emissions (N2O)
Emissions (CH4)      NaN          0.487024         0.862273
Emissions (CO2)    0.473618          NaN          0.726495
Emissions (N2O)    0.6985          0.433061          NaN

```

Based on the Granger Causality tests conducted with different lag numbers, it appears that there is no significant evidence to suggest a causal relationship between emissions of methane (CH<sub>4</sub>) and

emissions of carbon dioxide (CO<sub>2</sub>) or emissions of nitrous oxide (N<sub>2</sub>O). The p-values for the likelihood ratio tests, ssr based F tests, ssr based chi2 tests, and parameter F tests are all greater than the typical significance level of 0.05, indicating that there is no strong evidence to support the hypothesis that emissions of CH<sub>4</sub> cause emissions of CO<sub>2</sub> or N<sub>2</sub>O, or vice versa.

#### 4.4.2.3 Model Performance

```
Mean val Emissions (CH4): 9126.3638 -- RMSE: 327.5263 -- RMSE < 10% Mean: True
Mean val Emissions (CO2): 8178020.4741 -- RMSE: 734842.3886 -- RMSE < 10% Mean: True
Mean val Emissions (N2O): 482.8229 -- RMSE: 20.8869 -- RMSE < 10% Mean: True
```

The model's predictions for Emissions (CH<sub>4</sub>) are accurate, as the root mean squared error (RMSE) of 327.5263 is less than 10% of the mean value of 9126.3638. This is indicated by the "True" value in the "RMSE < 10% Mean" column, suggesting that the model's predictions for Emissions (CH<sub>4</sub>) are reliable.

Similarly, the model's predictions for Emissions (CO<sub>2</sub>) are also accurate, as the RMSE of 734842.3886 is less than 10% of the mean value of 8178020.4741. This is indicated by the "True" value in the "RMSE < 10% Mean" column, suggesting that the model's predictions for Emissions (CO<sub>2</sub>) are also reliable.

Furthermore, the model's predictions for Emissions (N<sub>2</sub>O) are accurate, as the RMSE of 20.8869 is less than 10% of the mean value of 482.8229. This is indicated by the "True" value in the "RMSE < 10% Mean" column, suggesting that the model's predictions for Emissions (N<sub>2</sub>O) are reliable.

#### 4.4.3 Statistical test for United States

##### 4.4.3.1 *Table 5. Augmented Dickey-Fuller (ADF) test for United States*

```
Augmented Dickey-Fuller Test: Emissions (CH4)
```

```
Test Statistic      -1.745922e+01
p-value             4.607128e-30
No. of lags Used    8.000000e+00
No. of obsv used    1.100000e+01
Confidence Interval (1%) -4.223238e+00
Confidence Interval (5%) -3.189369e+00
Confidence Interval (10%) -2.729839e+00
dtype: float64
```

```
Augmented Dickey-Fuller Test: Emissions (CO2)
```

```
Test Statistic      -4.588729
p-value             0.000135
```

```

No. of lags Used          6.000000
No. of obsv used         13.000000
Confidence Interval (1%)  -4.068854
Confidence Interval (5%)  -3.127149
Confidence Interval (10%) -2.701730
dtype: float64
-----
Augmented Dickey-Fuller Test: Emissions (N2O)

Test Statistic          -4.000486
p-value                  0.001408
No. of lags Used        7.000000
No. of obsv used        12.000000
Confidence Interval (1%) -4.137829
Confidence Interval (5%) -3.154972
Confidence Interval (10%) -2.714477
dtype: float64
-----

```

For CH<sub>4</sub>, the Augmented Dickey-Fuller Test resulted in a Test Statistic of -17.45922 and a p-value of 4.607128e-30. This suggests strong evidence against the null hypothesis of non-stationarity, and thus, we can conclude that the CH<sub>4</sub> emissions time series is stationary. For CO<sub>2</sub>, the Test Statistic is -4.588729 with a p-value of 0.000135, indicating strong evidence against the null hypothesis of non-stationarity. Therefore, we can conclude that the CO<sub>2</sub> emissions time series is stationary. For N<sub>2</sub>O, the Test Statistic is -4.000486 with a p-value of 0.001408. This suggests strong evidence against the null hypothesis of non-stationarity, and thus, we can conclude that the N<sub>2</sub>O emissions time series is stationary.

#### 4.4.3.2 Table 6. Granger Causality test for United States

```

Causality Matrix (1 Time Lag):
      Emissions (CH4) Emissions (CO2) Emissions (N2O)
Emissions (CH4)      NaN          0.445646          0.004567
Emissions (CO2)    0.000299          NaN          0.022222
Emissions (N2O)    0.072691          0.703584          NaN
Causality Matrix (2 Time Lags):
      Emissions (CH4) Emissions (CO2) Emissions (N2O)
Emissions (CH4)      NaN          0.483862          0.010935
Emissions (CO2)    0.01187          NaN          0.063244
Emissions (N2O)    0.300636          0.225918          NaN

```

The Granger causality test is used to determine whether one time series is useful in forecasting another. The results of the Granger causality tests suggest that there is evidence of unidirectional causality between emissions of methane (CH<sub>4</sub>) and emissions of carbon dioxide (CO<sub>2</sub>), with CO<sub>2</sub> causing CH<sub>4</sub> emissions. However, there is not enough evidence to support the claim that emissions of CH<sub>4</sub> cause emissions of CO<sub>2</sub>. Regarding the relationship between CH<sub>4</sub> and emissions of nitrous

oxide (N<sub>2</sub>O), the tests suggest that there is no evidence of causality between N<sub>2</sub>O and CH<sub>4</sub> emissions. There seems to be no causal relationship between CO<sub>2</sub> emissions and N<sub>2</sub>O emissions.

#### 4.4.3.3 Model Performance

```
Mean val Emissions (CH4): 138998.4559 -- RMSE: 2035.61 -- RMSE < 10% Mean: True
Mean val Emissions (CO2): 38555165.5516 -- RMSE: 1992189.3619 -- RMSE < 10%
Mean: True
Mean val Emissions (N2O): 9575.5724 -- RMSE: 115.552 -- RMSE < 10% Mean: True
```

The models' predictions for Emissions (CH<sub>4</sub>) are accurate, as the RMSE of 2035.61 is less than 10% of the mean value of 138998.4559. This is indicated by the "True" value in the "RMSE < 10% Mean" column, suggesting that the model's predictions for Emissions (CH<sub>4</sub>) are reliable.

Similarly, the model's predictions for Emissions (CO<sub>2</sub>) are also accurate, as the RMSE of 1992189.3619 is less than 10% of the mean value of 38555165.5516. This is indicated by the "True" value in the "RMSE < 10% Mean" column, suggesting that the model's predictions for Emissions (CO<sub>2</sub>) are also reliable.

Furthermore, the model's predictions for Emissions (N<sub>2</sub>O) are accurate, as the RMSE of 115.552 is less than 10% of the mean value of 9575.5724. This is indicated by the "True" value in the "RMSE < 10% Mean" column, suggesting that the model's predictions for Emissions (N<sub>2</sub>O) are reliable.

#### 4.4.4 Statistical test for Belarus

##### 4.4.4.1 *Table 7. Augmented Dickey-Fuller (ADF) test for Belarus*

```
Augmented Dickey-Fuller Test: Emissions (CH4)
```

```
Test Statistic          -4.444960
p-value                  0.000247
No. of lags Used        7.000000
No. of obsv used        12.000000
Confidence Interval (1%) -4.137829
Confidence Interval (5%) -3.154972
Confidence Interval (10%) -2.714477
dtype: float64
```

```
Augmented Dickey-Fuller Test: Emissions (CO2)
```

```
Test Statistic          -2.181525
p-value                  0.213019
No. of lags Used        8.000000
No. of obsv used        11.000000
Confidence Interval (1%) -4.223238
```

```

Confidence Interval (5%)      -3.189369
Confidence Interval (10%)     -2.729839
dtype: float64
-----
Augmented Dickey-Fuller Test: Emissions (N2O)

Test Statistic                -4.610962
p-value                      0.000123
No. of lags Used              0.000000
No. of obsv used              19.000000
Confidence Interval (1%)      -3.832603
Confidence Interval (5%)      -3.031227
Confidence Interval (10%)     -2.655520
dtype: float64
-----

```

For CH<sub>4</sub>, the Augmented Dickey-Fuller Test resulted in a Test Statistic of -4.444960 and a p-value of 0.000247. This suggests strong evidence against the null hypothesis of non-stationarity, and thus, we can conclude that the CH<sub>4</sub> emissions time series is stationary. For CO<sub>2</sub>, the Test Statistic is -2.181525 with a p-value of 0.213019. This p-value is greater than 0.05, indicating weak evidence against the null hypothesis of non-stationarity. Therefore, we cannot conclude that the CO<sub>2</sub> emissions time series is stationary. For N<sub>2</sub>O, the Test Statistic is -4.610962 with a p-value of 0.000123. This suggests strong evidence against the null hypothesis of non-stationarity, and thus, we can conclude that the N<sub>2</sub>O emissions time series is stationary.

#### 4.4.4.2 Table 8. Granger Causality test for Belarus

```

Causality Matrix (1 Time Lag):
      Emissions (CH4) Emissions (CO2) Emissions (N2O)
Emissions (CH4)      NaN          0.322688         0.472648
Emissions (CO2)    0.009836          NaN          0.808566
Emissions (N2O)    0.002166         0.345468          NaN
Causality Matrix (2 Time Lags):
      Emissions (CH4) Emissions (CO2) Emissions (N2O)
Emissions (CH4)      NaN          0.366547         0.4924
Emissions (CO2)    0.203645          NaN          0.803678
Emissions (N2O)    0.07065         0.266199          NaN

```

The results of the Granger causality tests suggest that there is evidence of unidirectional causality between emissions of methane (CH<sub>4</sub>) and emissions of carbon dioxide (CO<sub>2</sub>), with CH<sub>4</sub> causing CO<sub>2</sub> emissions. However, there is not enough evidence to support the claim that emissions of CO<sub>2</sub> cause emissions of CH<sub>4</sub>.



Regarding the relationship between CH<sub>4</sub> and emissions of nitrous oxide (N<sub>2</sub>O), the tests suggest that there is not enough evidence to conclude that CH<sub>4</sub> causes N<sub>2</sub>O emissions or vice versa. There seems to be no causal relationship between CO<sub>2</sub> emissions and N<sub>2</sub>O emissions.

#### 4.4.4.3 Model Performance

```
Mean val Emissions (CH4): 4583.6409 -- RMSE: 15.3235 -- RMSE < 10% Mean: True
Mean val Emissions (CO2): 1028080.3522 -- RMSE: 20471.6145 -- RMSE < 10% Mean: True
Mean val Emissions (N2O): 738.1418 -- RMSE: 47.6727 -- RMSE < 10% Mean: True
```

The models' predictions for Emissions (CH<sub>4</sub>) are accurate, as the RMSE of 15.3235 is less than 10% of the mean value of 4583.6409. This is indicated by the "True" value in the "RMSE < 10% Mean" column, suggesting that the model's predictions for Emissions (CH<sub>4</sub>) are reliable.

Similarly, the model's predictions for Emissions (CO<sub>2</sub>) are also accurate, as the RMSE of 20471.6145 is less than 10% of the mean value of 1028080.3522. This is indicated by the "True" value in the "RMSE < 10% Mean" column, suggesting that the model's predictions for Emissions (CO<sub>2</sub>) are also reliable.

Furthermore, the model's predictions for Emissions (N<sub>2</sub>O) are accurate, as the RMSE of 47.6727 is less than 10% of the mean value of 738.1418. This is indicated by the "True" value in the "RMSE < 10% Mean" column, suggesting that the model's predictions for Emissions (N<sub>2</sub>O) are reliable.

#### 4.4.5 Statistical test for Ukraine

##### 4.4.5.1 *Table 9. Augmented Dickey-Fuller (ADF) test for Ukraine*

```
Augmented Dickey-Fuller Test: Emissions (CH4)
```

```
Test Statistic      -2.278486
p-value             0.179001
No. of lags Used    8.000000
No. of obsv used    11.000000
Confidence Interval (1%) -4.223238
Confidence Interval (5%) -3.189369
Confidence Interval (10%) -2.729839
dtype: float64
```

```
-----
Augmented Dickey-Fuller Test: Emissions (CO2)
```

```
Test Statistic      -4.697866
p-value             0.000085
No. of lags Used    0.000000
No. of obsv used    19.000000
Confidence Interval (1%) -3.832603
```

```

Confidence Interval (5%)      -3.031227
Confidence Interval (10%)     -2.655520
dtype: float64

-----
Augmented Dickey-Fuller Test: Emissions (N2O)

Test Statistic                -4.205096
p-value                       0.000646
No. of lags Used              8.000000
No. of obsv used              11.000000
Confidence Interval (1%)      -4.223238
Confidence Interval (5%)      -3.189369
Confidence Interval (10%)     -2.729839
dtype: float64

-----

```

For CH<sub>4</sub>, the Augmented Dickey-Fuller Test resulted in a Test Statistic of -2.278486 and a p-value of 0.179001. This p-value is greater than 0.05, indicating weak evidence against the null hypothesis of non-stationarity. Therefore, we cannot conclude that the CH<sub>4</sub> emissions time series is stationary. For CO<sub>2</sub>, the Test Statistic is -4.697866 with a p-value of 0.000085. This suggests strong evidence against the null hypothesis of non-stationarity, and thus, we can conclude that the CO<sub>2</sub> emissions time series is stationary. For N<sub>2</sub>O, the Test Statistic is -4.205096 with a p-value of 0.000646. This suggests strong evidence against the null hypothesis of non-stationarity, and thus, we can conclude that the N<sub>2</sub>O emissions time series is stationary.

#### 4.4.5.2 Table 10. Granger Causality test for Ukraine

```

Causality Matrix (1 Time Lag):
      Emissions (CH4) Emissions (CO2) Emissions (N2O)
Emissions (CH4)      NaN          0.045828         0.008014
Emissions (CO2)    0.536406          NaN          0.073034
Emissions (N2O)    0.435092          0.006313          NaN
Causality Matrix (2 Time Lags):
      Emissions (CH4) Emissions (CO2) Emissions (N2O)
Emissions (CH4)      NaN          0.316337         0.020615
Emissions (CO2)    0.181442          NaN          0.011137
Emissions (N2O)    0.60806          0.039756          NaN

```

The Granger causality test is used to determine whether one time series is useful in forecasting another. The results of the Granger causality tests suggest that there is evidence of unidirectional causality between emissions of methane (CH<sub>4</sub>) and emissions of carbon dioxide (CO<sub>2</sub>), with CO<sub>2</sub> causing CH<sub>4</sub> emissions. However, there is not enough evidence to support the claim that emissions of CH<sub>4</sub> cause emissions of CO<sub>2</sub>. Regarding the relationship between CH<sub>4</sub> and emissions of nitrous oxide (N<sub>2</sub>O), the tests suggest that there is evidence of unidirectional causality between N<sub>2</sub>O and

CH<sub>4</sub> emissions, with N<sub>2</sub>O causing CH<sub>4</sub> emissions. There seems to be no causal relationship between CO<sub>2</sub> emissions and N<sub>2</sub>O emissions.

#### 4.4.5.3 Model Performance

```
Mean val Emissions (CH4): 9918.9218 -- RMSE: 161.9123 -- RMSE < 10% Mean: True
Mean val Emissions (CO2): 2001741.7963 -- RMSE: 131391.5505 -- RMSE < 10% Mean: True
Mean val Emissions (N2O): 1070.707 -- RMSE: 31.4954 -- RMSE < 10% Mean: True
```

The models' predictions for Emissions (CH<sub>4</sub>) are accurate, as the RMSE of 161.9123 is less than 10% of the mean value of 9918.9218. This is indicated by the "True" value in the "RMSE < 10% Mean" column, suggesting that the model's predictions for Emissions (CH<sub>4</sub>) are reliable.

Similarly, the model's predictions for Emissions (CO<sub>2</sub>) are also accurate, as the RMSE of 131391.5505 is less than 10% of the mean value of 2001741.7963. This is indicated by the "True" value in the "RMSE < 10% Mean" column, suggesting that the model's predictions for Emissions (CO<sub>2</sub>) are also reliable.

Furthermore, the model's predictions for Emissions (N<sub>2</sub>O) are accurate, as the RMSE of 31.4954 is less than 10% of the mean value of 1070.707. This is indicated by the "True" value in the "RMSE < 10% Mean" column, suggesting that the model's predictions for Emissions (N<sub>2</sub>O) are reliable.

#### 4.4.6 Statistical test for Nigeria

##### 4.4.6.1 *Table 11. Augmented Dickey-Fuller (ADF) test for Nigeria*

```
Augmented Dickey-Fuller Test: Emissions (CH4)
```

```
Test Statistic      -1.210884
p-value             0.668896
No. of lags Used    7.000000
No. of obsv used    12.000000
Confidence Interval (1%) -4.137829
Confidence Interval (5%) -3.154972
Confidence Interval (10%) -2.714477
dtype: float64
```

```
Augmented Dickey-Fuller Test: Emissions (CO2)
```

```
Test Statistic      -1.260850
p-value             0.646888
No. of lags Used    7.000000
No. of obsv used    12.000000
Confidence Interval (1%) -4.137829
Confidence Interval (5%) -3.154972
Confidence Interval (10%) -2.714477
```

```

dtype: float64
-----
Augmented Dickey-Fuller Test: Emissions (N2O)

Test Statistic      -2.416367
p-value             0.137170
No. of lags Used    2.000000
No. of obsv used    17.000000
Confidence Interval (1%) -3.889266
Confidence Interval (5%) -3.054358
Confidence Interval (10%) -2.666984
dtype: float64
-----

```

For CH<sub>4</sub>, the Augmented Dickey-Fuller Test resulted in a Test Statistic of -1.210884 and a p-value of 0.668896. This p-value is greater than 0.05, indicating weak evidence against the null hypothesis of non-stationarity. Therefore, we cannot conclude that the CH<sub>4</sub> emissions time series is stationary. For CO<sub>2</sub>, the Test Statistic is -1.260850 with a p-value of 0.646888. This p-value is greater than 0.05, indicating weak evidence against the null hypothesis of non-stationarity. Therefore, we cannot conclude that the CO<sub>2</sub> emissions time series is stationary. For N<sub>2</sub>O, the Test Statistic is -2.416367 with a p-value of 0.137170. This p-value is greater than 0.05, indicating weak evidence against the null hypothesis of non-stationarity. Therefore, we cannot conclude that the N<sub>2</sub>O emissions time series is stationary.

#### 4.4.6.2 Table 12. Granger Causality test for Nigeria

Causality Matrix (1 Time Lag):			
	Emissions (CH <sub>4</sub> )	Emissions (CO <sub>2</sub> )	Emissions (N <sub>2</sub> O)
Emissions (CH <sub>4</sub> )	NaN	0.585345	0.155826
Emissions (CO <sub>2</sub> )	0.069497	NaN	0.584566
Emissions (N <sub>2</sub> O)	0.044967	0.032342	NaN
Causality Matrix (2 Time Lags):			
	Emissions (CH <sub>4</sub> )	Emissions (CO <sub>2</sub> )	Emissions (N <sub>2</sub> O)
Emissions (CH <sub>4</sub> )	NaN	0.435209	0.86063
Emissions (CO <sub>2</sub> )	0.056697	NaN	0.84249
Emissions (N <sub>2</sub> O)	0.033459	0.082581	NaN

The results of the Granger causality tests suggest that there is evidence of unidirectional causality between emissions of methane (CH<sub>4</sub>) and emissions of carbon dioxide (CO<sub>2</sub>), with CO<sub>2</sub> causing CH<sub>4</sub> emissions. However, there is not enough evidence to support the claim that emissions of CH<sub>4</sub> cause emissions of CO<sub>2</sub>. Regarding the relationship between CH<sub>4</sub> and emissions of nitrous oxide (N<sub>2</sub>O), the tests suggest that there is evidence of unidirectional causality between CH<sub>4</sub> and N<sub>2</sub>O emissions, with CH<sub>4</sub> causing N<sub>2</sub>O emissions. There seems to be no causal relationship between CO<sub>2</sub> emissions and N<sub>2</sub>O emissions.

#### 4.4.6.3 Model Performance

```
Mean val Emissions (CH4): 40251.2775 -- RMSE: 1288.749 -- RMSE < 10% Mean: True
Mean val Emissions (CO2): 3067217.6172 -- RMSE: 34478.1651 -- RMSE < 10% Mean: True
Mean val Emissions (N2O): 1085.2707 -- RMSE: 13.8881 -- RMSE < 10% Mean: True
```

The models' predictions for Emissions (CH<sub>4</sub>) are accurate, as the RMSE of 1288.749 is less than 10% of the mean value of 40251.2775. This is indicated by the "True" value in the "RMSE < 10% Mean" column, suggesting that the model's predictions for Emissions (CH<sub>4</sub>) are reliable.

Similarly, the model's predictions for Emissions (CO<sub>2</sub>) are also accurate, as the RMSE of 34478.1651 is less than 10% of the mean value of 3067217.6172. This is indicated by the "True" value in the "RMSE < 10% Mean" column, suggesting that the model's predictions for Emissions (CO<sub>2</sub>) are also reliable.

Furthermore, the model's predictions for Emissions (N<sub>2</sub>O) are accurate, as the RMSE of 13.8881 is less than 10% of the mean value of 1085.2707. This is indicated by the "True" value in the "RMSE < 10% Mean" column, suggesting that the model's predictions for Emissions (N<sub>2</sub>O) are reliable.

#### 4.4.7 Statistical test for Colombia

##### 4.4.7.1 *Table 13. Augmented Dickey-Fuller (ADF) test for Columbia*

```
Augmented Dickey-Fuller Test: Emissions (CH4)
```

```
Test Statistic      -2.258468
p-value             0.185713
No. of lags Used    0.000000
No. of obsv used    19.000000
Confidence Interval (1%) -3.832603
Confidence Interval (5%) -3.031227
Confidence Interval (10%) -2.655520
dtype: float64
```

```
Augmented Dickey-Fuller Test: Emissions (CO2)
```

```
Test Statistic      -3.646264
p-value             0.004936
No. of lags Used    8.000000
No. of obsv used    11.000000
Confidence Interval (1%) -4.223238
Confidence Interval (5%) -3.189369
Confidence Interval (10%) -2.729839
dtype: float64
```

Augumented Dickey-Fuller Test: Emissions (N<sub>2</sub>O)

```
Test Statistic      -0.951684
p-value             0.770476
No. of lags Used    8.000000
No. of obsv used    11.000000
Confidence Interval (1%) -4.223238
Confidence Interval (5%) -3.189369
Confidence Interval (10%) -2.729839
dtype: float64
```

For CH<sub>4</sub>, the Augmented Dickey-Fuller Test resulted in a Test Statistic of -2.258468 and a p-value of 0.185713. This p-value is greater than 0.05, indicating weak evidence against the null hypothesis of non-stationarity. Therefore, we cannot conclude that the CH<sub>4</sub> emissions time series is stationary.

For CO<sub>2</sub>, the Test Statistic is -3.646264 with a p-value of 0.004936. This p-value is less than 0.05, indicating strong evidence against the null hypothesis of non-stationarity. Therefore, we can conclude that the CO<sub>2</sub> emissions time series is stationary.

For N<sub>2</sub>O, the Test Statistic is -0.951684 with a p-value of 0.770476. This p-value is greater than 0.05, indicating weak evidence against the null hypothesis of non-stationarity. Therefore, we cannot conclude that the N<sub>2</sub>O emissions time series is stationary.

#### 4.4.7.2 Table 14. Granger Causality test for Columbia

Causality Matrix (1 Time Lag):

	Emissions (CH <sub>4</sub> )	Emissions (CO <sub>2</sub> )	Emissions (N <sub>2</sub> O)
Emissions (CH <sub>4</sub> )	NaN	0.304739	0.226432
Emissions (CO <sub>2</sub> )	0.00117	NaN	0.000815
Emissions (N <sub>2</sub> O)	0.693649	0.685263	NaN

Causality Matrix (2 Time Lags):

	Emissions (CH <sub>4</sub> )	Emissions (CO <sub>2</sub> )	Emissions (N <sub>2</sub> O)
Emissions (CH <sub>4</sub> )	NaN	0.238399	0.011257
Emissions (CO <sub>2</sub> )	0.049761	NaN	0.001776
Emissions (N <sub>2</sub> O)	0.183198	0.68878	NaN

Based on the Granger causality tests results, there is no evidence to suggest that emissions of CH<sub>4</sub> or CO<sub>2</sub> cause emissions of N<sub>2</sub>O. On the other hand, there is some evidence to suggest that emissions of CH<sub>4</sub> may cause emissions of CO<sub>2</sub>. However, the evidence is not particularly strong, and more research would be necessary to establish a definitive causal relationship.

#### 4.4.7.3 Model Performance

```
Mean val Emissions (CH4): 20598.8263 -- RMSE: 1505.2005 -- RMSE < 10% Mean:
True
Mean val Emissions (CO2): 2592625.4498 -- RMSE: 198072.5219 -- RMSE < 10% Mean:
True
```

Mean val Emissions (N<sub>2</sub>O): 623.2535 -- RMSE: 51.7667 -- RMSE < 10% Mean: True

The models' predictions for Emissions (CH<sub>4</sub>) are accurate, as the RMSE of 1505.2005 is less than 10% of the mean value of 20598.8263. This is indicated by the "True" value in the "RMSE < 10% Mean" column, suggesting that the model's predictions for Emissions (CH<sub>4</sub>) are reliable.

Similarly, the model's predictions for Emissions (CO<sub>2</sub>) are also accurate, as the RMSE of 198072.5219 is less than 10% of the mean value of 2592625.4498. This is indicated by the "True" value in the "RMSE < 10% Mean" column, suggesting that the model's predictions for Emissions (CO<sub>2</sub>) are also reliable.

Furthermore, the model's predictions for Emissions (N<sub>2</sub>O) are accurate, as the RMSE of 51.7667 is less than 10% of the mean value of 623.2535. This is indicated by the "True" value in the "RMSE < 10% Mean" column, suggesting that the model's predictions for Emissions (N<sub>2</sub>O) are reliable.

## Chapter 5

### 5.0 Conclusion

In this chapter, further examination of Experiments and Results is presented, with the purpose of outlining key discoveries, observations, insights and limitations faced. The aim is to offer a comprehensive summary of the major findings, followed by recommendations on how to address the challenges that have been identified.

Ultimately, this analysis seeks to provide valuable insights so researchers can identify problem areas that still require intervention, significantly track changes in the climate system and postulate possible adaptation techniques which mitigate climate change impacts.

### 5.1 Findings and Limitation

The Granger Causality test conducted on the data set revealed a causal effect of N<sub>2</sub>O emissions on CH<sub>4</sub> and CO<sub>2</sub> emissions. A study by Yusuf et al. (2020) agrees with this notion, while Tumendelger et al. (2019) strongly disagrees, going contrary to existing beliefs and challenging the norm in literature stating that the variability of emission factors is high, meaning that causation could be due to a wide range of operational parameters instead and as such, more research is needed to factor in operational parameters during such studies.

Seven countries of focus were selected based on the World Economic Situation and Prospects report by the United Nations. Three countries were picked from the developed economies, two were picked from the economies in transition and two were picked from developing economies. They include Spain, Japan, United States, Belarus, Ukraine, Nigeria and Colombia.

In Spain, carbon dioxide emissions peaked in 2007, followed by a downward trend since then. Similarly, methane (CH<sub>4</sub>) emissions decreased over time, and started to rise in 2012. Nitrous oxide (N<sub>2</sub>O) emissions followed the CH<sub>4</sub> pattern, but at a relatively lower magnitude. Predictions indicate a steady increase in emissions of CH<sub>4</sub> from 13,498.45 Gg in 2020 to 13,524.75 Gg in 2030 and CO<sub>2</sub> from 2,355,785 Gg in 2020 to 2,428,219 Gg in 2030 over time, with stable N<sub>2</sub>O emissions, only slightly increasing from 687.04 Gg in 2020 to 684.56 Gg in 2030.

In Japan the emissions of methane (CH<sub>4</sub>) and nitrous oxide (N<sub>2</sub>O) steadily declined. Carbon dioxide (CO<sub>2</sub>) emissions decreased in 2009, increased to peak levels in 2013, and has since been on a decline. Predictions indicate a steady increase in emissions of CH<sub>4</sub> from 9,507.02 Gg in 2020



to 10,308.95 Gg in 2030 and CO<sub>2</sub> from 8,847,428 Gg in 2020 to 8,704,751 Gg in 2030 over time, with N<sub>2</sub>O emissions, slightly increasing from 507.93 Gg in 2020 to 556.84 Gg in 2030.

In the United States, Carbon dioxide (CO<sub>2</sub>) emissions start at high levels, and show a decline trend. A significant decline was observed in 2020 due to the pandemic. Nitrous oxide (N<sub>2</sub>O) emissions were low and fluctuating, but also experienced a reduction in 2020. Predictions indicate a steady increase in emissions of CH<sub>4</sub> from 136,376.54 Gg in 2020 to 137,962.22 Gg in 2030, and CO<sub>2</sub> from 38,751,303.65 Gg in 2020 to 39,561,435.44 Gg in 2030 over time. N<sub>2</sub>O emissions are expected to slightly decline from 9,557.20 Gg in 2020 to 9,548.13 Gg in 2030.

In Belarus, increasing levels of methane (CH<sub>4</sub>), carbon dioxide (CO<sub>2</sub>), and nitrous oxide (N<sub>2</sub>O) emissions were observed from the year 2010. These emissions levels have not shown a significant reduction, despite the COVID-19 pandemic in 2020. Predictions indicate a steady decrease in emissions of CH<sub>4</sub> from 4579.107946 Gg in 2020 to 4452.871106 Gg in 2030, and CO<sub>2</sub> from 1.002316e+06 Gg in 2020 to 9.535096e+05 Gg in 2030 over time. N<sub>2</sub>O emissions are also expected to decline from 693.219728 Gg in 2020 to 642.481489 Gg in 2030.

In Ukraine, a downward trend in CH<sub>4</sub> and CO<sub>2</sub> emissions over time were observed, while N<sub>2</sub>O emissions maintained a consistent upward trend. Predictions indicate a steady increase in emissions of CH<sub>4</sub> from 10194.515239 Gg in 2020 to 12905.881501 Gg in 2030, and CO<sub>2</sub> from 1.824629e+06 Gg in 2020 to 2.006878e+06 Gg in 2030. N<sub>2</sub>O emissions are also predicted to slightly reduce from 1059.149865 Gg in 2020 to 976.604660 Gg in 2030.

In Nigeria, CH<sub>4</sub> emissions remained relatively stable over the observed period, while N<sub>2</sub>O and CO<sub>2</sub> emissions displayed an increasing trend. A decline in all emissions was observed in 2009, with N<sub>2</sub>O emission being comparatively lower than the others, then an upward trend resumed. Predictions indicate a steady increase in emissions of CH<sub>4</sub> from 41840.642523 Gg in 2020 to 42829.154240 Gg in 2030, and CO<sub>2</sub> from 3.033347e+06 Gg in 2020 to 3.054608e+06 Gg in 2030. N<sub>2</sub>O emissions are expected to slightly decline from 1086.221344 Gg in 2020 to 1075.731444 Gg in 2030.

In Colombia, CH<sub>4</sub>, CO<sub>2</sub>, and N<sub>2</sub>O emissions have exhibited an overall increasing trend. A decline across all three emissions was observed between 2012 and 2016, after which all three emissions increased again, exceeding their previous levels. Predictions indicate a decrease in emissions of

CH<sub>4</sub> from 19334.172608 Gg in 2020 to 18732.354228 Gg in 2030, and CO<sub>2</sub> from 2.404636e+06 Gg in 2020 to 2.287395e+06 Gg in 2030 and N<sub>2</sub>O from 573.839110 Gg in 2020 to 556.221861 Gg in 2030.

The United States which was recorded to have the highest emission rates across all three emissions was still predicted to maintain the lead, via a huge margin in 2030. Although some work is being put in to mitigate the effects of climate change in the United States such as broad-based climate policies, policies on transportation systems, and policies on the use of electricity which affects climate change (Basseches et al., 2022), efforts should be intensified to bring a decline in the already high values. Some limitations were observed during the duration of this project. Firstly, this project was limited in scope to only the data available. Secondly time factor prevented the researcher from exploring a higher number of countries with the model built.

## 5.2 Future Work

An extensive analysis only 7 countries was carried out using the built machine learning model. In future, other countries can be considered using the same model to identify current trends and predicted patterns.

## References

- Aamir, M., Bhatti, M. A., Bazai, S. U., Marjan, S., Mirza, A. M., Wahid, A., Hasnain, A. & Bhatti, U. A. 2022. Predicting the Environmental Change of Carbon Emission Patterns in South Asia: A Deep Learning Approach Using BiLSTM. *Athmosphere*, 13, 1-14.
- Abdullah, L. & Pauzi, H. M. 2015. Methods In Forecasting Carbon Dioxide Emissions: A Decade Review. *Jurnal Teknologi*, 75, 67-82.
- Acheampong, A. O. & Boateng, E. B. 2019. Modelling carbon emission intensity: Application of artificial neural network. *Journal of Cleaner Production*, 225, 833-856.
- Aftab, S., Ahmed, A., Chandio, A. A., Korankye, B. A., Ali, A. & Fang, W. 2021. Modeling the nexus between carbon emissions, energy consumption, and economic progress in Pakistan: Evidence from cointegration and causality analysis. *Energy Reports*, 7, 4642-4658.
- Alam, T. & Alarjani, A. 2021a. A Comparative Study of CO<sub>2</sub> Emission Forecasting in the Gulf Countries Using Autoregressive Integrated Moving Average, Artificial Neural Network, and Holt-Winters Exponential Smoothing Models. *Advances in Meteorology*, 2021.
- Alam, T. & Alarjani, A. 2021b. Forecasting CO<sub>2</sub> Emissions in Saudi Arabia Using Artificial Neural Network, Holt-Winters Exponential Smoothing, and Autoregressive Integrated Moving Average Models. 2021 International Conference on Technology and Policy in Energy and Electric Power (ICT-PEP). Jakarta, Indonesia: IEEE.
- Alamgir, H. M., K, C. R., Sondoss, E. & J, R. M. 2020. Hybrid deep learning model for ultra-short-term wind power forecasting. 2020 IEEE International Conference on Applied Superconductivity and Electromagnetic Devices. IEEE.
- Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A. & Aljaaf, A. J. 2019. Supervised and Unsupervised Learning for Data Science. In: Berry, M. W., Mohamed, A. & Yap, B. W. (eds.) *Unsupervised and Semi-Supervised Learning*.
- Basseches, J. A., Bromley-Trujillo, R., Boykoff, M. T., Culhane, T., Hall, G., Healy, N., Hess, D. J., Hsu, D., Krause, R. M., Prechel, H., Roberts, J. T. & Stephens, J. C. 2022. Climate policy conflict in the U.S. states: a critical review and way forward. *Climate Change*, 170.
- Belousova, K. 2023. Ukraine reduced carbon emissions by 62.5%. Infographics [Online]. Available: <https://ecopolitic.com.ua/en/news/ukraina-zmenshila-vikidi-vuglecju-na-62-5-infografika-2/> [Accessed April 2023].
- Bhatti, U. A., Wu, G., Bazai, S. U., Nawaz, S. A., Baryalai, M., Bhatti, M. A., Hasnain, A. & Nizamani, M. M. 2022. A Pre- to Post-COVID-19 Change of Air Quality Patterns in Anhui Province Using Path Analysis and Regression. *Polish Journal of Environmental Studies*, 31, 4029–4042.

- Bokonda, P. L., Ouazzani-Touhami, K. & Souissi, N. Predictive analysis using machine learning: Review of trends and methods. International Symposium on Advanced Electrical and Communication Technologies, 2020. IEEE.
- Bora, N. 2021. Understanding ARIMA Models for Machine Learning [Online]. Available: <https://www.capitalone.com/tech/machine-learning/understanding-arima-models/> [Accessed April 2023].
- Brooks, C. & Thompson, C. 2017. Chapter 5 :Predictive Modelling in Teaching and Learning.
- Chatfield, C. 1987. The Holt-Winters Forecasting Procedure. Journal of the Royal Statistical Society. Series C (Applied Statistics), 27, 264-279.
- Chigora, F., Thaban, N. & Mutambara, E. 2019. Forecasting CO2 Emission for Zimbabwe's Tourism Destination vibrancy: A Univariate Approach using Box-Jenkins ARIMA Model. African Journal of Hospitality, Tourism and Leisure, 8, 1-15.
- Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F. & Campbell, J. P. 2020. Introduction to Machine Learning, Neural Networks, and Deep Learning translational vision science & technology, 9.
- Cornegruta, S., Bakewell, R., Withey, S. & Montana, G. Modelling Radiological Language with Bidirectional Long Short-Term Memory Networks. Seventh International Workshop on Health Text Mining and Information Analysis, 2016.
- Darkwah, W. K., Odum, B., Addae, M. & Koomson, D. 2018. Greenhouse Effect: Greenhouse Gases and Their Impact on Global Warming. Journal of Scientific Research and Reports, 17, 1-9.
- Deetman S, Marinova S, van der Voet E, van Vuuren D P, Edelenbosch O and Heijungs R 2020 Modelling global material stocks and flows for residential and service sector buildings towards 2050 J. Cleaner Prod. 245 118658
- Denchak, M. 2019. Greenhouse Effect 101 [Online]. Available: <https://www.nrdc.org/stories/greenhouse-effect-101#whatis> [Accessed April 2023].
- Doll, J. E. & Baransk, M. 2011. Climate Change and Agriculture Fact Series. In: Michigan State University (ed.).
- Faruque, O., Rabby, A. J., Hossain, A., Islam, R., Rashid, M. U. & Muyeen, S. M. 2022. A comparative analysis to forecast carbon dioxide emissions. Energy Reports, 8, 8046–8060.
- Fatimetou Zahra Mohamad Mahmoud (2017) The application of Predictive Analytics: Benefits, Challenges & how it can be improved, International Journal of Scientific and Research Publications, Volume 7(5), May 2017, ISSN 2250-3153.
- Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M. et al., (2020) Global Carbon Budget 2020, Earth Syst. Sci. Data, 12, 3269–3340, <https://doi.org/10.5194/essd-12-3269-2020>, 2020.
- Haines, A., Kovats, R.S., Campbell-Lendrum, D. and Corvalan, C. (2006) Climate Change and Human Health: Impacts, Vulnerability and Public Health. Public Health, 120, 585-596.

- Hayes, A. 2022. Autoregressive Integrated Moving Average (ARIMA) Prediction Model [Online]. Available: <https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp> [Accessed April 2023].
- Hu, Y. 2023. Bidirectional Analysis Model of Green Investment and Carbon Emission Based on LSTM Neural Network Xiangtan, China Xiangtan University.
- Hung, C. 2023. Deep learning in biomedical informatics. In: Zheng, Y. & Wu, Z. (eds.) Intelligent Nanotechnology. Elsevier Inc.
- ICPP (2021) Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Retrieved from <https://www.ipcc.ch/2021/08/09/ar6-wg1-20210809-pr/> on April 3, 2023
- Intergovernmental Panel on Climate Change 2022. Climate change 2022 impacts, adaptation and vulnerability summary for Policymakers. Cambridge, UK and New York, USA.
- Javatpoint. 2022. Artificial Neural Network Tutorial [Online]. Available: <https://www.javatpoint.com/artificial-neural-network> [Accessed April 2023].
- Jennings, C. L., Montgomery, D. & Kulahci, M. 2016. Introduction to Time Series Analysis and Forecasting., Wiley.
- Jovel, J. & Greiner, R. 2021. An Introduction to Machine Learning Approaches for Biomedical Research. Front. Med., 8.
- Juniarko, P., Ridho, H. & Gunawan, N. 2015. The Prediction of Carbon Dioxide Emission Using ARIMA for Support Green Energy Development in Surabaya Municipality. New, Renewable Energy and Energy Conservation Conference and Exhibition. Indonesia.
- Kone, A. C. & Buke, T. 2010. Forecasting of CO<sub>2</sub> emissions from fuel combustion using trend analysis. Renewable and Sustainable Energy Reviews, 14, 2906-2915.
- Kumari, S. & Singh, S. K. 2022. Machine learning-based time series models for effective CO<sub>2</sub> emission prediction in India. Environ Sci Pollut Res Int.
- Kweku, D. W., Bismark, O., Maxwell, A., Desmond, K. A., Danso, K. B., Oti-Mensah, E. A., Quachie, A. T. & Adormaa, B. B. 2017. Greenhouse Effect: Greenhouse Gases and Their Impact on Global Warming Journal of Scientific Research & Reports 17, 1-9.
- Lai, Y. & Dzombak, D. A. 2020. Use of the Autoregressive Integrated Moving Average (ARIMA) Model to Forecast Near-Term Regional Temperature and Precipitation Weather and Forecasting, 35, 959–976.
- Lakshay Swani, Pratika Tyagi (2017) Predictive Modelling Analytics through Data Mining, International Research Journal of Engineering & Technology (IRJET), Volume: 04 Issue: 09| Sep-2017, e-ISSN: 2395 -0056, P-ISSN: 2395-0072.
- Li, S., Siu, Y. W. & Zhao, G. 2021. Driving Factors of CO<sub>2</sub> Emissions: Further Study Based on Machine Learning. Front. Environ. Sci., 9, 1 - 16.

- Li, Y., Ma, G., Yang, J., Wang, H., Feng, J. & Ma, Y. 2020. Dynamic equivalent modeling for power converter based on LSTM neural network in wide operating range. 2020 The International Conference on Power Engineering (ICPE 2020). Guangzhou, China: ScienceDirect.
- Liu, Y., Yang, Q., Li, Y. & Zhang, L. 2020. Application of Machine Learning in Organic Chemistry. Chinese Journal of Organic Chemistry, 40.
- Lotfalipour, M. R., Falahi, M. A. & Bastam, M. 2013. Prediction of CO<sub>2</sub> Emissions in Iran using Grey and ARIMA Models. International Journal of Energy Economics and Policy, 3, 229-237.
- Magazzino, C. & Mele, M. 2022. A new machine learning algorithm to explore the CO<sub>2</sub> emissions-energy use-economic growth trilemma. Annals of Operations Research, 1-19.
- Malik, A. & Lan, J. 2016. The role of outsourcing in driving global carbon emissions. Economic Systems Research, 28, 168-182.
- Meng, Y. & Noman, H. 2022. Predicting CO<sub>2</sub> Emission Footprint Using AI through Machine Learning. Atmosphere, 13.
- Mitić, P., Ivanović, O. M. & Zdravković, A. 2017. A Cointegration Analysis of Real GDP and CO<sub>2</sub> Emissions in Transitional Countries. Sustainability, 9, 567-574.
- Mohammed Redha Qader, Khan, S., Kamal, M., Usman, M. & Haseeb, M. 2022. Forecasting carbon emissions due to electricity power generation in Bahrain. Environ Sci Pollut Res Int., 29, 17346–17357.
- Moor, J. 2006. The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years AI Mag, 28, 87.
- NASA Earth Observatory (2020) 2020 Tied for Warmest Year on Record. Retrieved from <https://earthobservatory.nasa.gov/images/147794/2020-tied-for-warmest-year-on-record>
- National Aeronautics and Space Administration. 2023. Global Temperature [Online]. Available: <https://climate.nasa.gov/vital-signs/global-temperature/> [Accessed April 2023].
- Ning, L., Pei, L. & Li, F. 2021. Forecast of China's Carbon Emissions Based on ARIMA Method. Discrete Dynamics in Nature and Society, 2021.
- Nyoni, T. & Wellington G., B. 2019. Prediction of CO<sub>2</sub> Emissions in India using ARIMA Models. Dynamic Research Journals (DRJ), 4, 01-10.
- On 3rd April, 2023.
- Park, Y.-S. & Lek, S. 2016. Chapter 7 - Artificial Neural Networks: Multilayer Perceptron for Ecological Modeling. Developments in Environmental Modelling, 28, 123-140.
- Pathak H, Jain N, Bhatia A, Patel J and Aggarwal P K (2010) Carbon footprints of Indian food items Agric. Ecosyst. Environ. 139 66–73

Patz, J.A., Campbell-Lendrum, D., Holloway, T. and Foley, J.A. (2005) Impact of Regional Climate Change on Human Health. *Nature*, 438, 310-317.

Projectpro. 2022. How to Build ARIMA Model in Python for time series forecasting? [Online]. Available: <https://www.projectpro.io/article/how-to-build-arma-model-in-python/544> [Accessed March 2023].

Rahman, A. & Hassan, M. 2017. Modeling and Forecasting of Carbon Dioxide Emissions in Bangladesh Using Autoregressive Integrated Moving Average (ARIMA) Models. *Open Journal of Statistics*, 7, 560-566.

Rehman, A., Ma, H., Ozturk, I., Murshed, M. & Dagar, V. 2021. The dynamic impacts of CO<sub>2</sub> emissions from different sources on Pakistan's economic progress: a roadmap to sustainable development. *Environment, Development and Sustainability*, 23, 17857–17880

Ritchie, H., Roser, M. & Rosado, P. 2020. CO<sub>2</sub> and Greenhouse Gas Emissions. *Our World in Data*.

Saa, E. D. & Ranathunga, L. 2020. Comparison between ARIMA and Deep Learning Models for Temperature Forecasting. *arXiv*.

Safa, M., Nejat, M., Nuthall, P. & Greig, B. 2016. Predicting CO<sub>2</sub> emissions from farm inputs in wheat production using artificial neural networks and linear regression models - Case study in Canterbury, New Zealand. *International Journal of Advanced Computer Science and Applications*, 7, 268-274.

Saleh, C., Leuveano, R. a. C., Rahman, M. N. A., Deros, B. M. & Dzakiyullah, N. R. 2015. Prediction Of CO<sub>2</sub> Emissions Using An Artificial Neural Network: The Case of The Sugar Industry. *American Scientific Publishers*, 211, 2079-2083.

Sghir, N., Adadi, A. & Lahmer, M. 2022. Recent advances in Predictive Learning Analytics: A decade systematic review (2012–2022). *Education and Information Technologies*

Shi, Z. (2018) Impact of Climate Change on the Global Environment and Associated Human Health. *Open Access Library Journal*, 5, 1-6. doi: 10.4236/oalib.1104934.

Siami-Namin, S., Tavakoli, N. & Namin, A. S. 2019. The Performance of LSTM and BiLSTM in Forecasting Time Series. *IEEE International Conference on Big Data*.

Singh, P. K., Pandey, A. K., Ahuja, S. & Kiran, R. 2021. Multiple-Forecasting Approach: A Prediction of CO<sub>2</sub> Emission from the Paddy Crop in India. *Environmental Science and Pollution Research*, 29, 25461--25472.

Thanh, H. V., Sugai, Y. & Sasaki, K. 2022. Application of artificial neural network for predicting the performance of CO<sub>2</sub> enhanced oil recovery and storage in residual oil zones. *Scientific Reports*, 10.

Tumendelger, A., Alshboul, Z. & Lorke, A. 2019. Methane and nitrous oxide emission from different treatment units of municipal wastewater treatment plants in Southwest Germany. *Plos One*, 14.

United Nations. 2023. SDG Goals [Online]. Available: <https://unstats.un.org/sdgs/report/2020/goal-13/> [Accessed March 2023].

Verma, Y. 2021. Complete Guide To SARIMAX in Python for Time Series Modeling. Available from: <https://analyticsindiamag.com/complete-guide-to-sarimax-in-python-for-time-series-modeling/> [Accessed April 2023].

Wani, S., Yadav, A. A., Panchal, M. M. & Pandey, P. V. 2022. Predicting CO<sub>2</sub> Emission Using Machine Learning. International Journal for Research in Engineering Application & Management (IJREAM), 08, 84-88.

Xu, Z., Liu, L. & Wu, L. 2021. Forecasting the carbon dioxide emissions in 53 countries and regions using a non-equigap grey model. Environmental Science and Pollution Research, 28, 15659–15672.

Yu, D., Soh, W., Noordin, A. A., Yahya, H. D. H. & Latif, B. 2022. The impact of innovation on CO<sub>2</sub> emissions: The threshold effect of financial development. Front. Environ. Sci., 10.

Yusuf, A. M., Abubaker, A. B. & Mamman, S. O. 2020. Relationship between greenhouse gas emission, energy consumption, and economic growth: evidence from some selected oil-producing African countries. Environmental Science and Pollution Research, 27, 15815-15823.

Zareba, M., Danek, T. & Stefaniuk, M. 2022. Unsupervised Machine Learning Techniques for Improving Reservoir Interpretation Using Walkaway VSP and Sonic Log Data. Energies, 16.

Zhong, W. & Haigh, J. D. 2013. The greenhouse effect and carbon dioxide. Weather, 68, 100-105.



## Appendix

### Dataset

Variable	Description	Variable Type
Area	Country	String
Item	Source of Emission	String
Element	Type of Emission	String
Unit	Emissions = Kilotonnes	String
Year	Total Emissions for Each Year, 2000 - 2020	Float

## Research Proposal

IMAT5314 Project Terms of Reference (ToR)

Student Name: Giwa-Daramola Inioluwa

P-number: P2712256

Programme: MSc Data Analytics

Email address: P2712256@my365.dmu.ac.uk Project

Title: Predictive Modelling for Emissions per country using machine learning.

Supervisor: Sean Xavier Laurence

Email: [sean.laurence@dmu.ac.uk](mailto:sean.laurence@dmu.ac.uk)

### 1.0 INTRODUCTION

Climate change is one of the greatest challenges facing humanity in the 21st century. According to the National Aeronautics and Space Administration (NASA), the earth's global surface temperature in 2020 was the second-highest on record since 1880 (NASA, 2020). The Earth's climate has been changing throughout its history, but the current warming trend is of particular

concern due to its rapid rate and human-induced causes. From rising sea levels to more frequent and severe extreme weather events, the impacts of climate change are already being felt around the world (Shi, 2018). As such, monitoring climate change is crucial in order to understand its impacts on ecosystems, economies, and societies, and to develop effective strategies for mitigating and adapting to these impacts. The need to identify the sources of greenhouse gas emissions and predict future emissions has become increasingly important in order to mitigate the impacts of climate change (Deetman et al., 2020). One approach to this problem is to use predictive modelling and machine learning techniques

Predictive modeling, involves the use of machine learning algorithms to make predictions hinged on historical data. This technology presents a workable solution to this problem (Lakshay and Pratika, 2017). The analysis of historical data on emission changes, patterns, trends and other relevant factors can aid the creation of predictive models that can generate accurate predictions of future emissions (Fatimetou, 2017). These predictions can then be used by policymakers to design effective policies and interventions to reduce emissions at a global scale.

This study aims to adopt a comprehensive approach that traverses various regions and continents to improve our understanding of emission patterns. By using machine learning techniques, researchers can identify complex patterns and relationships between different variables that may not be immediately apparent using traditional statistical methods.

### 1.1 Background of the study

Climate change is a global challenge that has become increasingly urgent to address. Greenhouse gas emissions, primarily carbon dioxide (CO<sub>2</sub>), are the main contributors to global warming and climate change (Shi, 2018). According to the Intergovernmental Panel on Climate Change (IPCC), the global average temperature has increased by 1.1°C since pre-industrial times, and it is projected to continue to increase in the coming decades if emissions continue to rise at the current rate (IPCC, 2021).

One way to address this challenge is to reduce greenhouse gas emissions. To achieve this, it is important to understand the sources and drivers of emissions, as well as the factors that influence them. Countries play a crucial role in this regard, as they are responsible for a significant proportion of global emissions. In 2019, global CO<sub>2</sub> emissions from fossil fuels and industry were about 36.4

billion tonnes, and the top five emitters (China, United States, India, Russia, and Japan) accounted for more than half of these emissions (Friedlingstein et al., 2020).

Some projects such as the Forecasting of transportation-related energy demand and CO<sub>2</sub> emissions in Turkey with different machine learning algorithms by Umit Agbulut, and Machine learning-based time series models for effective CO<sub>2</sub> emission prediction in India by Surbhi Kumanr and Suni Kumar Singh have done project related to the prediction of CO<sub>2</sub> emissions in India and Turkey. Kumari and Singh used three statistical models which were the autoregressive-integrated moving average (ARIMA) model, the seasonal autoregressive-integrated moving average with exogenous factors (SARIMAX) model, and the Holt-Winters model, and two machine learning models, i.e., linear regression and random forest model and a deep learning based long short-term memory (LSTM) model. Agbulut used Deep Learning (DL), Support Vector Machine (SVM) and Artificial Neural Networks (ANN). An exploration of clustering and classification machine learning algorithms will inform new ideas and aid the creation of predictive models.

Climate change is a pressing global issue with significant implications for ecosystems, economies, and societies around the world. Monitoring climate change is essential for comprehending its effects and creating efficient mitigation and adaptation methods (Haines et al., 2006). Through continued research and data collection, we can work towards a more sustainable future for ourselves and for future generations.

## 1.2 Statement of the problem

The most recent past decade, which includes year 2011 through year 2020 was observed to be the hottest decade. With current trends, global temperatures are expected to rise by another 1.5°C between 2030 and 2052, which would lead to more frequent and severe climate-related disasters (IPCC, 2021). The release of emissions such as CO<sub>2</sub>, CH<sub>4</sub>, or N<sub>2</sub>O etc., into the atmosphere as a result of human activities has significantly contributed to current global temperatures, negatively affecting the health of individuals directly and indirectly (Patz et al., 2005). This creates the need for a level of urgency when dealing with this global challenge. Governments, organizations and global stakeholders have outlined some best practices which can contribute to the reduction of gas emissions and its effects, however, accurate and reliable information on the country-by-country

emissions is needed for setting and achieving targets which reduce the emission of greenhouse gases; which are the main drivers of climate change.

With climate change comes negative occurrences which are already being observed in different countries, all around the globe. Rising sea levels, due to melting ice sheets and glaciers, are causing more frequent and severe flooding and storm surges, which threaten coastal communities and infrastructure. Changes in temperature and rainfall patterns are also affecting food security, with some estimates suggesting that global crop yields could decline by up to 25% by 2050 (Pathak et al., 2010). Additionally, climate change is exacerbating other global challenges, such as biodiversity loss, water scarcity, and public health risks.

Given the significant impacts of climate change, it is essential to monitor its effects in order to understand the scope of the problem and develop effective solutions. This requires collecting and analyzing data on a range of indicators, such as temperature, precipitation, sea level, and greenhouse gas concentrations. By monitoring these indicators over time, researchers can track changes in the climate system and identify areas where intervention is needed to lessen greenhouse gas emissions and encourage adaptation to the effects of climate change.

### 1.3 Aims and Objectives

The main aim of this project is to identify the causes of the country level emission patterns, and build solutions which can better forecast expected emission trends. This will aid with the policy making and decision taking.

1. Build predictive models to forecast future emissions trends by location.
2. Analyze the changes in emissions across different countries.
3. Present an analysis of emission trends by types (e.g., CO<sub>2</sub>, CH<sub>4</sub>, or N<sub>2</sub>O).
4. Developing an understanding of the detrimental effect of the different types of emissions based on predictions for the next 10 years.

### 1.4 Research Questions

1. How can machine learning models be effectively used to forecast future emissions trends?
2. What emission disparity can be observed amongst different countries?

3. How do different emission types (e.g., CO<sub>2</sub>, CH<sub>4</sub>, or N<sub>2</sub>O) vary across countries, and what factors contribute most to these differences?
4. How can we explain the detrimental effect of the different types of emissions based on predictions for the next 10 years?

### 1.5 Project Structure

This introduction section of the project highlights the goal of this project. It identifies the aim, objectives, as well as the research questions which will be answered during the course of this research. An in-depth review of literature which focuses on climate change, green gas emissions and machine learning algorithms that will aid predictive modelling will be discussed in the Literature Review. The Methodology section will highlight all methods used to achieve my goal, while the Result section will describe the outcomes observed during this study.

Finally, key findings will be concluded and summarized in the Conclusion section of the project

### 1.6 Conclusion

The scope and impact of climate change on the globe is vast, creating the need for diverse solutions which tackle this problem at a global scale.

# Ethics Application Form



## Research Ethics Screening Questions

This form will help you establish whether an application for research ethics approval is required. If you are unsure, please discuss with your research supervisor or Faculty Research Ethics Committee.

**Name:** GIWA-DARAMOLA INIOLUWA

**Project Title:** PREDICTIVE MODELLING FOR EMISSIONS PER COUNTRY USING MACHINE LEARNING

### Screening

PART 1: Screening Question	Further Guidance	Yes	No
Does your research involve accessing, storing or disseminating security sensitive information such as that relating to extremism or terrorism?	For further details, see DMU's policy on security sensitive research.	<input type="checkbox"/>	<input checked="" type="checkbox"/>

**If 'yes', please contact your Faculty Research Ethics Committee (see below)**

**If 'no', move onto Part 2.**

PART 2: Screening Question	Further Guidance	Yes	No
Does your research involve only secondary data that are: a. Curated for public access, or b. Data for which consent was provided for reuse in research beyond that for which it was originally collected, and those data are anonymized?	Examples of data already in the public domain include those curated for public access and display or for re-use by researchers, material held in archives or private collections. Refer to the Research Ethics Code of Practice for further details regarding secondary data.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Is your project exclusively based on published literature or library/archival materials which have been specifically curated for general public access or display?	Examples include literature reviews, meta-analysis or re-use of published data.	<input checked="" type="checkbox"/>	<input type="checkbox"/>

**If 'yes' to any of the questions in Part 2, an ethics application IS NOT required. See actions below:**

- **Staff:** Please retain a copy of this form in your project file.
- **Post-graduate research students:** Please confirm the declaration at the end of this form. Your supervisor will then need to forward the form to the Faculty Research Ethics Committee and Doctoral College.
- **Undergraduate/Taught Postgraduate students:** Please confirm the declaration at the end of this form. Your programme team will inform you what to do next.

**If 'no' to both of the above questions, please proceed to Part 3 (next page).**

## Research Ethics Screening Questions

PART 3: Screening Question	Further Guidance	Yes	No
Does your research involve secondary data not covered by Part 1?	This may include identifiable data that are not in the public domain, or where consent was not provided for reuse of data beyond the purposed for which it was originally collected.	<input type="checkbox"/>	<input type="checkbox"/>
Will your research involve human participants?	<p>This includes any procedures or activities (including questionnaires, surveys, interviews and focus groups) involving living participants and the use of data or tissue obtained from living individuals, whether directly or indirectly. This includes data about people obtained from databases, public websites and social media, unless held in publicly available collections or from archaeological sites and curated explicitly for the purposes of research. This may also include the use of data and tissue obtained from deceased individuals.</p> <p>Tissue (or material derived from human tissue e.g. cell lines) which has been obtained from commercial sources will not normally require ethical review. Please seek advice from <a href="mailto:ethics@dmu.ac.uk">ethics@dmu.ac.uk</a>.</p>	<input type="checkbox"/>	<input type="checkbox"/>
Will your research involve the use of animals?	This includes any research involving live animals (e.g. intervention, manipulation of conditions or observation) and animal tissue. Tissue (or material derived from animal tissue e.g. cell lines) which has been obtained from commercial sources will not normally require ethical review.	<input type="checkbox"/>	<input type="checkbox"/>
Does this project involve the use of sensitive or restricted materials?	<p>Please refer to the <a href="#">Sensitive Research Policy</a>. Research covered by this policy includes:</p> <ul style="list-style-type: none"> <li>• Research into illegal activities, including the collection of source data, e.g. crime statistics;</li> <li>• Research which requires access to web sites normally prohibited on university servers; including, but not limited to; pornography, or the sites of any of the organisations proscribed by the UK Government.</li> <li>• Research into extremism and radicalisation.</li> </ul>	<input type="checkbox"/>	<input type="checkbox"/>
Does the project have the potential to cause environmental damage or harm?	This includes the use of equipment, materials, techniques or research results that could have an impact on the environment, ecology, hydrology or on local communities during or after the research project.	<input type="checkbox"/>	<input type="checkbox"/>
Does your project require external ethical review?	<p>External review will be required for research involving NHS patients, staff, data or facilities (includes contractors providing services under contract with care services or commissioners).</p> <p>Further examples include:</p> <ul style="list-style-type: none"> <li>• UK Social Care organisations or service users (includes contractors providing services under contract with care services or commissioners).</li> <li>• Clinical Trials of Investigational Medicinal Products, Medical Devices or Gene Therapy Medicinal Products.</li> <li>• The prison service, offenders or participants on probation. Any member of the research team or participants who are members of the Armed Forces or their entitled dependents.</li> <li>• The administration of ionising radiation.</li> <li>• Participants aged 16 or over who may lack capacity to give fully informed consent or who may lose capacity to give</li> </ul>	<input type="checkbox"/>	<input type="checkbox"/>

## Research Ethics Screening Questions

	<p>fully informed consent during the study.</p> <ul style="list-style-type: none"> <li>• Research regulated under the Animals (Scientific Procedures) Act 1986 (ASPA).</li> <li>• Research that may be regulated under the Human Fertility &amp; Embryology Authority</li> </ul> <p>Please seek advice from <a href="mailto:ethics@dmu.ac.uk">ethics@dmu.ac.uk</a> before proceeding any further if your research involves any of the above.</p>		
--	--	--	--

**If 'yes' to any of the Part 2 questions, an ethics application IS required.**  
Please submit your application using the WorkTribe ethics module.

### Declaration (for students use, where required by their programme)

<p>I take full responsibility for the information provided in this application, and:</p> <ol style="list-style-type: none"> <li>An ethics application <u>IS</u> required. I confirm I will not commence research before receiving a favorable ethics opinion, OR;</li> <li>I am satisfied that an ethics application is <u>NOT</u> required. I have sought the advice as necessary, and will monitor my research as it progresses and reassess the need for an ethics application periodically.</li> </ol> <p>If you cannot confirm the above, please contact your Faculty Research Ethics Committee.</p>	<input type="checkbox"/>
As the student's supervisor, I am satisfied with the outcome of this assessment.	<input type="checkbox"/>

### Faculty Research Ethics Committees:

Faculty of Business and Law (BAL)	<a href="mailto:BALResearchEthics@dmu.ac.uk">BALResearchEthics@dmu.ac.uk</a>
Faculty of Arts, Design and Humanities (ADH)	<a href="mailto:adhethics@dmu.ac.uk">adhethics@dmu.ac.uk</a>
Faculty of Computing, Engineering and Media (CEM)	<a href="mailto:ethics.CEM@dmu.ac.uk">ethics.CEM@dmu.ac.uk</a>
Faculty of Health and Life Sciences (HLS)	<a href="mailto:hlsfro@dmu.ac.uk">hlsfro@dmu.ac.uk</a>