Forecasting Carbon Dioxide Emission from Energy Consumption within the Industrial Sector in U.S.

MSc Research Project
M.Sc. Data Analytic, MSCDAD_JAN21A_I

Vineet Sawant
Student ID: x19237758

School of Computing
National College of Ireland

Supervisor: Bharathi Chakravarthi

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Vineet Manoj Sawant ………………………………………………………………………………………………
Student ID:	…x1923775………………………………………………………………………………………………
Programme:	M.Sc. Data Analytics …………………………………………………………… **Year:** 2021-2022 …………………………..
Module:	M.Sc. Research Project ………………………………………………………………………………………….………
Supervisor:	Bharathi Chakravarthi ………………………………………………………………………………………….………
Submission Due Date:	31/1/2022 ………………………………………………………………………………………….………
Project Title:	Forecasting Carbon Dioxide Emission from Energy Consumption within the Industrial Sector in U.S. ………………………………………………………………………………………….………
Word Count:	7337 ……………………………………… **Page Count** 22 …………………………………………….……..

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: Vineet Manoj Sawant ………………………………………………………………………………………………

Date: 31/1/2022 ………………………………………………………………………………………………

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):	

Forecasting Carbon Dioxide Emission from Energy Consumption within the Industrial Sector in U.S.

Vineet Sawant
x19237758

Abstract

With carbon dioxide emission on the ascent as of late it's become more significant than any other time in recent memory for stakeholders to gauge future emission patterns and agree on approaches and methodologies, they carry out to limit the impacts. The U.S. has been one of the top emitters of carbon dioxide and its industrial sector has the highest share in it. Reduced carbon dioxide emissions would aid in the action required to address the climate change crisis. The data for 13 industrial sectors in the US was collected from the U.S. Energy Information Administration (EIA) website. Time series forecasting models like Simple Exponential Smoothing (SES), Holt-Winter Smoothing (HW), ARIMA, Prophet and LSTM were applied on the time series data of 13 industrial sectors. The models were compared using MAE, MAPE and RMSE and the best model was used to perform the short-term (six months) forecast for each sector.

Keywords: **SES, HW, ARIMA, Prophet, LSTM, Carbon dioxide forecasting**

1 Introduction

1.1 Background

Global warming and climate change have become key concerns for all international governments. Droughts, irregular weather patterns, and the melting of polar ice caps, which has resulted in the extinction of polar bears and a rise in sea levels, are all devasting impacts of global warming, which is connected to an increase in the average temperature of the globe. Carbon dioxide, one of the greenhouse gases that causes the greenhouse effect, has grown in recent years as a result of changes and acceleration in human activities such as urbanization, deforestation, and the widespread use of fossil fuels. Carbon dioxide emissions are increasing globally, and governments must prepare and mitigate for this. Recent research has looked at carbon emissions and attempted to anticipate future carbon emission patterns in various sectors, such as Shanghai's aviation industry (Yang and O'Connell 2020) and China's building industry (Zhou et al. 2019), but no studies have looked at the US industrial sector. In terms of carbon emissions, the United States of America (USA) placed second in 2020, while China ranked first. In 2019, the industrial sector in the United States accounted for 23% of total greenhouse gas emissions, which were mostly attributable to energy generation from fossil fuels. It is vital to analyse which industries contribute the most to greenhouse gas emissions because of their huge impact. This research project aims to close the gap by projecting carbon dioxide emissions generated by energy consumption in the US industrial sector. Forecasting future carbon dioxide emission patterns is critical so that policymakers may develop strategies to mitigate these emissions. With rules in place, stakeholders can keep an eye on emissions and work to reduce them. Another important aspect of carbon dioxide forecasting is improving public knowledge of the subject matter at stake.

1.2 Research Question

This research project's research question is:
"Which of the forecasting model performs best in short-term (six months) carbon emission forecasting for each of the United States' industrial sectors?"

1.3 Research Objective

The following are the objectives of this research:
- Identifying the best forecasting model by comparing the performance of all models.
- Using the best model to forecast future emissions.
- Making the data public so that relevant policies may be put in place.

The research project report is structured as follows. Section 2 looks at the related work regrading forecasting carbon emission. Section 3 describes the research methodology. Section 4 discusses design specification. Section 5 discusses the implementation followed by the evaluation in Section 6 and finally Section 7 concludes the research report and discusses the potential of any future work.

2 Related Work

This section of the research project examines current work on carbon dioxide forecasting during the last decade. It delves into the algorithms employed and the forecasting accuracy they attained. This section also looks for any gaps that the study proposal may cover, as well as emphasizes key findings from the studies under review.

Chinese officials set a goal of cutting carbon dioxide emissions by 40-45 percent per unit GDP in 2020 and 60-65 percent per unit GDP in 2030, compared to 2005 levels. Yan et al. (2020) undertook study with the goal of forecasting China's carbon dioxide emission intensity by 2030. The authors investigated whether China would be able to fulfil its targeted carbon dioxide reduction target by 2030 using forecasting approaches such as the Auto Regressive Integrated Moving Average (ARIMA), Traditional grey model, Discrete grey model (DGM), and rolling DGM. The authors utilized MAPE as a metric to compare the performance of the forecasting models. All of the forecasting methods utilized had a MAPE of less than 2%, although ARIMA performed better than the three grey models, which had a MAPE of 0.60 percent. The study indicated that, based on the forecasted results of the best performing model (ARIMA), China will be able to reduce carbon dioxide emissions by 57.65% in 2030, which is lower than the objective of 60-65%. ARIMA (Autoregressive Integrated Moving Average) models are one of the most widely used approaches for studying non-stationary time series, and Rahman and Hasan (2017) utilized them to assess carbon dioxide emissions in Bangladesh. Several ARIMA models were built and tested using a number of selection criteria, with the optimal model having the lowest value for these criteria. The ARIMA (0,2,1) model was shown to have the best match for predicting carbon dioxide emissions. Yang and O'Connell (2020) used an Autoregressive Integrated Moving Average (ARIMA) in research to anticipate carbon dioxide emissions from the Shanghai aviation sector . The authors attempted to assess the impact of increased connection between Shanghai and the rest of the globe on carbon emissions. The authors developed a seasonal ARIMA since the aviation business is impacted by trends and seasonality. The authors employed ME, MPE, MSE, RMSE, MASE, MAE and MAPE to measure and validate the models. The authors also compared the ARIMA model prediction to Holt-Winter and TBATS to assess the ARIMA model's predicting ability. The ARIMA (0,1,1) (0,1,1) [12] model was shown to be the best for predicting in the test. Atique et al. (2020) used a seasonal variant of the ARIMA model on time series data to anticipate daily solar energy output. The ARIMA model was compared to two other machine learning algorithms, SVM and Artificial Neural Network (ANN). Even though the machine learning algorithms outperformed SARIMA, there was still potential for improvement. Chigora, Thabani, and Mutambara (2019) used the ARIMA model to anticipate carbon emissions in Zimbabwe in their study. Several ARIMA models were investigated for predictive purposes, and the authors discovered that ARIMA (10,1,0) was the best one to predict carbon emissions in Zimbabwe for the next ten years. Thabani and Wellington (2019) did a study that used ARIMA models to anticipate carbon dioxide emissions in India. The study revealed that, among the many ARIMA models studied, ARIMA (2,2,0) was better fitted to predict carbon dioxide emissions in India for the next 13 years. Another study, conducted by Lotfalipour, Falahi, and Bastam (2013), aimed to forecast carbon dioxide emissions in Iran using data from the British Petroleum website. To develop the forecast, the authors used two models (Grey Model and ARIMA). The RMSE, MAE, and MAPE measures were used to compare the models. The GM (1,1) model outperformed the ARIMA (1,1,2) model. Tudor (2016) employed automation forecasting techniques to forecast carbon dioxide emissions in Bahrain over a ten-year timeframe. The author employed different forecasting approaches for the study, including Holt-Winters, ARIMA the exponential smoothing state space model (ETS), structural time series (STS), BATS/TBAT and neural network (NNAR). The NNAR model achieved the best forecast result in the research when using RMSE as a criterion. The HS test results further boost trust in the NAAR models' predicting ability.

Amarpuri et al. (2019) utilized a CNN-LSTM hybrid deep learning model to forecast carbon dioxide emissions in India for 2020. The model is built on two deep learning approaches developed from ANN (Artificial Neural Network): Convolution Neural Network and Long Short-Term Memory The authors examined the RMSE and MAPE results of the proposed model to an exponential smoothing model, and concluded that the hybrid model performed better. Ameyaw et al. (2019) employed an LSTM model to predict carbon dioxide emissions from fuel combustion in the United States, Canada, China, and Nigeria. Because of its capacity to store contextual information, the LSTM model was chosen for the study. hao et al. (2018) forecasted carbon dioxide emissions in the United States using a MIDAS-BP model, which is a hybrid model made up of a mixed data sampling regression model (MIDAS) and a back propagation neural network (BP). The MIDAS-BP model was also capable of providing real-time predictions and was ideal for making forecasts for the short, medium, and long range, making it a promising option in the field of carbon dioxide forecasting. Mason et al. employed a neural network (RNN) to forecast Ireland's electricity demand, energy output, and CO2 levels in 2018. The authors concentrated their research on short-term forecasting utilizing Eirgrid time series data for two months, one of which was utilized for training and the other for testing. The study's main goal was to see if a CMA-ES-trained neural network could make accurate predictions. Four common measures were used to assess the forecast accuracy (MAE, MSE, RMSE and MAPE). The study's findings show that CMA-ES provides the most accurate forecasting for the training data and performs better on two of the three forecasting difficulties when tested with previously unknown test data in the forecasting conditions. Khan & Khan (2019) recommended the adoption of fuzzy-based modelling tools over traditional methodologies due to the intricate link between carbon dioxide emissions and temperature rise. Techniques such as ANFIS, ANN, and fuzzy time series modelling were applied. The ANFIS and ANN models were evaluated using the RMSE and correlation coefficient (R^2)/MSE, respectively. The fuzzy time series analysis' performance was evaluated using MAE, MSE, MAPE and RMSE. Jin (2021) employed artificial neural networks such as the BP-neural network, RBF, and Elman neural network to forecast carbon dioxide emission in various Chinese provinces, accounting for the non-linear connection between the characteristics of the carbon dioxide emission data.

Ming et al. (2014) utilized a hybrid model with a short sample size to predict carbon dioxide emissions associated with energy consumption in China throughout the last 20 years (1992-2011). The Grey Model GM is a version of the hybrid model (1,1). To assess the hybrid model's efficiency, the authors used a linear model and a GM (1,1) to compare predicting accuracy using MdAPE (median absolute percentage error), MAPE (mean absolute percentage error), and MaxAPE (maximum absolute percentage error). Zhou et al. (2019) studied China's construction industry's historical and future developments. This study developed a better prediction model that incorporated a weighted algorithm with an Elman neural network (ENN) modified by the Adaptive Boosting algorithm for evaluating future carbon emissions in China's construction sector (Adaboost).

In research by Kallio et al. (2021), the authors used four different machine learning approaches to anticipate carbon dioxide content in an indoor office. Another research, by Leerbeck et al. (2020), looked at short-term carbon dioxide emission predictions in relation to the power system. Sánchez Lasheras et al. (2020) attempted several strategies to anticipate PM10 concentration based on past historical data for the Port of Gijon in Spain in forecast research. The study used the ARIMA, Vector Autoregressive Moving Average (VARMA), MLP (Multi-layer Perceptron), Support Machine Vector as Regressor (SVMR), and multivariate adaptive regression splines. The models were assessed using the RMSE value as a criterion.

3 Research Methodology

3.1 Introduction

Data mining is the process of collecting valuable information from a large volume of data in order to acquire new insights. The three most prominent data mining approaches are KDD (Knowledge Discovery in Databases), CRISP-DM (Cross-industry standard process for data mining), and SEMMA (Sample, Explore, Modify, Model, and Assess). In terms of the number of stages, each approach is distinct. The CRISP-DM methodology was chosen for this study because it offers flexibility and structure to the project, is simple to understand, and has well-designed steps that cover all aspects of the project.

3.2 Overview of CRISP-DM

CRISP-DM contains six phases or processes that span the whole life-cycle of a data mining project (as depicted in Figure 1). Business understanding, data understanding, data preparation, modelling, evaluation, and deployment are the six steps of the CRISP-DM methodology.

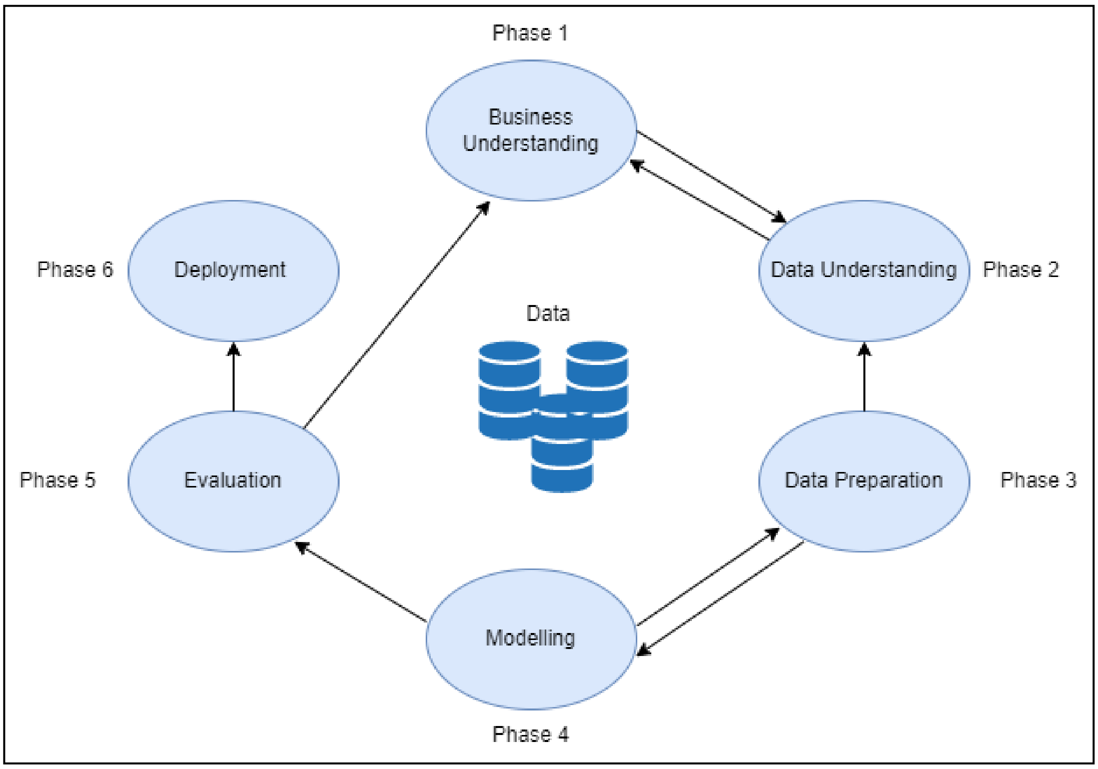

Figure 1: CRISP-DM Phases

3.2.1 Business Understanding

It is the first and most important phase of the project, during which the research objectives are specified and highlighted, followed by the development of a business strategy to attain those goals. The fight against climate change has already begun, with individuals banding together to develop various solutions that would aid in improving the environment. There is an immediate need to mitigate global warming by implementing sustainable living practices. The goal of the research project is to predict carbon dioxide emissions from various industrial sectors in the United States so that sectors with rising emissions may be determined, and this information, can aid in the implementation of policies to reduce emissions. The business plan is to forecast carbon dioxide emissions by employing five time series models and then selecting the right model with the highest forecasting accuracy for future projections.

3.2.2 Data Understanding

This is the second phase of the CRISP-DM, where the acquired data is extensively evaluated in order to gain a better knowledge and insight into the data. The dataset for the research project was obtained from the website of the United States Energy Information Administration (EIA)[1], which has made the dataset accessible for usage. The dataset contains monthly data of carbon dioxide emission from thirteen industrial sectors in the U.S. measured in million metric tons from January 1973 to April 2021. The dataset has 15 columns and 580 rows and no NA values. The column ‘Total Industrial Sector CO2 Emissions’ is the sum of all the emissions of all the sectors for that month. This column is not used in the time series analysis.

[1] https://www.eia.gov/totalenergy/data/browser/?tbl=T11.04#/?f=M

3.2.3 Data Preparation

There are no irregularities in the dataset, such as wild characters, missing values, or NA values. As a result, preparing the data for modelling does not require much work. Data must be stationary as a precondition for time series models to work. As a result, the time series data is examined for stationarity, and differencing is employed to make non-stationary data stationary.

3.2.4 Modelling

Modelling is the fourth phase of CRISP-DM, in which the completely cleaned data is used by the selected models to generate the necessary output. The five models selected for the purpose of the research project are Simple Exponential Smoothing (SES), Holt-Winter Exponential Smoothing (HW), Autoregressive Integrated Moving Average (ARIMA), Prophet and Long Short-Term Memory (LSTM). These time series models are trained on 80% of the data and tested on the remaining 20% data.

i. SES - Simple Exponential Smoothing is a time series forecasting approach for univariate data without a trend or seasonality. It is also known as Single Exponential Smoothing.
ii. HW - This approach is also known as Triple Exponential Smoothing, is named after two of the system's creators, Charles Holt and Peter Winters. Triple Exponential Smoothing is a variant of Exponential Smoothing that explicitly accounts for seasonality in univariate time series.
iii. ARIMA - One of the most often used time series forecasting methods is ARIMA. It provides a forecast for the future based on the time series' previous values. The order of the AR term (p), the order of the MA term (q), and the number of differencing (d) are the three input parameters for an ARIMA model.
iv. Prophet - Prophet is a novel forecasting model built by Facebook's data science team, and it's open-source software in R and Python. Prophet, in contrast to ARIMA, doesn't need missing values to be interpolated since it can manage to provide high accuracy with inadequate information while giving predominant expectation exactness. It is intended for business anticipating circumstances that Facebook has encountered on an hourly, every day, week after week, and month to month premise. One of the advantages of Prophet is that it doesn't need earlier information, making it simple to utilize and able to naturally identify occasional patterns in information while giving effective results.
v. LSTM - While the LSTM approach was first established in the late 1990s, it is just now becoming a realistic and strong forecasting technology. Traditional forecasting approaches such as ARIMA and HWES remain popular and effective, but they lack the broad generalizability that memory-based models such as LSTM provide. The LSTM addresses a major flaw in recurrent neural networks: short memory. The LSTM manages to keep, forget, or ignore data points depending on a probabilistic model by using a sequence of 'gates,' each with its own RNN (Ameyaw et al. 2019)

3.2.5 Evaluation

The performance of all the models employed in the research is evaluated in the fifth step of CRISP-DM. The performance measures applied in this research are RMSE (Root Mean Square Error), MAE (Mean Absolute Error), and MAPE (Mean Absolute Percentage Error), which are used to evaluate a model's performance to that of other models in order to determine which is the best. Lower RMSE, MAPE, and MAE values are preferable since they indicate that the model is more accurate in forecasting.

3.2.6 Deployment

This phase for the research project entails introducing and presenting a notable report , a designs manual, and a manual for ICT solutions.

3.3 Conclusion

The reasoning for picking CRISP-DM as the research methodology in this project is legitimized and talked about momentarily alongside the six phases present in the CRISP-DM in correspondence to the project.

4 Design Specification

The implementation of the research project is done using Jupyter notebooks in Anaconda Environment and Google Colab. The dataset contains thirteen time series data of various industrial sectors in U.S . Each time series data has a Jupyter notebook associated with it. In the Jupyter notebook the data is read into dataframe , visualized , checked for stationarity using dickey fuller test and then given to models to be trained and tested. The time series models like SES , HW , ARIMA and Prophet are trained and tested in the Jupyter notebook provided by the anaconda environment. For the time series model LSTM, the Jupyter notebook provided by Google Colab is utilised so as to harness the support and libraries available for deep learning.

5 Implementation

5.1 Introduction

This section briefly discusses the implementation flow and the model implementation done to achieve the objectives of the research project. Figure 2 below shows the implementation flow followed in the project to achieve the future forecasting results.

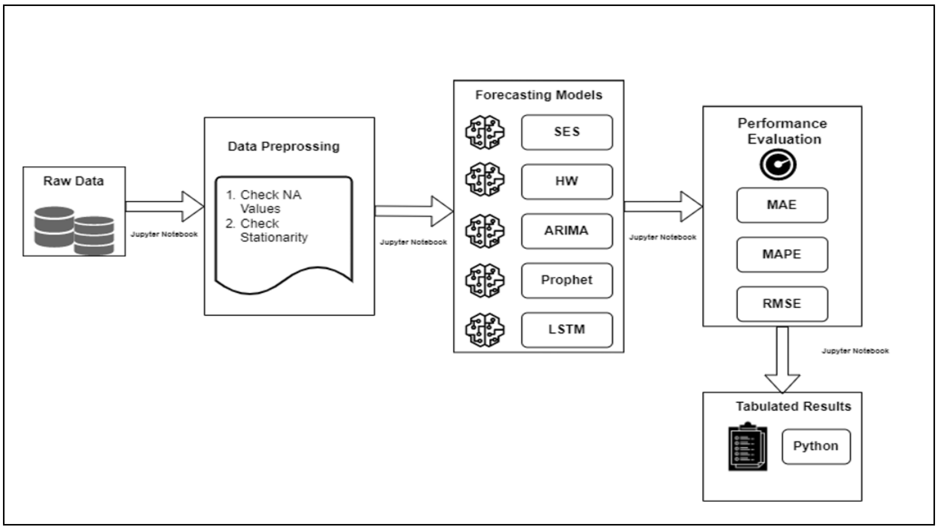

Figure 2: Implementation Flow

5.2 Acquiring Data

The dataset is downloaded from the U.S. Energy Information Administration (EIA) that contains the monthly carbon dioxide emissions measured in metric tonnes from January 1973 to April 2021. This raw dataset contains 580 rows and 15 columns which contain the date and the carbon dioxide emission from 13 sectors along with a column for the total carbon dioxide emission. The figure 3 shows the raw dataset that is used for the purpose of the research project. The table in figure 4 lists all the carbon dioxide emitting sectors in the dataset.

Month	Coal Industrial Sector CO2 Emissions	Coal Coke Net Imports CO2 Emissions	Nat
1973 January	33.236	-0.127	
1973 February	30.609	-0.014	
1973 March	31.408	-0.229	
1973 April	30.904	-0.074	
1973 May	31.429	-0.323	
1973 June	29.806	-0.028	
1973 July	29.249	-0.13	
1973 August	28.846	-0.167	
1973 September	27.997	-0.136	
1973 October	30.832	0.198	
1973 November	31.702	0.139	
1973 December	34.828	0.04	
1974 January	32.855	0.43	
1974 February	30.818	0.337	
1974 March	31.168	0.419	
1974 April	30.787	0.504	
1974 May	29.955	0.589	
1974 June	28.533	0.385	
1974 July	28.323	0.408	
1974 August	29.187	0.439	
1974 September	28.45	0.744	
1974 October	30.327	1.036	
1974 November	27.05	0.555	
1974 December	26.567	0.558	
1975 January	29.883	0.971	

Figure 3: Raw Data

No	Carbon dioxide emitting sectors
1	Coal Industrial Sector CO2 Emissions
2	Coal Coke Net Imports CO2 Emissions
3	Natural Gas Industrial Sector CO2 Emissions
4	Distillate Fuel Oil Industrial Sector CO2 Emissions
5	HGL Industrial Sector CO2 Emissions
6	Kerosene Industrial Sector CO2 Emissions
7	Lubricants Industrial Sector CO2 Emissions
8	Motor Gasoline, Excluding Ethanol, Industrial Sector CO2 Emissions
9	Petroleum Coke Industrial Sector CO2 Emissions
10	Residual Fuel Oil Industrial Sector CO2 Emissions
11	Other Petroleum Products Industrial Sector CO2 Emissions
12	Petroleum, Excluding Biofuels, Industrial Sector CO2 Emissions
13	Industrial Share of Electric Power Sector CO2 Emissions

Figure 4: Emitting Sectors

5.3 Data Pre-processing

Raw data can contain impurities such as missing values (NA), special characters, or incorrect values, it cannot be used directly for training models; therefore, acquired data is always pre-processed to remove data impurities. Jupyter Notebook which is an open-source web application is used for cleaning the dataset used in our research project.

The dataset contains time series data of carbon dioxide emission for 13 industrial sectors in the US. The dataset is available as an excel sheet , the dataset is read into the Jupyter notebook and stored in a dataframe. The columns of the dataframe represent the 13 industrial sector time series data. The columns of the dataframe were checked for NA values. None of the time series data used for the analysis contained NA values. For time series models it is a perquisite that the time series be stationary which means the mean and variance is constant over time. If the time series data is not stationary, we need to convert it into a stationary one. In python it is done using the function diff().

5.4 Model

The time series analysis models are developed using python programming language by using Jupyter notebook. Some python libraries like numpy , pandas , matplotlib , sklearn are used in the development of the time series analysis models. The data used by the models is split into 80/20 where 80% of the data is used to train the model and the remaining 20% is used for testing.

5.4.1 Simple Exponential Smoothing

Simple exponential smoothing (ses) model is used to forecast time series data were there is no clear trend or seasonality. The model is constructed by importing the SimpleExpSmoothing model from the statsmodel library. The model is trained and tested using the fit() function and using the predict() function the future values are forecasted. The test and predicted values are plotted and the values of MAE, MAPE and RMSE are calculated using the sklearn and numpy libraries which can be used to compare the various models.

5.4.2 Holt-Winter Exponential Smoothing

The Holt-Winter exponential smoothing model (hw) is used to forecast time series data were there is trend and seasonality. The model is constructed by importing the ExpotentialSmoothing model from the statsmodel library. The model is trained and tested using the fit() function and using the predict() function the future values are forecasted. The test and predicted values are plotted and the values of MAE, MAPE and RMSE are calculated.

5.4.3 Autoregressive Integrated Moving Average

The Autoregressive Integrated Moving Average (ARIMA) model is constructed by importing the ARIMA model from the statsmodel library. To fit the ARIMA model we need to pass the training data and the order of (p,d,q). The order of 'd' is determined by the number of times we difference the time series to make it stationary. To check the stationarity of a time series we use the dickey fuller test where the p-value is checked , if the p- value is less than 0.05 the null hypothesis is rejected and it is inferred that the time series is stationary. To determine the value of 'p' and 'q' the PACF (Partial autocorrelation function) and ACF (Autocorrelation function) plots are plotted respectively using the plot_pacf() and plot_acf() functions from the statsmodel library. Once the order of (p,d,q) is determined the ARIMA model is trained and fitted using the fit() function. Using the predict() function the future values are forecasted. The test and predicted values are plotted to get a visual representation of how the model fit the data. The values for MAE , MAPE and RMSE are calculated.

5.4.4 Prophet

The Prophet model is constructed by importing the Prophet model from the fbprophet library. The month and the timeseries data are stored in a dataframe and the column names are renamed to 'ds' and 'y'. The data is divided into 80% for training and 20% for training. The model is instantiated using the Prophet() function. The model is trained and fitted on the training dataset using the fit() function. To make future predictions a dataframe containing the future dates is created using the make_future_dataframe() function in which the periods and the frequency are passed. The frequency for the predictions is set to 'MS' (Month Start). The future dataframe is then passed to the predict() function to get the forecast. The test and predicted values are plotted and the values of MAE, MAPE and RMSE are calculated.

5.4.5 Long Short-Term Memory

The Long short-term memory (LSTM) model is constructed in Jupyter notebook using Google Colab. The time series data is loaded from the excel into a dataframe in the Jupyter notebook. The seed() is set to make the model reproducibility The ‘Month’ column is converted to the datetime and the dataframe index is set to the ‘Month’ column and index frequency is set to ‘MS’. The time series data is split into training and testing data. The training data consists of 574 values and the test data contains the remaining values. The values in the train and test set are transformed using the MinMaxScaler which transforms the values in the range of 0 and 1. The TimeseriesGenerator is used to generate the sample input and output component used by the model. The LSTM model is trained and fitted on the train dataset values using 50 epochs and the loss function is plotted. The predict() function is used to forecast the values and the test and predicted values are plotted and the values of MAE, MAPE and RMSE are calculated.

5.5 Conclusion

The Implementation section provides a high-level overview of the functional flow, data cleaning, model construction method, libraries utilized, and functions for obtaining forecasts from the models.

6 Evaluation

6.1 Introduction

In this section the forecasting accuracy of the timeseries models is compared for predicting the carbon dioxide emission for the next six months. The metrics used to compare the various models are MAE, MAPE and RMSE. (Yang and O’Connell 2020).

MAE : The MAE is the average of the absolute difference between predicted and actual values.

MAPE : The ratio of the average absolute difference between predicted and actual values divided by the actual value is known as MAPE.

RMSE : It is the square root of the mean square error. It is also known as the standard deviation of errors that occurs while predicting future values . This statistic is more sensitive to the presence of outliers in the data.

The equations of the evaluation metrics are shown in the Figure 5 below (Yang and O’Connell 2020) , in the equation x_i represents the actual value and y_i represents the predicted value and n represents the total number of observations.

$$RMSE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}\left(y_i - x_i\right)^2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i|$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{x_i - y_i}{x_i}\right|\%$$

Figure 5: Metric Equations

6.2 Experiment 1

In this experiment the carbon dioxide emission data for Coal Industrial Sector is used. The data is spilt into 80% training and 20% testing. The table below shows the evaluation metrics used to determine the performance accuracy between all the models used.

Metrics	SES	HW	ARIMA(2,0,2)	Prophet	LSTM
MAE	0.30821127965667716	0.30821127965667716	0.30821127965667716	0.6775099620049856	0.6178692976012846
MAPE	1.2816207243382343	1.2816207243382343	1.2816207243382343	0.06752018700139785	0.07705232961984561
RMSE	0.3923507941886213	0.3923507941886213	0.39327562977237335	0.8357309363297689	0.7228725589501662

Table 1: Coal Sector Evaluation Metrics Table

In order for a time series model to have a higher accuracy in forecasting the values of the evaluation metrics should be low. We will use the lowest RMSE value for choosing the model. From Table 1 it can be seen that SES and HW models have lower values of RMSE so it has a higher accuracy in forecasting the future values and is used to forecast the next six months of carbon dioxide emission.

6.3 Experiment 2

In this experiment the carbon dioxide emission data for Coal Coke Net Imports Industrial Sector is used. The data is spilt into 80% training and 20% testing. The table below shows the evaluation metrics used to determine the performance accuracy between all the models used.

Metrics	SES	HW	ARIMA(1,0,1)	Prophet	LSTM

MAE	0.4689302102334523	0.4689302102334523	0.4689302102334523	0.15382046501513	0.13422283903757723
MAPE	4.843316789186837	4.843316789186837	4.843316789186837	1.5153360553687192	2.2492043801038757
RMSE	0.49045784699737405	0.49045784699737405	0.4648958925504697	0.19003493855901668	0.15193803322507823

Table 2: Coal Coke Net Imports Sector Evaluation Metrics Table

From Table 2 it can be seen that LSTM has the lowest RMSE value and hence it will be used to forecast the carbon dioxide emission for the next six months.

6.4 Experiment 3

In this experiment the carbon dioxide emission data for Natural Gas Industrial Sector is used. The data is spilt into 80% training and 20% testing. The table below shows the evaluation metrics used to determine the performance accuracy between all the models used.

Metrics	SES	HW	ARIMA(1,0,1)	Prophet	LSTM
MAE	1.8201934275041027	1.8201934275041027	1.8201934275041027	11.379330427969437	4.078506159782413
MAPE	0.9958084441226088	0.9958084441226088	0.9958084441226088	0.26777394435153073	0.08551805161975835
RMSE	2.2544055081595595	2.2544055081595595	2.2542140068585077	12.149796014157605	4.7671682101946455

Table 3: Natural Gas Sector Evaluation Metrics Table

From Table 3 it can be seen that ARIMA(1,0,1) has the lowest RMSE value and hence will be used to forecast the carbon dioxide emission for the next six months.

6.5 Experiment 4

In this experiment the carbon dioxide emission data for Distillate Fuel Oil Industrial Sector is used. The data is spilt into 80% training and 20% testing. The table below shows the evaluation metrics used to determine the performance accuracy between all the models used.

Metrics	SES	HW	ARIMA(1,0,1)	Prophet	LSTM
MAE	1.4164825754821546	1.4164825754821546	1.4164825754821546	0.9510217266619141	1.607771641343832

MAPE	0.187489855 76877317	0.187489855 76877317	0.187489855 76877317	0.146209213 30582873	0.182332 19222064 076
RMSE	1.772098904 4233193	1.772098904 4233193	1.586858743 1958687	1.305724969 8082918	1.780853 87081783 34

Table 4: Distillate Fuel Oil Sector Evaluation Metrics Table

From Table 4 it can be seen that Prophet has the lowest RMSE value and hence will be used to forecast the carbon dioxide emission for the next six months.

6.6 Experiment 5

In this experiment the carbon dioxide emission data for HGL Industrial Sector is used. The data is spilt into 80% training and 20% testing. The table below shows the evaluation metrics used to determine the performance accuracy between all the models used.

Metrics	SES	HW	ARIMA(1,0,1)	Prophet	LSTM
MAE	0.448520769 9239518	0.448520769 9239518	0.448520769 9239518	1.324092246 857593	1.072406 52855237 34
MAPE	1.000786424 202593	1.000786424 202593	1.000786424 202593	0.300800197 37944946	0.211948 05633770 178
RMSE	0.613058802 6859037	0.613058802 6859037	0.612800716 5061814	1.597276694 8486867	1.149753 57330059 6

Table 5: HGL Sector Evaluation Metrics Table

From Table 5 it can be seen that ARIMA(1,0,1) has the lowest RMSE value and hence will be used to forecast the carbon dioxide emission for the next six months.

6.7 Experiment 6

In this experiment the carbon dioxide emission data for Kerosene Industrial Sector is used. The data is spilt into 80% training and 20% testing. The table below shows the evaluation metrics used to determine the performance accuracy between all the models used.

Metrics	SES	HW	ARIMA(3,0,3)	Prophet	LSTM
MAE	0.012252963 492402876	0.012252963 492402876	0.012252963 492402876	0.098791150 66610612	0.012510 65687586 8642
MAPE	17961438856 95.2222	17961438856 95.2222	17961438856 95.2222	21155922335 42.973	1.181566 50769797 56
RMSE	0.017769633 760077327	0.017769633 760077327	0.017784828 719972716	0.140644668 1593087	0.017209 73859985

					7163

Table 6: Kerosene Sector Evaluation Metrics Table

From Table 6 it can be seen that LSTM has the lowest RMSE value and hence will be used to forecast the carbon dioxide emission for the next six months.

6.8 Experiment 7

In this experiment the carbon dioxide emission data for Lubricants Industrial Sector is used. The data is spilt into 80% training and 20% testing. The table below shows the evaluation metrics used to determine the performance accuracy between all the models used.

Metrics	SES	HW	ARIMA(1,0,1)	Prophet	LSTM
MAE	0.052549136712864514	0.052549136712864514	0.052549136712864514	0.6775099620049856	0.0133861728509267
MAPE	10481967760.064526	10481967760.064526	10481967760.064526	0.06752018700139785	0.040088130982134956
RMSE	0.06529335115620415	0.06529335115620415	0.06514017202628151	0.8357309363297689	0.01861573466721162

Table 7: Lubricants Sector Evaluation Metrics Table

From Table 7 it can be seen that LSTM has the lowest RMSE value and hence will be used to forecast the carbon dioxide emission for the next six months.

6.9 Experiment 8

In this experiment the carbon dioxide emission data for Motor Gasoline, Excluding Ethanol, Industrial Sector is used. The data is spilt into 80% training and 20% testing. The table below shows the evaluation metrics used to determine the performance accuracy between all the models used.

Metrics	SES	HW	ARIMA(1,0,1)	Prophet	LSTM
MAE	0.07144140524774376	0.07144140524774376	0.07144140524774376	0.2469497293286886	0.09765758744875593
MAPE	0.9992875588215623	0.9992875588215623	0.9992875588215623	0.17507048580946455	0.0748024257262508
RMSE	0.09531905845379637	0.09531905845379637	0.09538154738650094	0.2894023477387521	0.12192825334671148

Table 8 : Motor Gasoline, Excluding Ethanol Sector Metrics Table

From Table 8 it can be seen that SES and HW has the lowest RMSE value and hence will be used to forecast the carbon dioxide emission for the next six months.

6.10 Experiment 9

In this experiment the carbon dioxide emission data for Petroleum Coke Industrial Sector is used. The data is spilt into 80% training and 20% testing. The table below shows the evaluation metrics used to determine the performance accuracy between all the models used.

Metrics	SES	HW	ARIMA(1,0,1)	Prophet	LSTM
MAE	1.3016697512645448	1.3016697512645448	1.3016697512645448	0.6775099620049856	1.2204178130179637
MAPE	0.9753937155648135	0.9753937155648135	0.9753937155648135	0.06752018700139785	0.4327263818135479
RMSE	1.6794783854502344	1.6794783854502344	1.6690994942869095	0.8357309363297689	1.4822662641209419

Table 9: Petroleum Coke Sector Evaluation Metrics Table

From Table 9 it can be seen that Prophet has the lowest RMSE value and hence will be used to forecast the carbon dioxide emission for the next six months.

6.11 Experiment 10

In this experiment the carbon dioxide emission data for Residual Fuel Oil Industrial Sector is used. The data is spilt into 80% training and 20% testing. The table below shows the evaluation metrics used to determine the performance accuracy between all the models used.

Metrics	SES	HW	ARIMA(1,0,1)	Prophet	LSTM
MAE	0.07083438343573034	0.07083438343573034	0.07083438343573034	0.41449589055752073	0.44245361525689564
MAPE	2041328272458.0078	2041328272458.0078	2041328272458.0078	1.600743302427558	2.348580411925893
RMSE	0.09781940066238648	0.09781940066238648	0.09845948231097001	0.5041562083168408	0.4716541345973428

Table 10: Residual Fuel Oil Sector Evaluation Metrics Table

From Table 10 it can be seen that SES and HW has the lowest RMSE value and hence will be used to forecast the carbon dioxide emission for the next six months.

6.12 Experiment 11

In this experiment the carbon dioxide emission data for Other Petroleum Products Industrial Sector is used. The data is spilt into 80% training and 20% testing. The table below shows the evaluation metrics used to determine the performance accuracy between all the models used.

Metrics	SES	HW	ARIMA(1,0,1)	Prophet	LSTM
MAE	1.190362068965517	1.190362068965517	1.190362068965517	1.1797245293402367	1.520254206717014
MAPE	1.0195258408518184	1.0195258408518184	1.0195258408518184	0.13087129687453608	0.161081033134941 02
RMSE	1.493514698033019	1.493514698033019	1.4930508386370942	1.423326767440241	1.66064249269251 52

Table 11: Other Petroleum Products Sector Evaluation Metrics Table

From Table 11 it can be seen that Prophet has the lowest RMSE value and hence will be used to forecast the carbon dioxide emission for the next six months.

6.13 Experiment 12

In this experiment the carbon dioxide emission data for Petroleum, Excluding Biofuels, Industrial Sector is used. The data is spilt into 80% training and 20% testing. The table below shows the evaluation metrics used to determine the performance accuracy between all the models used.

Metrics	SES	HW	ARIMA(1,0,1)	Prophet	LSTM
MAE	2.509142120326585	2.509142120326585	2.509142120326585	2.9994301816254794	2.0353738808830584
MAPE	0.9998738677590953	0.9998738677590953	0.9998738677590953	0.10241145742306423	0.0834235119827204
RMSE	3.2996552197767217	3.2996552197767217	3.292529188623172	3.675886110307766	3.334315971890593

Table 12: Petroleum, Excluding Biofuels Sector Evaluation Metrics Table

From Table 12 it can be seen that ARIMA(1,0,1) has the lowest RMSE value and hence will be used to forecast the carbon dioxide emission for the next six months.

6.14 Experiment 13

In this experiment the carbon dioxide emission data for Industrial Share of Electric Power Sector is used. The data is spilt into 80% training and 20% testing. The table below shows the evaluation metrics used to determine the performance accuracy between all the models used.

Metrics	SES	HW	ARIMA(1,0,1)	Prophet	LSTM
MAE	2.547927118097881	2.547927118097881	2.547927118097881	3.648941338221681	5.069255708058674
MAPE	1.014829638149213	1.014829638149213	1.014829638149213	0.10614445238145588	0.17385970023907213
RMSE	3.155311518909509	3.155311518909509	3.160930227438713	4.7209075750559055	6.177956841109287

Table 13: Electric Power Sector Evaluation Metrics Table

From Table 13 it can be seen that SES and HW has the lowest RMSE value and hence will be used to forecast the carbon dioxide emission for the next six months.

6.15 Discussion

For the all the experiments conducted there wasn't one time series model which worked for all . The classical models like SES and HW outperformed other models in experiments 1, 8, 10 and 13. While the ARIMA model was the best performing model in experiments 3, 5 and 12. While modern time series forecasting models like Prophet model had better forecasting accuracy in experiments 4, 9 and 11. Artificial recurrent neural network (RNN) like LSTM was able to perform well in experiments 2, 6 and 7. The modern forecasting techniques like Prophet and LSTM weren't the choice of many authors while performing time series analysis and forecasting but the findings of our experiment show promising results which can be explored further. To improve the results of the experiments further the time series data can be transformed by taking the log of the data before it is differenced to make it stationary. The use of Auto ARIMA can be done to find the best performing ARIMA model in the class. While using the LSTM model the stationarity of the data wasn't checked or fixed as the it is not a required that the data be stationary, as the model is capable of making forecast from the non-stationary nature of the data. By making the data stationary before training the LSTM model can help reduce forecasting error and increase the accuracy of the model.

6.16 Results

The results from experiment 1 to 13 provide the best time series forecasting model for each industrial sector. The model used to forecast the carbon dioxide emission for the next six months for each industrial sector has the lowest RMSE value which is used as a comparison metric for the accuracy of the models. The Table 14 has the next six months carbon dioxide forecasting for each sector and the models used for forecasting.

Experiment Number	Sector	Forecast	Model

1	Coal Industrial Sector	2021-05-01 -0.091551 2021-06-01 -0.183102 2021-07-01 -0.274653 2021-08-01 -0.366203 2021-09-01 -0.457754 2021-10-01 -0.549305	SES and HW
2	Coal Coke Net Imports	2021-05-01 -0.24235069 2021-06-01 -0.2454184 2021-07-01 -0.25194994 2021-08-01 -0.25557628 2021-09-01 -0.2571615 2021-10-01 -0.26368517	LSTM
3	Natural Gas Industrial Sector	2021-05-01 -0.029391 2021-06-01 -0.058781 2021-07-01 -0.088172 2021-08-01 -0.117562 2021-09-01 -0.146953 2021-10-01 -0.176343	ARIMA (1,0,1)
4	Distillate Fuel Oil Industrial Sector	2021-05-01 7.1672 2021-06-01 6.408850 2021-07-01 5.682151 2021-08-01 6.306502 2021-09-01 7.074946 2021-10-01 8.437516	Prophet
5	HGL Industrial Sector	2021-05-01 0.003642 2021-06-01 0.007284 2021-07-01 0.010926 2021-08-01 0.014568 2021-09-01 0.018210 2021-10-01 0.021852	ARIMA (1,0,1)
6	Kerosene Industrial Sector	2021-05-01 6.18812606e-03 2021-06-01 4.03763172e-03 2021-07-01 2.01572973e-03 2021-08-01 1.35072857e-03 2021-09-01 9.11830664e-05 2021-10-01 -6.39010876e-04	LSTM
7	Lubricants Industrial Sector	2021-05-01 0.3524816 2021-06-01 0.35345731 2021-07-01 0.35433705 2021-08-01 0.35521024 2021-09-01 0.35647898 2021-10-01 0.35776452	LSTM
8	Motor Gasoline, Excluding Ethanol, Industrial Sector	2021-05-01 0.000203 2021-06-01 0.000406 2021-07-01 0.000609 2021-08-01 0.000812 2021-09-01 0.001015 2021-10-01 0.001218	SES and HW
9	Petroleum Coke Industrial Sector	2021-05-01 7.101282 2021-06-01 6.405419	Prophet

		2021-07-01 6.602608 2021-08-01 6.639935 2021-09-01 6.416279 2021-10-01 7.566746	
10	Residual Fuel Oil Industrial Sector	2021-05-01 -0.052579 2021-06-01 -0.105158 2021-07-01 -0.157737 2021-08-01 -0.210315 2021-09-01 -0.262894 2021-10-01 -0.315473	SES and HW
11	Other Petroleum Products Industrial Sector	2021-05-01 11.113201 2021-06-01 10.784507 2021-07-01 10.911049 2021-08-01 10.938043 2021-09-01 10.006031 2021-10-01 10.286149	Prophet
12	Petroleum, Excluding Biofuels, Industrial Sector	2021-05-01 -0.056745 2021-06-01 -0.113489 2021-07-01 -0.170234 2021-08-01 -0.226978 2021-09-01 -0.283723 2021-10-01 -0.340467	ARIMA (1,0,1)
13	Industrial Share of Electric Power Sector	2021-05-01 0.045649 2021-06-01 0.091299 2021-07-01 0.136948 2021-08-01 0.182598 2021-09-01 0.228247 2021-10-01 0.273896	SES and HW

Table 14 : Experiment Results

7 Conclusion and Future Work

The research focused on finding the best time series model for short-term (six months) forecasting of carbon dioxide emission in the various industrial sectors in the US which would aid the policy makers to implement new policies and take actions on the various sectors. The forecasted data can be used to monitor the various sectors and take preventive actions on high emitting sectors as a response to the rising global climate crises. The raw data was collected from the U.S. EIA website and processed in Jupyter notebook. The time series data was checked for stationarity and non-stationary data was made stationary using differencing. Using SES, HW, ARIMA, Prophet and LSTM models time series forecasting was done on each of the industrial sectors to find which model had the best accuracy. The models were compared using the evaluation metrics of MAE, MAPE and RMSE. The model with the lowest RMSE value was chosen for the short-term forecasting of carbon dioxide emission in each industrial sector. The study found that the traditional ARIMA model wasn't always the best, and that contemporary forecasting models such as Prophet and LSTM produced promising outcomes.

For future work, the auto arima function can be used to select the best ARIMA model in the class. Other models like Stacked LSTM, Bidirectional LSTM and Grey Model can be used for forecasting. Multivariate timeseries forecasting can also be used in the future depending on the availability of the dataset.

Acknowledgment

I would like to express my gratitude to my research supervisor Bharathi Chakravarthi for his support and guidance as well as the priceless constructive feedback provided during the duration of my research project. I would additionally prefer to offer my thanks to my family , friends and research group colleagues for constantly reassuring and supporting me.

References

Amarpuri, L, Yadav, N, Kumar, G & Agrawal, S (2019) 'Prediction of CO2 emissions using deep learning hybrid approach: A Case Study in Indian Context', 2019 Twelfth International Conference on Contemporary Computing (IC3), Contemporary Computing (IC3), 2019 Twelfth International Conference on, pp. 1–6. doi: 10.1109/IC3.2019.8844902.

Ameyaw, B, Yao, L, Oppong, A & Agyeman, JK (2019) 'Investigating, forecasting and proposing emission mitigation pathways for CO2 emissions from fossil fuel combustion only: A case study of selected countries', Energy Policy, 130, pp. 7–21. doi: 10.1016/j.enpol.2019.03.056.

Atique, S, Noureen, S, Roy, V, Bayne, S & Macfie, J (2020) 'Time series forecasting of total daily solar energy generation: A comparative analysis between ARIMA and machine learning techniques', 2020 IEEE Green Technologies Conference (GreenTech), Green Technologies Conference (GreenTech), 2020 IEEE, pp. 175–180. doi: 10.1109/GreenTech46478.2020.9289796.

Chigora, F., Thabani, N. and Mutambara, E. (2019) 'Forecasting CO 2 emission for Zimbabwe's tourism destination vibrancy: A univariate approach using Box- Jenkins ARIMA model', African Journal of Hospitality, Tourism and Leisure, 8(2), p. 15p. Available at: https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,cookie,shib&db=edselc&AN=edselc.2-52.0-85062966150&site=eds-live&scope=site (Accessed: 10 August 2021).

Cristiana Tudor (2016) 'Predicting the Evolution of CO2 Emissions in Bahrain with Automated Forecasting Methods', Sustainability, 8(9), p. 923. doi: 10.3390/su8090923.

Jin, H. (2021) 'Prediction of direct carbon emissions of Chinese provinces using artificial neural networks', PLoS ONE, 16(5), pp. 1–17. doi: 10.1371/journal.pone.0236685.

Kallio, J, Tervonen, J, Räsänen, P, Mäkynen, R, Koivusaari, J & Peltola, J (2021) 'Forecasting office indoor CO2 concentration using machine learning with a one-year dataset', Building and Environment, 187. doi: 10.1016/j.buildenv.2020.107409.

Khan, M.Z and Khan, M.F (2019) 'Application of ANFIS, ANN and fuzzy time series models to CO2 emission from the energy sector and global temperature increase', International Journal of Climate Change Strategies and Management, 11(5), pp. 622–642. doi: 10.1108/IJCCSM-01-2019-0001.

Leerbeck, K, Bacher, P, Junker, RG, Goranović, G, Corradi, O, Ebrahimy, R, Tveit, A & Madsen, (2020) ‘Short-term forecasting of CO2 emission intensity in power grids by machine learning’, Applied Energy, 277. doi: 10.1016/j.apenergy.2020.115527.

Lotfalipour, M. R., Falahi, M. A. & Bastam, M. (2013), ‘Prediction of CO2 emissions in Iran using grey and ARIMA models’, International Journal of Energy Economics and Policy 3(3), 229–237.

Mason, K., Duggan, J. and Howley, E. (2018) ‘Forecasting energy demand, wind generation and carbon dioxide emissions in Ireland using evolutionary neural networks’, Energy, 155, pp. 705–720. doi: 10.1016/j.energy.2018.04.192.

Meng, M., Niu, D. and Shang, W. (2014) ‘A small-sample hybrid model for forecasting energy-related CO2 emissions’, Energy, 64, pp. 673–677. doi: 10.1016/j.energy.2013.10.017.

Nyoni, Thabani and Bonga, Wellington Garikai (2019) ‘Prediction of CO2 Emissions in India Using ARIMA Models’. DRJ - Journal of Economics & Finance, 4(2), pp. 01-10., Available at SSRN: https://ssrn.com/abstract=3346378

Rahman, A. and Hasan, 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. doi: 10.4236/ojs.2017.74038.

Sánchez Lasheras, F, García Nieto, PJ, García Gonzalo, E, Bonavera, L & de Cos Juez, FJ (2020) ‘Evolution and forecasting of PM10 concentration at the Port of Gijon (Spain)’, Scientific Reports, 10(1), pp. 1–12. doi: 10.1038/s41598-020-68636-5.

U.S Energy Information Administration 2008, Carbon Dioxide Emissions from Energy Consumption: Industrial Sector, viewed on 12 August 2012 < https://www.eia.gov/totalenergy/data/browser/?tbl=T11.04#/?f=M>

Yan Li, Yigang Wei and Zhang Dong (2020) ‘Will China Achieve Its Ambitious Goal? — Forecasting the CO2 Emission Intensity of China towards 2030’, Energies, 13(2924), p. 2924. doi: 10.3390/en13112924.

Yang, H. and O, C. J. F. (2020) ‘Short-term carbon emissions forecast for aviation industry in Shanghai’, Journal of Cleaner Production, 275. doi: 10.1016/j.jclepro.2020.122734.

Zhao, X, Han, M, Ding, L & Calin, AC (2018) ‘Forecasting carbon dioxide emissions based on a hybrid of mixed data sampling regression model and back propagation neural network in the USA’, Environmental Science and Pollution Research, 25(3), p. 2899. doi: 10.1007/s11356-017-0642-6.

Zhou, J, Xu, X, Li, W, Guang, F, Yu, X & Jin, B (2019) ‘Forecasting CO_2 Emissions in China’s Construction Industry Based on the Weighted Adaboost-ENN Model and Scenario Analysis’, Journal of Energy. doi: 10.1155/2019/8275491.