Effects of Carbon Dioxide (CO2) Emissions on People's Death and Global Warming

MSc Research Project
Data Analytics

Jose Geo Vattolly
Student ID: x22139508

School of Computing
National College of Ireland

Supervisor: Taimur Hafeez

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	…….………….Jose Geo Vattolly…………………………………………………………
Student ID:	……………………x22139508……………………………………………………………..……
Programme:	…………MSc..Data Analytics…………………….… **Year:** ………2023……….
Module:	………………………MSc Research Project…………..…………………….………
Supervisor:	………………………Taimur Hafeez…………..……………………………….………
Submission Due Date:	…………………………14th December 2023…………………………………….………
Project Title:	…Effects of Carbon dioxide(CO2) Emission on people Death and Global warming……..………
Word Count:	…………8000+..…………… **Page Count**……………………19……………….……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ……………………………Jose Geo Vattolly……………………………………………

Date: …………………………14th December 2023………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	☒
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☒
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☒

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Effects of Carbon Dioxide (CO2) Emissions on people's Death and global warming

Jose Geo Vattolly
x22139508

Abstract

A thorough study forecasting and analyzing the most important environmental indicators around the world are presented in this report, backed up by state-of-the-art Machine learning and Deep Learning technologies. The new strategy, based on the forecasting of anomalies, emissions of CO2, and premature death due to pollution, uses cutting-edge tools for dealing with a complicated relationship between environmental factors and human health. Our model has a wide variety of information from across continents so that it can form robust forecasting frameworks. Using machine learning, we've been able to predict accurately global anomalies and provide insight into climatic fluctuations and their possible consequences. In addition, we can find out in detail what regions are contributing to environmental degradation through our deeper learning algorithms which precisely predict CO2 emissions from all continents. By projecting the rate of mortality attributable to pollution, this study has expanded its scope into the field of health. By incorporating advanced Deep Learning architectures in our algorithms, we are able to identify complex patterns that link pollution levels with negative health impacts and give important information on the use of Public Health measures. The findings of this study will help us better understand global dynamics in the environment, and give a platform to inform policymaking. Our models' ability to predict and contribute to the development of tailored actions for environmental protection can be helpful in societies that are facing climate change and pollution challenges.

Keywords: ***Global Anomalies, CO2 Emissions, Pollution-Related Death Rates, Machine Learning, Deep Learning***

1 Introduction

In the 20th century, climate change is becoming a global crisis of historic proportions: rising levels of carbon dioxide CO2 emissions and their profound impact on human health as well as Earth's climatic stability. CO2 levels in the atmosphere are rising to dangerous levels, creating a chain reaction of repercussions that will spread across an intricate web of ecosystems and civilizations as global populations grow and economic activity expands. The initiative is aimed at exploring the internal dynamics of CO2 emissions, examining their impact on people's health and global warming as a whole.

In order to maintain a planet's delicate climate balance, CO2 is the most important natural component of Earth's atmosphere. However, this equilibrium has been broken by an acceleration of theincrease in anthropogenic CO2 emissions, mainly as a result of industrial development and energyuse. This imbalance has two effects: it affects people's health on a micro level and contributes to climate change at the macro level.

The effects of the higher CO2 content are beginning to be seen in a microscale, affecting human health. Owing to inadequate ventilation and the use of fossil fuels for energy, higher CO2 concentrations in internal environments are associated with numerous adverse effects on health. Individuals are unknowingly exposed to a cocktail of contaminants that risk their health, ranging from respiratory difficulties to cognitive impairment. This study will examine the extent to which such health consequences are complex, revealing some of the often overlooked relationships between indoor air quality and CO2 emissions as well as human health.

In parallel, there are significant effects on the world's climate with continued increases in atmospheric CO2 concentrations. CO2 is a powerful greenhouse gas that traps heat in the Earth's atmosphere and contributes to the enlargement of the greenhouse effect. As a result, global warming leads to various and often terrible consequences such as the melting of polar ice caps or worsening severe weather events. Through a systematic analysis, the research aims to find out how human CO2emissions are intimately linked with looming climate change threats and what impacts these will have on ecosystems, biodiversity, and Earth's overall stability.

To develop realistic measures to reduce the negative impacts of CO2 emissions, which are at a pivotal point in human history, it is necessary to understand both its own effects and those of other pollutants. The aim of this project is to provide a more accurate understanding of the complicated interactions between CO2 emissions, human health, and climate stability by bridging gaps in science knowledge and public awareness. We shall seek to lead the way towards sound practices, policies, andcooperative action which can safeguard both human well-being and our planet's climate balance.

1.1 Research Questions and Objectives:

This research seeks to address the following interrelated questions:

Research Question 1: To what end do heightened CO2 emissions contribute to changes in mortality rates across diverse countries and regions?

Research Question 2: what is the spatial and temporal distribution of CO2 levels across continents and how do these variations affect regional climate patterns?

Research Question 3: Can predictive models be used to develop forecast temperature anomalies resulting from elevated CO2 emissions, and how do they contribute to global climate change?

2 Related work:

2.1 Related work on predicting global warming:

According to Lin et al. (2023),In this study, ARIMA and SVM models are used to forecast the level of global warming, using the data of the previous history. The performance of the ARIMA model is better than the SVM model and the relation between the temperature, time and greenhouse gases such as N2O, SF6, CH4, and CO2 is between (0.6-0.7), this gives light to the greenhouse gas emissions caused due to global warming by humans. The results show us that excess use of natural resources by man influences the earth's biological environment and greenhouse gas emissions are increased, which is the main cause of climate change, The study also shows that after the industrial revolution, there has been a drastic in CO2 concentrations which where consistent with the established global temp trends. This can be used to develop climate change adaptations and mitigation policies at local, regional, and global levels. Similarly, Santos et al., (2022), the authors performing the study using an ARIMA model to axis the paleo climate info from about twelve thousand years and it is compared to the current worldwide warming cost by humans. Their study is likely to call into the general belief in human climate change. A face of temp decline is shown in the ARIMA model which contradicts IPCC's projections for worm temp. The study also shows the need for the scientific agreement to be measured especially when there is intra-factional dissension and gives a clear picture of the fragility of the anthropogenic warming theory. In this study, the necessity is marked to know the climate variability and a balanced view of the climate policy which is taken account of different science views.

In this research by Liu et al. (2023), The ARIMA (auto-regressive integrated moving average) is extremely used to analyze the study paper of global warming projections. It shows the importance of data smoothness which is proven by an augmented dickey-fuller (ADF) test to establish stationarity which is the peak condition for using ARIMA. The ARIMA (0,1,1) was chosen for its capacity for the rejection of the null hypothesis of nonstationary and its statistical significance. Later the model is used to anticipate the world's average temperatures for the years 2050 and 2100.The previously forecasted temperatures of 10.6 degrees centigrade in 2050 and 11.78 degrees centigrade in 2100 provide crucial insight into the upcoming temperature patterns. Model parameters such as alpha and gamma are presented which helps in the study of Global warming and its consequences. When the ARIMA model accuracy is compared to other models like GM (1,1) and Holt, it becomes crystal clear how important it is in climate research. These forecasts give confidence to academics and policymakers in assessing and responding to the issues caused by global warming.

In this survey Dossa & Miassi, (2023), the ARIMA model plays a very crucial role in the estimation of maize quantities within Togo. If the climate and population density have no variation over the following decade, it forecasts a uniform maize quantity, this tells that the model is reliable and shows the negative effect of high rainfall on maize yield which has been influenced by increased rainfall intensity. The temperature and level of N2O emissions are constant without significant changes over the same time period. Agricultural practices such as the use of chemical fertilizers have a good relationship with N2O and maize yield. Urban population density has a negligible effect on maize production, but the rural population density is expected to decrease, reflecting the phenomenon of rural migration. A quality indicator is used to measure the accuracy of the model ARIMA and its conclusions will have an impact on Togo's climate change adaptation and food security strategy. The study by Sameh & Elshabrawy, (2022) shows the ARIMA and SARIMAX models for forecasting climate change time series. ARIMA is an univariate model that incorporates previous data to anticipate climate-related trends. Mathematical foundations visualization and data preparation are all part of the study. ARIMA's performance is assessed using ACF, PACF, and lag plots, and its appropriateness is determined by the auto ARIMA function. The ARIMA model has a mean absolute error of 1.1388007203766655 and a mean squared error of 1.747035631319338, making it one of the best models for predicting temperature. In the study by Masum et al., (2022) ARIMA model was used in Chattogram, Bangladesh from 1953 to 2070 in the study to forecast rainfall and temperature trends. The ARIMA model is good at predicting, with average stationary R squared values of 0.952 for rainfall and 0.884 for temperature. Rainfall is increasing during the monsoon season, with a significant increase expected in later stages. Temperature, on the other hand, follows a rising trend, particularly between 2021 and 2050, before declining. The ARIMA model is used for predicting climate parameters, giving insights into shifting patterns of rainfall and temperature in the context of climate change.

2.2 Related work on predicting Air pollutants mainly due to Carbon dioxide:

In this research Zhu et al. (2022), indoor air quality in bedrooms was monitored focusing mainly on the effect of CO2 level on human health IOT-based machine was used with CO2 sensors in which MQTT is used for real time data conversion in a useful way. To forecast the CO2 level the researchers used long short-time memory (LSTM) which is a technique in deep learning. The system made a 5.5 percent error in computing CO2 in this study state, a crucial aspect for health and safety. The LSTM technique has performed well with an accurate R-square value of 0.981. The work in the future is expected to focus on temperature and the amount of humidity to improve the result of the LSTM model. The research by, Amarpuri et al., (2019) is done to predict the amount of CO2 in India. To predict the CO2 level LSTM and CNN(convolutional neural networks) models were used to forecast the emission of CO2 that has been raised in India and is expected to increase up to 2915 tones by the year 2020. The combination of LSTM and CNN has performed better in exponential smoothing methodology with the root mean square error (RMSE) of 52.06 and mean absolute percentage error(MAPE) of 1.49. This research has shown an effective method to forecast the CO2 emission in India and the upcoming future research is expected to find the connection between the Co2 emission and factors that affect the production of more amount of carbon. This study Zhou et al., (2022) aims to search approaches and frameworks in the field of carbon pricing prediction and evaluate whether they are effective within a range of environment trading schemes. The combination of VMD and LSTM methods with a long computation time provides the best results, though it is better for accurate predictions than individual VMD, and LSTM methods. This study gives light on the relevant integrating sample entropy which confirms LSTM as a better reliable forecasting model than GRU and DRU. The CEEMDAN forecast shows that min-max was appropriate to forecast the normalization of carbon prices, out spaces both EMD and EEMD. The research shows a discussion of potential improvements through parameter optimizations, noble models such as temporal convolutional methods (TCN), and clustering approaches such as K-means.

According to Kumari & Singh. (2022), The study evaluates their effectiveness by measuring nine metrics with the goal of predicting CO2 emissions based on a combination of statistics, machine learning, and deep learning models. The best models are LSTM, SARIMAX, and LH Holt Winters all of which outperform each other in different criteria. The LSTM model has been confirmed as having the greatest success by a statistical study involving, training Friedman's test. The report calls on governments to introduce carbon taxes, environmental protection standards and support for greener technologies with a view to reducing emissions study provides important insights and it should be followed up by future studies into the effects of outside factors on emissions. Comparisons with other study shows that the LSTM model has a higher ability to predict. Thus the study encourages productive steps to reduce CO2 emissions and gives options for future multivariant studies. Similarly, Han et al., (2023) This study compared the model for predicting CO2 concentrations in cars, ARIMA, and LSTM. The chosen LSTM architecture shell used one or two layers of LSTM. Both models accurately predict the CO2 variability during testing and ARIMA was set to be more efficient than all of them. The mean MAPE and RMSE values are lower for ARIMA compared to LSTM. Desperate LSTM success at capturing complex non-linear patterns, training data are hard to complete, the ARIMA system is a scope. The study demands that it is important to choose models according to data volumes and recommends the use of LSTM for more exclusive data sets as well as ARIMA for smaller ones. Thus, the objective of future research is to take predictive functions in vehicle ventilation control systems into account.

According to Singh & Kumari, (2022) In this survey, the research specialist has used machine learning, deep learning models, and statistical models to predict CO2 emission in India. The researchers have compared nine models to compare the performance of each model. ARIMA model forecast the CO2 level which has deviated from actual predictions from the year 2020 to 2030. The LSTM model has performed way better than the other models which is forecasted to the actual prediction making it a useful tool to predict the CO2 level with the mean square error of 3676.646 and the R square score of 0.99. Thus the survey helps the policy makers control the emission of CO2 levels to meet the UN-specific reduction measures. Similarly, Parvez et al., (2023) This research was done to forecast the CO2 emissions in Bangladesh from the year 2022 to 2024 based on the history of 1946 to 2021. The research specialist used the ARIMA (1,2,1) mode to predict the CO2 level and a result of 0.00922446 root mean square error and mean absolute error of 0.00609683 is obtained. The survey underscores the model's significant need for both short-term and long-term planning to control the emission of CO2 levels in the days to come in the atmosphere. In the study by Spyrou et al., (2022), the authors use a univariate LSTM model with variable batch sizes(100, 1000, 7000) to predict carbon monoxide levels in this study showing air quality in the port of Lgoumenitsa, Greece. Notably, batch size 7000 outperforms all others in terms of Root Mean Square Error(RMSE) and mean absolute error(MAE) in both the training and testing stages. The forecasts of the LSTM are compared to the ARIMA model, with the ARIMA model showing slightly superior results. Future research includes bidirectional LSTMs, a multivariate approach with added environmental elements, and the integration of data from different stations for a more detailed study, according to the scientists. The study results shed light on the potential of LSTM models for airquality prediction and also highlight areas for additional refinement and development.

The research by Javanmard et al., (2023) provides multiple objectives mathematical model that includes various machine learning methods in the Canadian transportation industry until 2048 to forecast energy demand and CO2 emissions. The report deals with the transportation sector's substantial position as a major energy consumer and CO2 emitter. For forecasts, many algorithms such as AR, ARIMA, SARIMA, MIDAS, and WOA are used. Individual algorithms are performed better than the integrated model, which achieves the MAPE index value of 0.003849 for energy demand and 0.00038 for CO2 emissions. Sensitivity analysis discloses the different effects of oil, gas, electricity, and renewable energy consumption on CO2 emissions, giving valuable insights for policymakers in developing successful emission-reduction policies.

The research by Kamoljitprapa & Sookkhee, (2022) offers ARIMA models for forecasting CO2 emissions in Thailand's economic sectors, with effective model fitting and prediction outcomes. Optimal models for total emissions, such as ARIMA(1, 1, 1)(1, 0, 0), have good accuracy, with R2 values surpassing 0.98. Residual analysis demonstrates model adequacy by meeting independence and normalcy criteria. Based on projected yearly net CO2 emissions of about 264,839 kt over the next decade, the report emphasizes the urgent need for the Thai government to address imminent climate consequences, emphasizing the significance of switching to greener energy technologies.

2.3 Related work on predicting the Death rate due to pollution:

According to Sridhar et al., (2023) the research was done on the Internet of Things sensing technology for predicting pollutant gases in developing countries experiencing massive climate problems. This was done using a mixed-learning approach to efficiently calibrate cheap air quality sensors. With the help of GSM/Wi-Fi technology, the IOT node, which is outfitted with a number of sensors delivers the actual air quality reports. The suggested platform displays big data analysis efficiency such as weather predictions by recording signals by the sensors at a rate of one sample in one second. The motto of the research is to develop low-price sensor equipment for environmental intelligence applications. Implementing algorithms such as NB, KNN, NN, and SVM performed better than the existing techniques in performance evaluation demonstrating the systems' predictive capabilities. Similarly Lee et al., (2021)The Random Forest Classifier manufactured with the Scikit-learn module is used in the paper. It includes static demographic/clinical parameters combined with dynamic vital sign data to predict preterm newborn death.

This study by Kim & Kim, (2022) shows a spatial unit random forest (RF) model for calculating mortality in a city caused due to heat, and it also includes mortality caused due to climate, demographic, and socioeconomic factors. Model evaluation describes that the grid search strategy improves the F1-score, accuracy, and area under the curve (AUC). The RF model with grid search has 90.3% accuracy, an F1-score of 94.75% accuracy, and an AUC is more than 0.8%. When compared to the previous research in urban disaster prediction, this study is said to be better than others. The significance of demographic and socio-economic data gives light to SHAP analysis, with young and elderly (YE) and aging ratio(AR) leading to mortality estimates. This study gives recommendations for urban heat adaptation measures, giving light to the significance of socially vulnerable populations which is taken into account. The research by Schachtschneider et al., (2023) uses an Echo State Network (ESN), which is a type of recurrent neural network, to forecast death rates in Germany based on climate storyline simulations. The ESN is trained using current climate model output and makes forecasts using future, warmer climate scenarios. This study's success confirms the ESN's Ability to forecast the fatality rate, with an increase during the summer owing to more severe heat waves. Assembled projections support the ESNs reliant on standard deviations which range from 0.5% to 1% of the target. Despite the fact that the ESN's climate structure limits interpretability, the algorithm short inputs data time series givesthe correct and accurate findings which indicate its use in simulating the link between temperature inputs and death rates. Similarly, the study by Singh et al., (2023) shows the complex link between lung cancer incidence rates and the Air Quality Index (AQI). The results show a substantial positive link between high AQI readings and a high risk of lung cancer. The suggested machine learning model which has been using XGBoost Regression, has the highest accurate range of 81-98%, followed by Random Forest being the second most accurate(79-97%). Due to potential confounding factors like smoking and occupational exposures, the study highlights the difficulties in proving causation in observational research.

3 Methodology:

3.1 Data collection and understanding:

Data collection and implementation of the methodology is an important aspect of any kind of data-related project. The collected data should be authentic.

3.2 Carbon dioxide (CO2) emissions:

Our World in Data website contains all kinds of records such as global problems in health, violence, education, energy, and human impact on the earth. The Carbon dioxide(CO2) record was collected from ourworldindata.org. In the Carbon dioxide dataset the CO2 emission of all countries and continents from the year 1949 to 2021.

3.3 Death rate data:

The death rate data was collected from Kaggle.com, which contains all the death rates of each continent and countries that died due to pollution risks. The death rate record contains the collection of data of all continents and countries from the year 1990 to 2019.

3.4 Global Temperature Anomalies Data:

Global temperature anomaly data is also required for this research to predict the temperature change due to pollution produced by human activities. The temperature anomalies dataset was collected from the data world website. The temperature anomalies dataset contains the record of the average highest temperature of every year from 1880 to 2020.

3.5 Data preprocessing:

Data Preprocessing is the process of transforming the data in a condition that it is fit to apply machine learning and deep learning models. It involves different stages such as removing the missing values or putting the mean, median, or mode values in place of the missing values, resolving the white spaces, performing calculations, and splitting the data into testing and testing sets.

The carbon dioxide record gathered has three columns one of them contains the name of the country or continent, the second column contains the year of carbon dioxide emission, and the third column contains the amount of carbon produced every year. This data was pre-processed by checking the null values and handling the outliers. Then the data was transformed in two ways one with log transformation to apply ARIMA and LSTM models and the second by square transformation to perform only ARIMA. For the death rate prediction due to carbon dioxide, the carbon dioxide data was combined with the death rate and all the unwanted columns such as “Deaths - Cause: All causes - Risk: Diet high in sodium - Sex: Both - Age: All Ages”, “Deaths - Cause: All causes - Risk: Diet low in whole grains - Sex: Both - Age: All Ages” and “Deaths - Cause: All causes - Risk: Air pollution - Sex: Both - Age: All Ages” were considered to predict the death rate in each country. The final column to predict the death rate due to pollution is Entity which contains the country and continent names, Code_Xcontains the country code in three alphabets, Year which contains the years of death rate and carbon dioxide emission, Lives Affected represents the number of deaths due to Air pollution, and the data of Annual Co2 emission and One Hot Encoder column transformation was applied on Entity column, and three machine learning models called Random Forest, XGBoost and Linear Regression was applied and a deep learning model was applied to it.

To predict the average global temperature anomalies the dataset contains twenty columns Hemisphere, Year, Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec, each column contains the average temperature anomalies of the month and all the months columns was set in a single column and the year was converted to date and the year column was converted to date format. Then the dataset was checked whether it was stationary or not as the Seasonality, Trends, Residuals, and ACF(Autocorrelation Function) and PACF(Partial Autocorrelation Function) graphs were plotted to find the correlation between time series and lags. Then the time series methodology ARIMA was applied to predict the future average temperature anomalies of the earth.

3.6 Ethical and legal compliance:

The Our World in Data collection on CO2 emissions from 1949 to 2021 is an essential resource for studying global emissions trends. Ethically, it is essential to use these data in a way that avoids misleading techniques and keeps the interpretation as transparent as possible. The privacy and sensitivity of emission data from individual countries need to be protected. The rights of copyright and intellectual property associated with the data set should be taken into consideration in order to ensure that Our World in Data's terms of use are respected.

A dataset from Kaggle, covering 1990 to 2019, contains helpful information about pollution- related mortality rates. Ethically, the proper handling of such sensitive information, the avoidance of sensationalism and the disclosure of any biases or limits in the data sets are essential. It is a legal requirement for scientists to follow Kaggle's terms and conditions of use, while ensuring the right to access and analyze this data in order to achieve their research objectives. The Global Temperature Anomalies dataset, obtained from data. The world is divided into three periods, from 1880 to 2020, that are important for anticipating climate change caused by pollution. The law requires transparency in relation to data sources and recognition of contributors. In additionto addressing uncertainty in the temperature anomaly data, researchers should be open and transparent about their methodology. Compliance with data. The legal requirements for the world's terms of use, related licensing arrangements, and adherence to copyright and intellectual property laws regarding third-party content contained in these data sets are necessary.

3.7 Model selection:

3.7.1 Carbon Dioxide Emission Data:

1) **ARIMA (Auto Regressive Integrated Moving Average):**
   Reasoning:
   - ARIMA is a forecasting model for time series data that captures linear correlations.
   - Suitable for the prediction of future values in line with historic patterns.
   - The variance, which is the most commonly needed feature of ARIMA, can be stabilized by a log transformation.
2) **LSTM (Long Short-Term Memory):**
   Reasoning:
   - (Long Short Term Memory) LSTM is a type of recurrent neurological network, which makes it very effective at collecting temporal Dependencies from an input stream consisting primarily of synchronous inputs**.**
   - It is suitable for the forecast of time series with complex patterns and dependencies.
   - Transformation of logs has been used to account for the scale of carbon dioxide emissions.
3) **Square Transformation for ARIMA:**
   Reasoning:
   - To meet the assumptions of the ARIMA model, it can be decided to use a linear transformation according to data properties.
   - In the case of situations which would have been more appropriate with a quadratic relationship, it could be used.

3.7.2 Prediction of Death Rate Due to Carbon Dioxide:

1) **XGBoost, Random Forest, Linear Regression:**
   Reasoning:
   - Complex correlation of features and interactions can be detected by ensemble models like Random Forest or XGBoost.
   - The Linear Regression will be used, given that it is capable of revealing important information for simpler parallel relationships.

- These models shall be applied to estimate death rates, based upon the combination of carbon dioxide and mortality data.

2) **Deep Learning Model:**
   Reasoning:
   - Deep learning techniques, such as neural networks, are an effective method of detecting complex patterns in data.
   - They're being used to predict death rates, potentially addressing non linear correlations that traditional models may not be able to detect.

3.7.3 Prediction of Average Global Temperature Anomalies:

1) **ARIMA (AutoRegressive Integrated Moving Average):**
   Reasoning:
   - In the future, ARIMA will be used to forecast temperature anomalies.
   - When the data is stationary, it is suitable for time series forecasting.
   - The assessment of stationarity is performed using ACF, PACF, and Graphs.
   Stationarity Check:
   - By comparing autocorrelation and partial autocorrelation the analysis of ACF and PACF is useful in establishing stationarity.
   - In order to ensure that these data are accepted by the ARIMA system, seasonality, trends, and residuals shall be considered.

3.7.4 Model Training:

A combination of conventional and modern methodologies is used for the modeling of CO2 emissions, death rates from CO2 as well as worldwide temperature anomalies. In the case of carbon dioxide emissions, both ARIMA and LSTM are capturing linear correlation with log transformation in order to address complex patterns. Ensemble models (XGBoost, Random Forest) are used for feature interactions, Linear Regression for simpler associations, and a deep learning model for non-linear correlations in death rate forecasts. For temperature anomaly, ARIMA shall be used again and stationarity confirmed using an ACF, PACF, or a graph. The system combines traditional statistical methods with sophisticated machine learning for reliable forecasts.

3.7.5 Model Performance Evaluation:

To evaluate the model's performance in an out-of-sample case, a validation set has been used. In order to evaluate the accuracy of the model's prediction, metrics such as the mean squared error, MSE, and R-squared have been used.

4 Design specification:

4.1 Feature selection:

4.1.1 Feature selection for the CO2 prediction:

The name of the country or continent, the year of carbon dioxide emissions, and the corresponding yearly carbon production were three columns in the CO2 record. The first steps were tosolve nulls and deal with outliers so that data could be prepared for modeling. Subsequently, the production numbers of carbon have been modified in two different ways. The first transformation applied an exponential scale so that ARIMA and LSTM models could be exploited. In the meantime,a square function was used for this second transformation in order to fit only an ARIMA model.These carefully selected modifications have been designed to optimize the information for different modeling methods and enhance their accuracy and effectiveness in subsequent studies.

4.1.2 Feature selection for death rate due to CO2:

In combining the data on carbon dioxide with death rates in order to remove superfluous columns related to specific causes, e.g. heavy sodium intake, low or no grains consumption, and air pollution, careful selection has been made of major elements of the Carbon Dioxide Death Rate Forecast Model. The emphasis was on estimating the mortality rates in individual countries, with an "Entity" column that included country and continent names as a main indicator. Three letters from the nation codes were included in "code_x" and there was a temporal element of death rates and CO2 emissions reflected in 'year' columns. Moreover, the number of deaths due to air pollution has been calculated in a column entitled "lives affected". In order to enhance model performance, the data preprocessing consisted of the application of yearly CO2 emissions, and One Hot Encoder was applied to the "Entity" column. A comprehensive strategy was developed at the end.

4.1.3 Feature selection for predicting Temperature anomalies:

To forecast the overall temperature anomalies, the dataset has undergone a series of significant changes. As early as last month, only the global temperature anomaly was aggregated into one category for all months. The year column was converted to the date format. Stationarity, which is a critical factor for the analysis of time series, has been evaluated using different diagnostics methods. Seasonality, trends, and residuals were investigated, and Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) graphs were created to investigate correlations between time series and lags. Following that, the ARIMA Autoregressive Integrated Moving Average methodology was applied to model and predict future average temperature anomalies on a global scale in conjunction with knowledge derived from an exploratory study.

4.2 Data Preprocessing:

4.2.1 Handling Missing Values:

- In the carbon dioxide database, recognize the missing values.
- Imputation methods include elimination, a mean, median, and mode of imputation.

4.2.2 Outlier Handling:

- Identifying the outliers in carbon dioxide data.

4.2.3 Data Transformation:

- Applying transformations such as log and square for ARIMA.
- Applying log transformation for LSTM.

4.2.4 Combining the dataset:

- Joining of carbon dioxide data with death rate dataset.

4.2.5 Renaming of Column:

- Renaming of columns such as Entity, Code_X, Year, and Lives Affected for better understanding.

4.2.6 One-Hot Encoding:

- Apply one-hot encoding to predict the death rate.

4.2.7 Model Application:

- Random forest, XGBoost, linear regression, and a deep learning model have been used to estimate death rates.

4.2.8 Transformation of data for Temperature Anomalies:

- Take a monthly temperature anomaly column and transform it into one column.

5 Implementation:

The steps are taken to predict the global anomalies, death rate prediction due to carbon dioxide, and the emission of carbon dioxide with the help of Machine Learning and deep learning models such as

Random Forest, XGBoost, Linear Regression, ARIMA(Autoregressive Integrated Moving Average), LSTM(Long Short-Term Memory), ANN(Artificial Neural Network).

1) **Importing needed Libraries:**

Importing the Python libraries that are needed for machine learning, data manipulation, and visualization is the initial step in implementing it.

2) **Loading the historical Dataset:**

Loading death rate, temperature anomalies, and CO2 data from Kaggle.com, data.world, and ourworldindata.org.

3) **Data preprocessing:**

In data preprocessing the carbon dioxide(co2) emission data the missing values were checked

Missing Values:	
Entity	0
Code	1360
Year	0
Annual CO2 emissions	0
dtype: int64	

Fig 1: Image for checking missing values

For the death rate prediction, the data was combined with the Carbon dioxide CO2 dataset topredict the death rate due to pollution and applied One Hot Encoder for the Entity column.

	Entity	Country Code	Year of observation	Death count
0	Afghanistan	AFG	1990	37231
1	Afghanistan	AFG	1991	38315
2	Afghanistan	AFG	1992	41172
3	Afghanistan	AFG	1993	44488
4	Afghanistan	AFG	1994	46634
…	…	…	…	…
6835	Zimbabwe	ZWE	2015	13246
6836	Zimbabwe	ZWE	2016	13131
6837	Zimbabwe	ZWE	2017	12926
6838	Zimbabwe	ZWE	2018	12745
6839	Zimbabwe	ZWE	2019	12667
6840 rows	x 4 columns			

Fig 2: Record of Death rate due to pollution

	Entity	Code	Year	Annual CO2 emissions
0	Afghanistan	AFG	1949	14656.0
1	Afghanistan	AFG	1950	84272.0
2	Afghanistan	AFG	1951	91600.0
3	Afghanistan	AFG	1952	91600.0
4	Afghanistan	AFG	1953	106256.0

Fig 3: Record of Carbon dioxide emission

	Entity	Code_x	Year	Lives Affected	Code_y	Annual CO2 emissions
0	Afghanistan	AFG	1990	37231	AFG	2024326.1
1	Afghanistan	AFG	1991	38315	AFG	1914301.0
2	Afghanistan	AFG	1992	41172	AFG	1482054.0
3	Afghanistan	AFG	1993	44488	AFG	1486943.0
4	Afghanistan	AFG	1994	46634	AFG	1453829.0

Fig 4: Record after combining death rate and CO2 emission record

For the temperature anomalies, all the months had different columns. So all the months were set in one order the year column was converted to date and time format and seasonality, trends, and residuals were checked in the dataset.

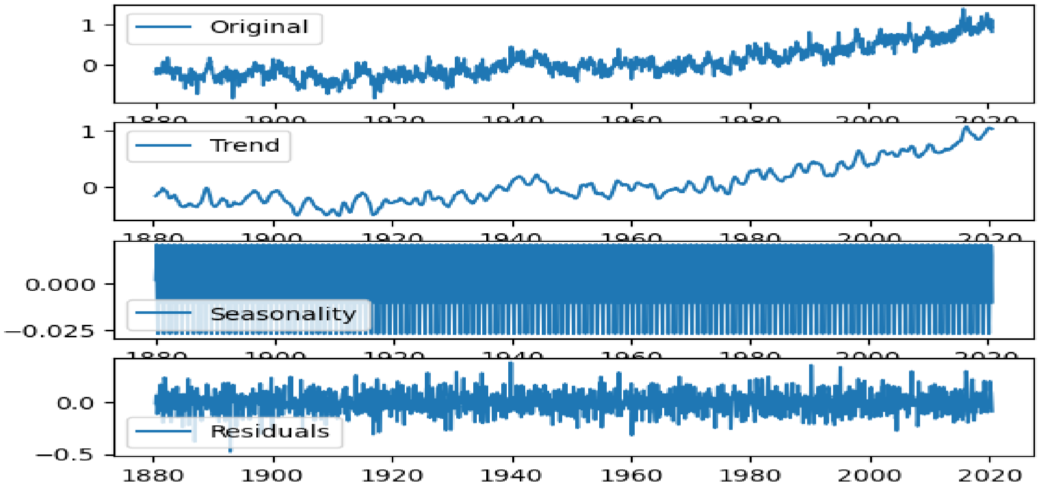

Fig 5: Graph of seasonality, trends, and residuals for temperature anomalies

4) visualization of data:

The below graph shows the visualization of the co2 emissions and temperature anomalies over years.

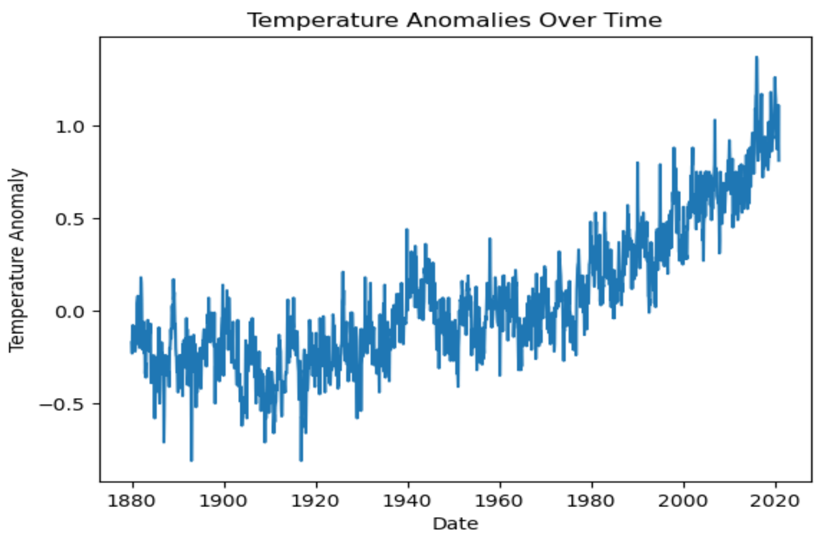

Fig 6: Temperature Anomalies Over Time

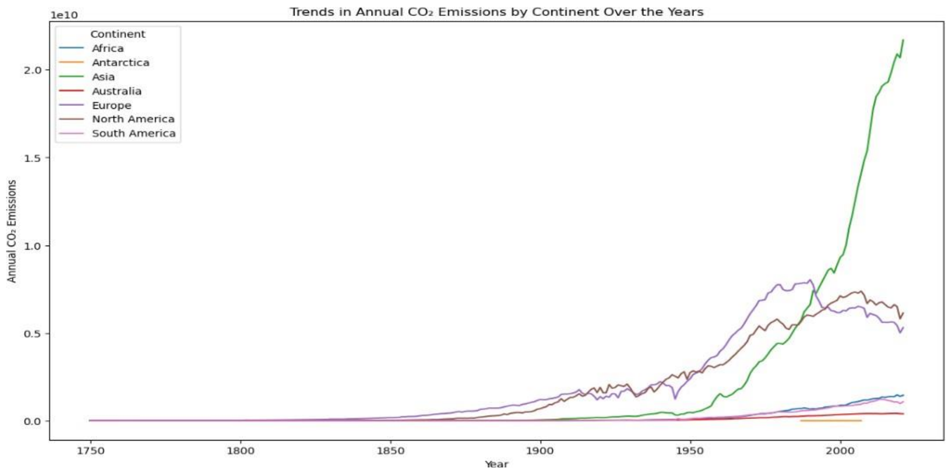

Fig 7: Trends in Annual CO2 Emissions by Continent Over the Years

5) **Feature Selection:**

Feature select for CO2 prediction:

The "Annual CO2 emission" column was used to predict the CO2 level of each continent using the LSTM model and ARIMA model.

Feature select for temperature Anomalies:

After arranging all the months in one column "Temperature Anomaly" the ARIMA model was applied to it.

Feature selection for death rate prediction due to pollution:

Country name in the column "Entity" and the CO2 emission record in the column "Annual CO2 emissions" were taken as the independent variable and the death rate record in "Lives Affected" was taken as the dependent variable.

6) **Data splitting:**

All three datasets were split into test and train to apply models to achieve the maximum score.

```
In [11]: def split_data(data):
             train_data = data.loc[data.index <= '2019-12-31']
             test_data = data.loc[(data.index >= '2020-01-01') & (data.index <= '2020-12-31')]
             return train_data, test_data
```

Fig 8: Splitting of data for temperature anomalies

```
    train_data, test_data = train_test_split(co2_data, test_size=0.2, shuffle=False)
    scaler = MinMaxScaler()
```

Fig 9: Splitting of data for CO2 prediction

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Fig 10: Splitting of data for death rate prediction

7) **Model Training:**

The ARIMA, and auto ARIMA model was then trained for the training set of temperature anomalies. Using auto ARIMA the best fit for (p,d,q) obtained is (2,1,1).

```
Performing stepwise search to minimize aic
 ARIMA(2,1,2)(1,0,1)[12] intercept   : AIC=-110.683, Time=0.64 sec
 ARIMA(0,1,0)(0,0,0)[12] intercept   : AIC=-95.577, Time=0.03 sec
 ARIMA(1,1,0)(1,0,0)[12] intercept   : AIC=-108.426, Time=0.12 sec
 ARIMA(0,1,1)(0,0,1)[12] intercept   : AIC=-111.922, Time=0.15 sec
 ARIMA(0,1,0)(0,0,0)[12]             : AIC=-97.333, Time=0.02 sec
 ARIMA(0,1,1)(0,0,0)[12] intercept   : AIC=-113.409, Time=0.09 sec
 ARIMA(0,1,1)(1,0,0)[12] intercept   : AIC=-111.839, Time=0.12 sec
 ARIMA(0,1,1)(1,0,1)[12] intercept   : AIC=-110.116, Time=0.27 sec
 ARIMA(1,1,1)(0,0,0)[12] intercept   : AIC=-114.915, Time=0.11 sec
 ARIMA(1,1,1)(1,0,0)[12] intercept   : AIC=-113.096, Time=0.21 sec
 ARIMA(1,1,1)(0,0,1)[12] intercept   : AIC=-113.139, Time=0.18 sec
 ARIMA(1,1,1)(1,0,1)[12] intercept   : AIC=-111.335, Time=0.47 sec
 ARIMA(1,1,0)(0,0,0)[12] intercept   : AIC=-109.975, Time=0.06 sec
 ARIMA(2,1,1)(0,0,0)[12] intercept   : AIC=-115.796, Time=0.20 sec
 ARIMA(2,1,1)(1,0,0)[12] intercept   : AIC=-113.995, Time=0.43 sec
 ARIMA(2,1,1)(0,0,1)[12] intercept   : AIC=-114.036, Time=0.34 sec
 ARIMA(2,1,1)(1,0,1)[12] intercept   : AIC=-112.186, Time=0.60 sec
 ARIMA(2,1,0)(0,0,0)[12] intercept   : AIC=-109.037, Time=0.13 sec
 ARIMA(3,1,1)(0,0,0)[12] intercept   : AIC=-114.080, Time=0.23 sec
 ARIMA(2,1,2)(0,0,0)[12] intercept   : AIC=-114.248, Time=0.24 sec
 ARIMA(1,1,2)(0,0,0)[12] intercept   : AIC=-115.414, Time=0.18 sec
 ARIMA(3,1,0)(0,0,0)[12] intercept   : AIC=-112.034, Time=0.06 sec
 ARIMA(3,1,2)(0,0,0)[12] intercept   : AIC=-111.894, Time=0.20 sec
 ARIMA(2,1,1)(0,0,0)[12]             : AIC=-115.182, Time=0.09 sec

Best model:  ARIMA(2,1,1)(0,0,0)[12] intercept
Total fit time: 5.188 seconds
```

Fig 11: Best fit for temperature anomalies data obtained using auto ARIMA

For the death rate prediction due to pollution, the training set data was trained using Random Forest, Linear Regression, XGBoost, and ANN(Artificial neural network)

```
5]: models = {
        'Linear Regression': LinearRegression(),
        'Random Forest': RandomForestRegressor(random_state=42),
        'XGboost': XGBRegressor()
    }
```

Fig 12: The model functions were set in an array

```
model = Pipeline([
    ('preprocessor', preprocessor),
    ('regressor', models[model_name])
])
```

Fig 13: The array was connected to a pipeline

```
model.fit(X_train, y_train)
```

Fig 14: Each model was trained using the training set

```
model = Sequential()
model.add(Dense(units=64, activation='relu', input_dim=X_train.shape[1]))
model.add(Dense(units=32, activation='relu'))
model.add(Dense(units=1))
model.compile(optimizer='adam', loss='mean_squared_error')

model.fit(X_train, y_train, epochs=50, batch_size=32, validation_split=0.2)
```

Fig 15: Training set data was Trained using ANN(Artificial neural network)

For the CO2 emission prediction, the training set data was trained using LSTM and ARIMA for all continents.

```
model = Sequential()
model.add(LSTM(50, input_shape=(seq_length, 1)))
model.add(Dense(1))
model.compile(loss='mean_squared_error', optimizer='adam')
model.fit(X_train, y_train, epochs=100, batch_size=64)
```

Fig 16: All the continent's CO2 emission training set data were trained using the above code of LSTM in a function

```
model = auto_arima(train_data, seasonal=True, m=12, stepwise=True, trace=True)
model.fit(train_data)
```

Fig 17: All the continent's CO2 emission training set data were trained using the above code of autoARIMA in a function

8) Prediction of Models:

All the values were predicted according to the validation set.

Prediction of temperature anomalies:

```
def plot_actual_vs_predicted(actual, predicted, title='Actual vs Predicted'):
    plt.figure(figsize=(10, 6))
    plt.plot(actual.index, actual, label='Actual', marker='o')
    plt.plot(predicted.index, predicted, label='Predicted (Last 12 Months)', marker='o', color='orange')
    plt.title(title)
    plt.xlabel('Date')
    plt.ylabel('Temperature Anomaly')
    plt.legend()
    plt.show()
```

Fig 18: Function that was used to plot the predicted temperature anomalies

```
In [31]: auto_arima_predictions.index = test_data.index
         plot_actual_vs_predicted(test_data, auto_arima_predictions, title='Auto ARIMA Model')
```

Fig 19: Function call for plotting the predicted graph and prediction of temperature anomalies

Prediction of Carbon dioxide emission:

```
forecast, conf_int = model.predict(n_periods=len(test_data), return_conf_int=True)
actual_data = test_data.values
```

Fig 20: The CO2 emissions of all continents were predicted by this line of code for ARIMA

```
y_pred = model.predict(X_test)
y_pred = scaler.inverse_transform(y_pred)
```

Fig 21: The CO2 emissions of all continents were predicted by this line of code for LSTM

Prediction of the death rate due to pollution:

```
y_pred = model.predict(X_test)
```

Fig 22:The Random Forest, XGBoost, and Linear Regression were predicted using the above code in a function for the death rate prediction due to pollution

9) Evaluation Metrics:

The ARIMA model accuracy was evaluated using Mean absolute error(MAE), Mean square error (MSE), and Root mean square error(RMSE)

Mean Squared Error:

MSE is calculated by averaging the squared projected errors in any prediction. Due to squared errors, huge forecasting disparities are squared to produce even larger inaccuracies. As a result, outliers sharply raise the MSE values.Because PROPHET's MSE was the lowest of all models, it demonstrates that the novel Prophet is capable of dealing with outliers successfully.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$$

Mean Absolute Error:

The average of all predicted errors, if all forecasted values are positive, is the mean absolute error. Because it considers the direct absolute forecast difference between the test and forecasts, MAE is not subject to outliers. Because MAE is particularly sensitive to outliers, it has an advantage over mean squared error. Because the data contains a lot of peak values, MAE was chosen as the major evaluation statistic.

$$\text{MAE} = \frac{\sum_{t=1}^{k}|\hat{y}_t - y_t|}{k},$$

Root Mean Squared Error: The Root Mean Squared Error yields the RMSE. The advantage of RMSE over MSE is that it uses the same units as the data, making it easier to read. The root mean square error (RMSE) is a prominent evaluation metric for time series projects.

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

The LSTM(Long Short-Term Memory), Random Forest, XGBoost, Linear Regression accuracy and ANN(Artificial neural network) are evaluated by Mean Squared Error, Mean Absolute Error, Root Mean Squared Error, and R-square.

R-square: In a regression model, the R-squared value, also known as the coefficient of determination, quantifies the proportion of variance in the dependent variable explained by the independent variables. It has a scale of 0 to 1, with 0 indicating no explanatory power and 1 indicating complete explanatory power. A greater R-squared indicates that the model fits the data better.

$$R^2 = \frac{SSR}{SStot} or R^2 = 1 - \frac{SSres}{SStot}$$

6 Evaluation:

6.1 Comparison of ARIMA and LSTM form the prediction of CO2 emissions:

The ARIMA model was applied in 4 ways for all the continents one of them was by applying only ARIMA by taking AutoRegressive, Integration, and moving average as(p,d,q) (1,1,1), the second way was by applying autoregression technique, the autoregression helps in finding the best combination of (p,d,q) and the third way was by applying Square root transformation and auto ARIMA and the fourth way was by using Log transformation and auto ARIMA.

Table 1:The results obtained in Africa using these four methodologies is given below:

Algorithms	Mean Squared error	Mean Absolute error	Root Mean Squared error
ARIMA for Africa	290531760.331297	1.2051250316340288e+17	347149107.9686118

Auto ARIMA for Africa	339259433.9861032	1.628588964059672e+17	403557798.09832346
Auto ARIMA with Log transformation for Africa	5435814095.230929	8.273068665910785e+19	9095641080.160751
Auto ARIMA with square transformation for Africa	211195072.87397638	5.634386201854745e+16	237368620.54312792

Here we can see auto ARIMA with square transformation is seen to be giving better results compared to other ways of implementing ARIMA with mean squared error 211195072.87397638, mean absolute error 5.634386201854745e+16 and root mean squared error 237368620.54312792.

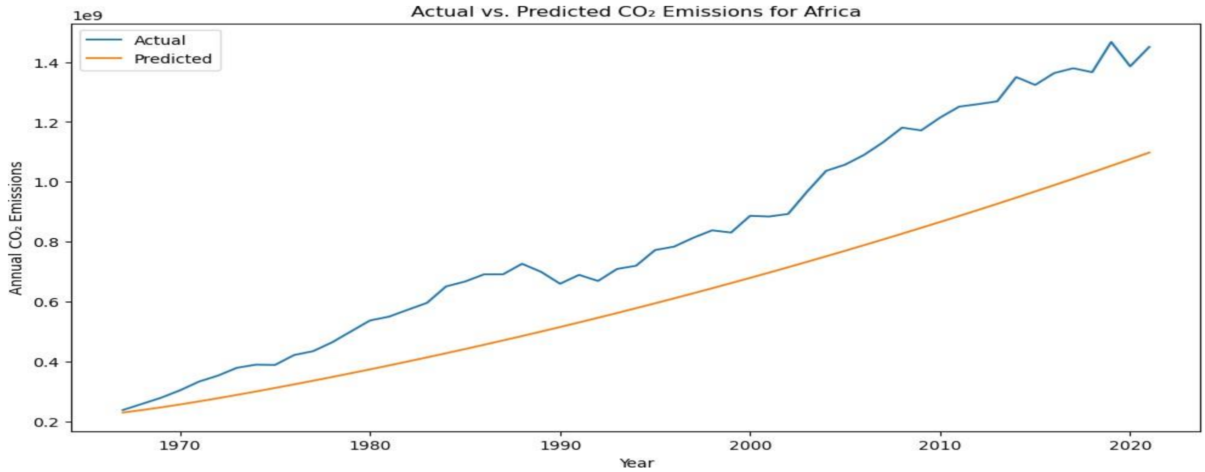

Fig 23: Comparison of predicted CO2 emission data and the actual data using auto ARIMA withsquare transformation for Africa.

When coming to the comparison of results in North America it seems to be the same as in Africa the result was better when ARIMA used square transformation and auto ARIMA.

Table 2: ARIMA and Auto ARIMA using Log and square transformation to predict CO2 emission in North America

Algorithms	Mean Squared error	Mean Absolute error	Root Mean Squared error
ARIMA for North America	2160261185.587062	5.33167749227389e+18	2309042548.8227563
Auto ARIMA for North America	651150994.9901432	5.2449058358417197e+17	724217221.270091
Auto ARIMA with Log transformation for North America	5435814095.230929	8.273068665910785e+19	9095641080.160751
Auto ARIMA with square transformation for North America	662676064.6131425	1.016603949809434e+18	1008267796.6737974

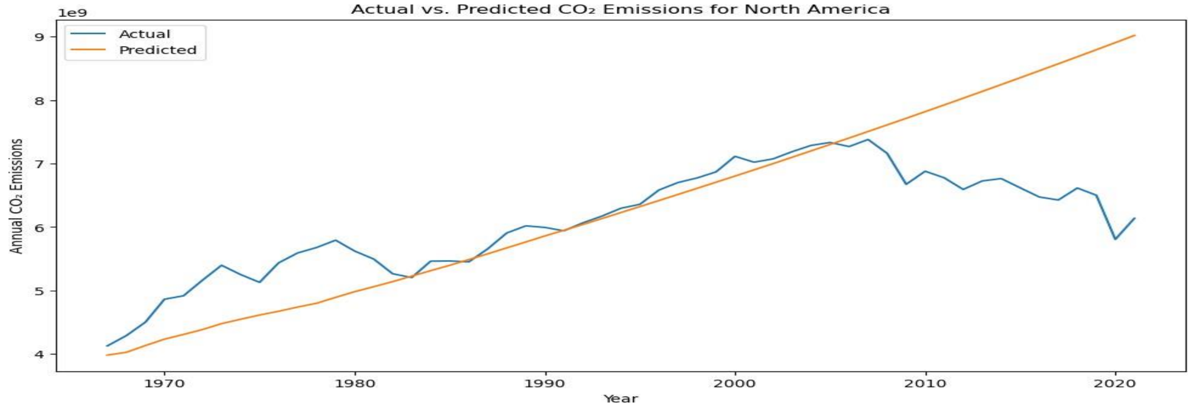

Fig 24: Comparison of predicted CO2 emission data and the actual data using auto ARIMA withsquare transformation for North America.

In the same way, all the continent's predictions of CO2 emission performed better when auto ARIMA and square transformation were applied. The better results obtained in South America, Europe, Australia, and Asia were 84789367.91980399, 5681130173.117291, 38456109.359867394, and 6173500177.719932 Root Mean Squared errors.

Compared to ARIMA, LSTM has given better results. LSTM was applied in two ways, one way was by just applying LSTM without any transformation and putting a single layer with 100 epochs and the second method was by applying log transformation with two layers and 70 epochs, the accuracy was improved when log transformation was apply to it in LSTM. The result obtained using LSTM with log transformation is given below.

Table 3: Prediction of CO2 prediction using LSTM for all the continents

Algorithm	Mean Squared error	Mean Absolute error	R-squared value
LSTM for Asia	2.3564833517148513	0.6000155924906578	0.9746838822849571
LSTM for Europe	0.01053538793285718	0.08264950142925974	0.9978587510632275
LSTM for North America	0.6503821173554817	0.4458126905330167	0.9874727276998669
LSTM for South America	0.6075929142721977	0.2716939328707382	0.9928710565032028
LSTM for Africa	0.5921187680114269	0.2607452187040052	0.9930747222830251
LSTM for Australia	0.954941086802987	0.2783716062941217	0.9870752854723714

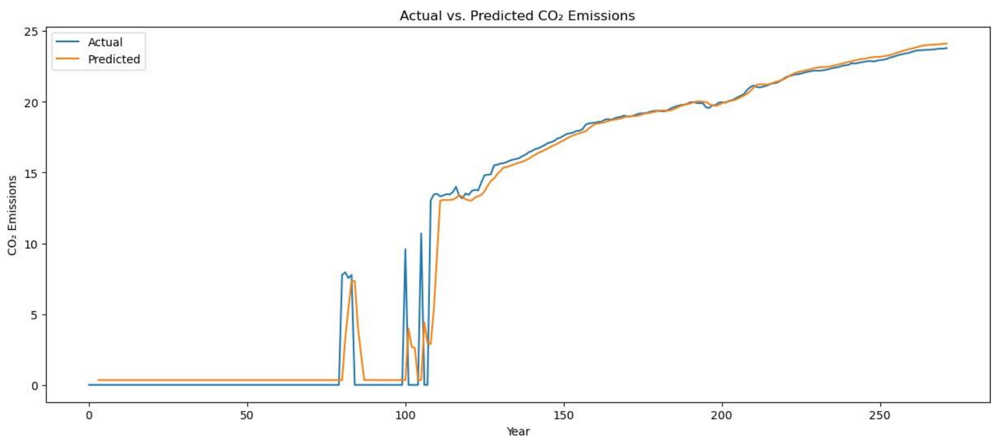

Fig 25: Comparison of predicted CO2 emission data and the actual data using LSTM with Logtransformation for Asia.

Graphs for rest of the continents were also same like Asia close to the actual CO2 emission line.

Comparing ARIMA and LSTM, LSTM is seen to give better accurate results compared to the ARIMA model.

6.2 Comparison of Random Forest, XGBoost, Linear Regression, and ANN for the prediction of the death rate due to carbon dioxide:

For the prediction of death rate Random forest and XGBoost is seen to be giving better results when compared to the other two techniques.

Table 4: Comparison of predicted CO2 emission data and the actual data using LSTM with Logtransformation for Asia.

Algorithms	Mean Squared error	Mean Absolute error	Root Mean Squared error	R-squared
Random forest	21037563.003558323	1279.533198156811	4586.672323543325	0.9998861755993983
Linear Regression	28877181370.73155	45615.532187226316	169932.87313151493	0.8437590960495704
XGBoost	54954612.266509205	3149.0177540620844	7413.137815156899	0.9997026663306735
ANN	3.9506031747132114e+18	287576743.4249185	1987612430.7100747	-0.02135105080151889

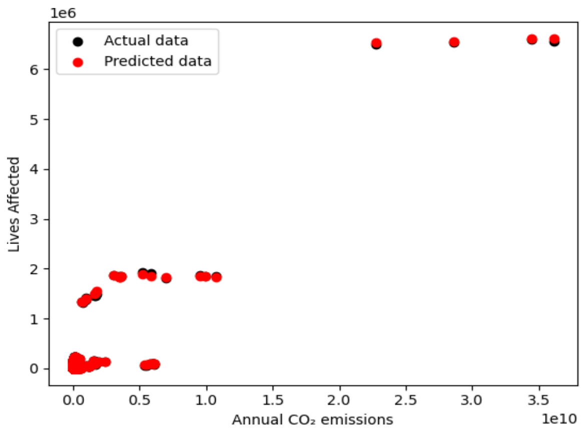

Fig 26: Prediction graph of Random forest

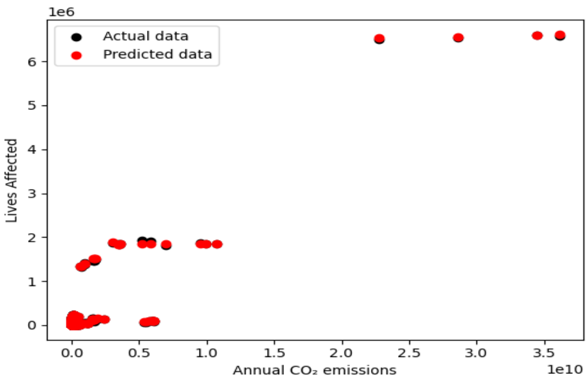

Fig 27:Predicted graph of XGBoost

6.3 Results obtained when ARIMA was applied to predict the temperature anomalies:

For the prediction of temperature anomalies, ARIMA was applied in two ways without auto ARIMA and with auto ARIMA. In ARIMA (1,1,1) was considered for AutoRegressive, for Integration, and moving average. When auto ARIMA was applied the best fit was shown to be (2,1,1) for AutoRegressive, for Integration, and moving average.

Table 5: Results obtained using ARIMA and Auto ARIMA in predicting temperature anomalies

Algorithm	Mean Squared error	Mean Absolute error	Root Mean Squarederror
ARIMA	0.13033416212658233	0.02503972779973708	0.15823946347146492
Auto ARIMA	0.11554501319511012	0.0194171273999903	0.13934535299029638

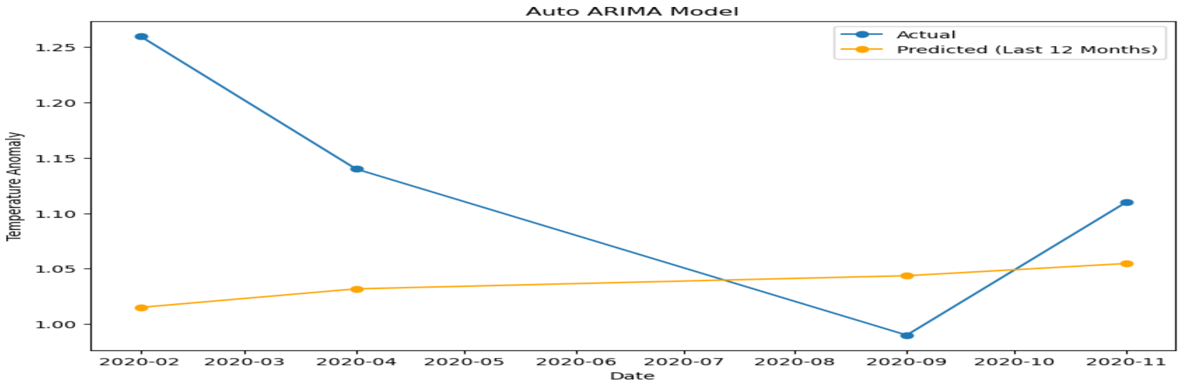

Fig 28: Prediction of temperature anomalies using Auto ARIMA

The result is seen to be better when auto ARIMA was applied by finding the best fit as (2,1,1), and the result obtained is shown in the above table with mean Squared error 0.11554501319511012, mean Absolute error 0.0194171273999903, and root mean Squared error 0.13934535299029638.

7 Conclusion:

The results of this study revealed the effectiveness of machine learning and deep learning methods to predict global anomalies such as CO2 emissions across continents, and pollution-related mortality rates. The complex link between environmental factors and public health impacts has been illustrated in the models developed. The forecast of these important indicators is a major step towards understanding the role played by people's activities in the environment and human well-being. These findings suggest that the challenges of today's climate change and pollution must be addressed by using advanced computational tools.

By integrating more extensive data sets and exploring new model architectures, research efforts could be focused on updating current models as we go forward. Furthermore, efforts should be made to develop real-time forecasting systems that can help policymakers to tackle the negative impacts of pollution and climate change before they happen. Accuracy and timeliness of predictions could be enhanced by the integration of satellite data, Internet of Things devices, or any emerging technologies. In addition, it could provide a comprehensive overview of working dynamics by looking at the socioeconomic factors that influence pollution and its consequences. The creation of more practical models that promote a more sustainable and healthy future for our planet will be facilitated by collaborative research between Interdisciplinary Institutions, as well as the development of a Global Data Sharing Framework.

Reference:

Amarpuri, L., Yadav, N., Kumar, G. and Agrawal, S., 2019, August. Prediction of CO 2 emissions using deep learning hybrid approach: A Case Study in Indian Context. In *2019 twelfth international conference on contemporary computing (IC3)* (pp. 1-6). IEEE.

Dossa, K.F. and Miassi, Y.E., 2023. Exploring the nexus of climate variability, population dynamics, and maize production in Togo: Implications for global warming and food security. *Farming System*, *1*(3), p.100053.

Han, J., Lin, H. and Qin, Z., 2023. Prediction and Comparison of In-Vehicle CO2 Concentration Based on ARIMA and LSTM Models. *Applied Sciences*, *13*(19), p.10858.

Javanmard, M.E., Tang, Y., Wang, Z. and Tontiwachwuthikul, P., 2023. Forecast energy demand, CO2 emissions and energy resource impacts for the transportation sector. *Applied Energy*, *338*, p.120830.

Kamoljitprapa, P. and Sookkhee, S., 2022, September.

Forecasting models for carbon dioxide emissions in major economic sectors of Thailand. In *Journal of Physics: Conference Series* (Vol. 2346, No. 1, p. 012001). IOP Publishing.

Kim, Y. and Kim, Y., 2022. Explainable heat-related mortality with random forest and SHapley Additive exPlanations (SHAP) models. *Sustainable Cities and Society*, *79*, p.103677.

Kumari, S. and Singh, S.K., 2022. Machine learning-based time series models for effective CO2 emission prediction in India. *Environmental Science and Pollution Research*, pp.1-16.

Lee, J., Cai, J., Li, F. and Vesoulis, Z.A., 2021. Predicting mortality risk for preterm infants using random forest. *Scientific reports*, *11*(1), p.7308.

Lin, Z., Kuang, J. and Li, W., 2023. Predictions and Research about Global Warming Based on ARIMA models. *Academic Journal of Environment & Earth Science*, *5*(3), pp.42-48.

Liu, Z., Cui, M., Sun, S., Wang, Q., Yang, W., Li, R. and Yue, X., 2023. Application of Multivariate Modelling to Global Warming Projections. *Academic Journal of Science and Technology*, *7*(2), pp.61-65.

Masum, M.H., Islam, R., Hossen, M.A. and Akhie, A.A., 2022. Time Series Prediction of Rainfall and Temperature Trend using ARIMA Model. *Journal of Scientific Research*, *14*(1), pp.215-227.

Parvez, S.M., Uddin, M.N., Uddin, J. and Mahmud, M.I., 2023. PREDICTION OF CARBON DIOXIDE (CO2) EMISSIONS IN.

Sameh, B. and Elshabrawy, M., Seasonal Autoregressive Integrated Moving Average for Climate Change Time Series Forecasting.

Santos, G.V.F., Cordeiro, L.G., Rojo, C.A. and Leismann, E.L., 2022. A Review of the Anthropogenic Global Warming Consensus: An Econometric Forecast Based on the ARIMA Model of Paleoclimate Series. *Applied Economics and Finance*, *9*(3), pp.102-112.

Schachtschneider, R., Saynisch-Wagner, J. and Thomas, M., 2023. Neural Network Based Estimates of the Climate Impact on Mortality in Germany-Application to Storyline Climate Simulations. *Available at SSRN 4442898*.

Singh, S.K. and Kumari, S., 2022. Machine learning-based time series models for effective CO2 emission prediction in India.

Singh, T., Kaur, A., Katyal, S.K., Walia, S.K., Dhand, G., Sheoran, K., Mohanty, S.N., Khan, M.I., Awwad, F.A. and Ismail, E.A., 2023. Exploring the relationship between air quality index and lung cancer mortality in India: predictive modeling and impact assessment. *Scientific Reports*, *13*(1), p.20256.

Spyrou, E.D., Tsoulos, I. and Stylios, C., 2022. Applying and comparing LSTM and ARIMA to predict CO levels for a time-series measurements in a port area. *Signals*, *3*(2), pp.235-248.

Sridhar, K., Radhakrishnan, P., Swapna, G., Kesavamoorthy, R., Pallavi, L. and Thiagarajan, R., 2023. A modular IOT sensing platform using hybrid learning ability for air quality prediction. *Measurement: Sensors*, *25*, p.100609.

Zhou, F., Huang, Z. and Zhang, C., 2022. Carbon price forecasting based on CEEMDAN and LSTM. *Applied Energy*, *311*, p.118601.

Zhu, Y., Al-Ahmed, S.A., Shakir, M.Z. and Olszewska, J.I., 2022. LSTM-based IoT-enabled CO2 steady-state forecasting for indoor air quality monitoring. *Electronics*, *12*(1), p.107.