**Unveiling the Future of Carbon Emissions Trading: A Machine Learning and Neural Network Perspective on Regional Markets.**

**GROUP - 07**

Sindhu Nagesha (017419987)

Prayag Nikul Purani (017416737)

Syed Faraaz Ahmed (017428619)

Sai Vivek Chunduri (017435301)

Department of Applied Data Science,

San Jose State University

DATA 270: Data Analytics Processes

Submitted to: Dr. Eduardo Chan

March 13, 2024

**Introduction**

* 1. **Background information and Execute Summary**

In the “atomic age of science and technology”, since the 1950’s the world has seen drastic modernization which has led to new jobs, laws, and new experiments. Innovations carry a big stigma behind carbon emissions. Due to 2019’s pandemic, the world saw a huge drop in carbon emission by **5.4% in 2020,** this hit the carbon credit market quite hard. Post-pandemic lifestyle has a huge impact on carbon emissions which affects our environment in all possible manners. Being the main cause, carbon dioxide gas emitted from the outlets of vehicles, industries and so on. Anthony et. al. (2020) in their paper discussed the carbon footprint produced by a machine learning or deep learning model provided during the time of its training, that in peak hours for the training of data, the model produces one-fourth equivalent of carbon dioxide for 100 to 1000 nodes. At this, we can imagine the amount of carbon dioxide being emitted in the training of large language models like GPT-4 which has nearly 1.75 billion neural networks.

The rise in carbon emissions leads to a new problem, “how to handle carbon trade?”

New techniques of carbon trade were developed which led to the invention of the “carbon credit score”. This created a new moment in the world to not go beyond the limits of each country. The nutshell of this system acts like a normal auction of the credit score provided to one particular country to sell their score and follows a bidding war between the countries that place their bid higher and win the race.

The main purpose of the project is to analyze the effect of the pandemic and post-pandemic influencing factors and carbon dioxide being emitted to nature based on the previous data and to predict the carbon credit score that will be required to match the metric ton of carbon dioxide released. The project aims to collaborate with real-time weather data to check the impact of carbon dioxide on air quality, rain timing and intensity, and so on. So, this will help to estimate the metric ton of carbon dioxide that will be emitted in the future, and also, how these changes are or will affect the weather throughout the United States of America.

* 1. **Project Requirements**

The functional requirements of the project include authenticated data by the government of the USA which represents real-time data. In the project, we will be focusing only on the data from the United States of America. As a time frame the dataset begins from the year 1980, the data consists of emissions not only state-wise but also data according to different sectors like industries, electricity, and automobiles. This is followed by incorporating the real-time weather data with a carbon emission dataset, to estimate the effect of carbon emission on weather in the whole United States of America.

Succeeding to data collaboration this study focuses on the creation of the models, which will be accompanied by sensitivity analysis, algorithmic trading, and continuous learning.

* 1. **Project Deliverables**

Current information on carbon emissions and trade in specific regional markets in the USA is very little. So, the data is gathered from different sections and will be processed to develop a one-stop shop for the government to foresee the requirement. Also, a well-collaborated study with the weather data will project a relationship to give a brief knowledge about the effect of carbon emission on weather along with the effect on the country’s economy . The main outcome of the project is to use Machine learning and Neural Network techniques to anticipate and investigate carbon emissions and credit transactions. The evaluation of the project is to gather relevant data from the particular geographical markets that aids the main output. Creating and refining Machine learning and Neural Network models for data analysis is the main focus, which is then tested against each other. Findings and analysis will shed light on the dynamics of the carbon trading system and result in suggestions for policy.

* 1. **Technology and Solution Survey**

This study proposes novel strategies to improve carbon credit prediction accuracy using a combination of advanced techniques. We aim to achieve this by Employing grey Relational Analysis (GRA) and Principal Component Analysis (PCA) to comprehensively examine the data and identify key factors influencing carbon credit values. The dataset consists of driving factors for CO2 emissions across various sectors: commercial, electric, industrial, residential, and transportation, along with weather data and GDP information. This project aims to comprehensively examine the key factors, utilizing machine learning algorithms to predict carbon credit and its effect on weather and Gross Domestic Product (GDP). Utilizing a combination of powerful models such as Convolutional Neural Networks with Iterative Bayesian Filter Adaptation (CNN-IBFA) for capturing complex relationships within the data. Bayes Optimized XGBOOST: This approach leverages XGBoost for predictions, further optimized using Bayesian Optimization as described by Zhu et al. (2023). Hybrid ICSO-SVM: Inspired by Lei & Yang (2020), this model combines the Imperialist Competitive Algorithm (ICSO) with Support Vector Machines (SVM) for robust forecasting. We will incorporate a well-established Artificial Neural Network (ANN) model, similar to the one employed by Acheampong & Boateng (2019) for CO2 emission prediction, to serve as a benchmark for comparison. Following Bhatt et al. (2023), we will utilize a Decision Tree Regressor to assess the critical issue of rising CO2 emissions and their impact on climate change. We will evaluate the effectiveness of our proposed models using standard metrics like Mean Absolute Error (MAE) and Mean Squared Error (MSE).

* 1. **Literature Survey Of Existing Research**

Alshatri and Hussain (2023) in their research, surveyed different papers published in IEEE, SpringerLink, and other journals. Their dataset comprises from 1st January 2005 to 1st July 2023, they have listed 20 factors and 13 target variables to predict the market and the prices in the auction and how it will impact the country's GDP. So, the 20 main factors affecting carbon credit prices are related to non-renewable fossil fuels such as coal, natural gas, petroleum, and other electricity markets. Apart from this, it is also dependent on the air quality and the AQI of the country. The credit also depends on macroeconomics such as the raw material import and export, the stock market, and the currency of the dealing countries. They have also, described that public awareness is also an important factor that could affect the prices because the web search index will analyze the works and if it has some words related to the climate, air quality, and so on then the analysis will tell that the awareness is growing and from the government, support will directly affect demand for carbon credits and increasing the value for those carbon credits.

Guðbrandsdóttir (2011) tells about predicting the prices of carbon credits in the European Union Emission Trading Scheme also, called EU ETS. The description of the dataset used tells us that the data is huge and has many features so to reduce the number of features the process used PCA for feature reduction and to select the important features and later they matched the features with the data from the British energy market and global equity indices. So, the start of the project is done with an analysis of time series to calculate gross return in the one-time frame in between certain intervals and keeping track of how a random variable concerning time and so after that the correlation between the variable is taken into consideration and the basic building block of principal components analysis and after selecting the features in the presence of multicollinearity the OLS estimation parameters where not good so, the next topic or the method used is Latent Root Regression (LRR) which is modified version of OLS and is good with synonyms so, the features with high relation can be easily combined. The implementation of a dumpy variable to decrease the number of dependent variables and to generate the test and train data set so the parameters can be adjusted for - non-predictive near singularities.

The data in this project also has the same features that were the outcome of the study from Alshatri N. and Hussain F. K. (2023) and there is also great overlap in the features and the data selected for the thesis in the current research. The last stage is the prediction phase and the algorithm used were base-line regressions and also, the implementation of PCA using backward elimination and forward selection regression results were compared. This thesis investigates the impact of market relationships on the evolution of carbon prices in the EU ETS, aiming to determine their role as significant driving forces. The study also explores the possibility of making predictions about carbon prices based on these identified relationships.

Lasse et al. (2020), suggested a beautiful paper regarding the carbon footprint for training a deep learning model. As we all know that know-a-days machine learning and artificial intelligence are taking a huge place in the world and to make the model efficient we need to train

it on huge data so, the authors of this project have focused on and have predicted the carbon footprint for training a deep learning model. If we take the example of a Chatgpt it has billions of nodes and the data used to train is the whole world wide web and the training time for the modal is nearly a year. So, back in 2020, the authors thought of building a performance matrix using their tool called “*carbontracker*” so, the outcome could not only build good models but also build an efficient one. The conclusion of the paper focuses on the operational region of low carbon intensity region, hyperparameters can be improved by substitution in grid search, and last energy and efficiency of hardware and setting can help to reduce carbon emissions.

Mao & Yu (2024) in a recent paper researched a hybrid forecasting approach for carbon emissions and how the economy of China is being impacted. The National Carbon Emission Trading Market's formal debut is a major accomplishment for China. Accurately predicting carbon pricing in this market is crucial for businesses to actively engage in the market as well as for the government to create wise laws. Predicting these prices, however, is not an easy task because of the inherent complexity of the market, which includes volatility and instability brought on by a variety of complicated elements. This research proposes a new method for carbon price predictions. It considers several variables that affect national carbon pricing. It presents a framework that integrates many methods, including feature selection, machine learning predictions, and data breakdown and reconstruction. The objective is to improve the efficiency, accuracy, and comprehensibility of carbon price estimates, particularly in light of the intricate and unpredictable nature of the carbon markets. The results of this study are rather intriguing: (1) the novel forecasting framework outperforms existing models in terms of accuracy, and (2) the variables influencing national carbon pricing vary with time. High-frequency series are influenced by short-term energy markets and economic indicators, but medium- and low-frequency series are more heavily influenced by financial markets and long-term economic situations. In addition to improving our knowledge of the factors influencing China's domestic carbon market pricing, this research offers governments and businesses insightful information they can use to develop efficient carbon price prediction tools. It's similar to having a plan in a market that is dynamic and always evolving.

Kunda & Phiri (2017) in their research try to analyze CO2 emissions trends across different sectors in Zambia from 1971-2014, using time series analysis. The main goal was to understand the sources contributing to rising CO2 levels. The findings showed transport, including cars and trucks, caused the highest growth in emissions over the period. Overall, CO2 emissions have substantially increased in Zambia, driven by the transport sector growth as well as manufacturing. The forecast indicates emissions will likely continue rising, especially in the transport and manufacturing industries. Older vehicles were found to be a major factor in transport emissions. The research highlights the specific industries like transport and manufacturing that have disproportionately contributed to rising CO2 levels in Zambia over the past decades. The authors suggest policy interventions targeting cleaner vehicles and sustainable transport, as well as emissions regulations for industry, are needed to curb this trajectory.

Liu et al. (2015) explore how the policy and development state of China’s carbon trade market are emerging and issues affecting the market trade development. In 2013, China started testing a carbon-trading market with five pilot schemes. This was a big step for China to reduce greenhouse gas emissions and show its interest in global carbon trading. However, there are challenges, like not having a clear plan for China's carbon market, mistakes in setting emission limits, and an underdeveloped market system. The Chinese government is more focused on changing companies' behavior than doing a lot of trading in the carbon market. Expanding the market quickly might be hard, and the prices of carbon are a big concern for the Chinese government. China's carbon market is still uncertain. There's a need for more studies to understand how it works, improve the market system, and set clear rules. The government also has to work on making sure companies follow the rules and have a department overseeing everything.

Lu et al. (2020) tell us about China which has the largest carbon emissions globally, trends in their carbon pricing, and trading volumes across markets which are impactful. However, there has been little previous research comprehensively modeling and predicting carbon price and trading volume trends across all of China's carbon markets. Hybrid models achieved the highest accuracy levels, with prediction accuracies of 98.40% and 97.89% respectively on the testing data sets. However, high prediction accuracy alone does not guarantee the stability and reliability of

model predictions over longer periods. Notably, the authors did claim that in addition to accuracy.

Mardani et al. (2020) in their paper succeeded in an efficient multi-stage methodology for predicting carbon dioxide (CO2) emissions in Group of 20 (G20) countries. The methodology utilizes a self-organizing map (SOM) clustering to group countries based on similarities in factors related to CO2 emissions. Adaptive neuro-fuzzy inference systems (ANFIS) and artificial neural network (ANN) models are then constructed for each SOM cluster to predict emissions. Additionally, singular value decomposition (SVD) is employed for dimensionality reduction and handling missing values. The results, based on real-world data, demonstrate the effectiveness of the approach for predicting G20 CO2 emissions, with the SOM-ANFIS-SVD combination providing the highest accuracy (0.065 mean average error). Comparison with other methods shows the superiority of the proposed technique. The analysis stresses the importance of understanding the relationship between economic development, CO2 emissions, and energy consumption for energy and economic policy-making in G20 countries.

Wang & Ye (2017) in their research developed a non-linear grey multivariable model (NGM) to forecast China's carbon emissions from fossil energy consumption, incorporating power exponential terms as exogenous variables to capture complex non-linear relationships with economic growth. Two non-linear programming models are constructed to optimize unknown parameters, minimizing mean absolute percentage error. Empirical analysis of Chinese GDP and emissions data (1953-2013) shows the NGM(1, N) model significantly outperforms traditional grey and autoregressive integrated moving average approaches in accuracy. The model adapts to large sample sizes by dividing data into stages. Scenario analysis applies the NGM(1, N) to quantify China's future emissions under varying economic growth, offering insights into energy and environmental policymaking. Expanding upon the standard GM(1, N) through power exponential terms better describes variable non-linearity. Results emphasize optimizing industrial structure, developing non-fossil energy, and improving efficiency as critical for China's low-carbon economic development. Proposed future work leverages big data technology to enable real-time emissions feedback and proactive energy structure adjustments aligned with China's carbon reduction goals. This study contributes an advanced non-linear modeling approach with policy insights for sustainable growth pathways.

Boateng et al. (2020) developed models to forecast China's carbon emissions stemming specifically from buildings. They used 5 predictive factors: urbanization, R&D spending, GDP, energy consumption, and population. Six machine learning algorithms. The models were trained on 140 observations and validated on 36 observations. The random forest algorithm achieved the highest prediction accuracy at 99.88%, followed by KNN, XGBoost, decision tree, AdaBoost, and SVR. The decision tree was the most time-efficient. Random forest was the top-performing model for accurately forecasting building-related emissions. However, KNN also produced accurate predictions quickly and promptly. The authors recommend random forest, K-nearest neighbors, and decision tree models to policymakers to enable quality, real-time forecasts of China's building emissions patterns. This can inform policy interventions on time.

Li et al. (2022) compared some machine learning algorithms that are used to predict the monetary value of carbon in the HBEA and GDEA provinces of China. The study used historical data on carbon prices and three energy prices (coal, natural gas, and petroleum) in these provinces from May 2014 to July 2021. The three energy prices were considered because the fluctuations in their costs can readily impact the carbon price. The models tested for this prediction are Multivariate machine learning models. During this study, the findings revealed that Multivariate LSTM performed best with minimum values in evaluation metrics for both HBEA and GDEA provinces. Based on this evaluation, they are convinced that Multivariate LSTM has predicted the carbon prices better than the other models used in the study.

Wu et al. (2015) researched to estimate the BRICS Countries carbon dioxide emissions and also to establish a relationship between socioeconomic factors like urban population, growth of the economy, consumption of energy, and emissions of carbon dioxide with an innovative multivariable grey model called GOM (1, N). They chose these BRICS countries because they are major contributors to global emissions. These grey prediction models are useful when analyzing systems with complex and uncertain data. The data was obtained from the World Bank website for the 2004 - 2010 timeframe which has variables like GDP data, Energy use, Urban population, and emissions of carbon dioxide. Results suggest that China had the highest growth rate in terms of carbon emissions followed by India, Brazil, Russia, and South Africa. Increased population in urban areas and energy use directly contributed to higher CO2 emissions. Growth in the economy has led to more carbon emissions in countries like China, India, and South Africa but less in Brazil and Russia. Overall, the GOM (1, N) model has effectively estimated the BRICS nation's carbon dioxide emissions based on the mentioned socioeconomic factors.

Acheampong & Boateng (2019) in their research used an artificial neural network (ANN) model to forecast each nation's(Australia, Brazil, China, India, and the USA) level of carbon emissions. Carbon emission intensity measures the amount of Carbon dioxide released per one unit of GDP. These countries were selected for research because they rank among the highest in the world for carbon dioxide emissions globally. The models used socioeconomic parameters like the growth of the economy, energy use, research and development (R&D), financial development, urbanization, FDI, industrialization, and trade openness as input. Data for the parameters except financial development were collected from the World Bank website for 1980-2015. Data for the Financial Development Index came from the IMF (International Monetary Fund) website for the period 1980Q1-2015Q4. The team used a multilayer perceptron (MLP) algorithm with backpropagation in ANN for better performance. Input data was normalized and then standardized before training. Sensitivity analysis with partial rank correlation coefficient (PRCC) is used to determine which input parameter mostly contributes to the carbon emissions of each country. Results showed that R&D had the largest effect on carbon emissions in Australia, urbanization in Brazil and USA, energy use in India, and population size in China. The team also suggests that the ANN models were highly accurate in predicting carbon emission intensity with negligible errors.

Zhu et al. (2023) proposed using the XGBOOST machine learning algorithm to forecast domestic carbon prices in China. Government organizations allocate specific carbon emission quotas to companies based on their previous carbon emissions data. By predicting prices, companies with surplus carbon credits can sell them to higher-emitting companies to achieve emission reductions cost-effectively. The research utilized carbon trading data from 2013-2021 for major Chinese cities- Beijing, Fujian, Guangdong, Hubei, Shanghai, Shenzhen, Tianjin, and Chongqing from the China Research Data Service Platform (CNRDS). The 10615 total samples were divided into 80% training (10508) and 20% testing (107) sets. The data was then standardized and the Pearson correlation coefficient was calculated to identify and remove highly correlated features via Principal Component Analysis. After that, XGBOOST was applied to make predictions and it is further optimized with Bayes Optimization. The accuracy is then evaluated using MAE and RMSE metrics. Using this, the final prediction results are evaluated, and the team concludes that the Bayes Optimized XGBOOST model is much better than traditional XGBOOST in both accuracy and stability. The research demonstrates the value of machine learning algorithms to predict carbon pricing accurately.

Bhatt et al. (2023) addressed the critical problem of increasing carbon dioxide (CO2) emissions and how they impact climate change in this research. They highlighted the need to accurately predict when Earth's atmospheric carbon dioxide concentration will likely reach 500 ppm, leading to irreparable environmental damage. For this research, they used a historical US dataset with 38 features related to emissions in the US, they then pre-processed the data and applied PCA to reduce it to 17 input parameters. Multiple machine learning models were tested, with the Decision Tree Regressor having the best prediction accuracy at 99%. The study estimated that by 2047, the 500 ppm threshold will have been reached based on current trends. To return to safer levels (316 ppm), they calculated a required CO2 emissions reduction rate of 6.37% and a reversal rate of about 23.38%. The team also mentioned that a 0.75 ppm increase in CO2 causes a 0.05°C temperature rise, meaning a projected total of 5.6°C atmospheric temperature increase by 2047. Overall, the study demonstrates using machine learning to accurately predict the timeline for reaching critical carbon emission thresholds, informing urgent climate action needs.

Lei & Yang (2020) address the pressing issue of growing residential energy use and the resulting CO2 emissions in China, the world's largest emitter of carbon dioxide. In this research, a comprehensive analysis of influencing elements affecting CO2 emissions associated with residential energy use is conducted, offering valuable insights for policymakers and stakeholders. To begin with, the study selects 18 preliminary indicators related to residential energy-related CO2 emissions and employs grey relational analysis to identify their correlations. This approach allows for the identification of key factors strongly associated with CO2 emissions in the residential sector. Subsequently, these influencing factors are classified into four categories based on their characteristics, and PCA is used to simplify the dataset. This process results in the extraction of four components, which serve as input data for the forecasting model. Innovatively, the study proposes a hybrid ICSO-SVM model for the prediction of CO2 emissions, marking the first application of this approach in the field. The model integrates the improved chicken swarm algorithm with support vector machine (SVM) optimization techniques. Through a testing case focused on Shanghai's residential sector, the forecasting results of the ICSO-SVM model are compared with those of other models. The study draws several conclusions based on these comparisons: (a) By considering local electricity and heat emissions coefficients, the study yields a more precise measurement of CO2 emissions related to residential energy use at the city level. (b) Utilizing grey relational analysis and principal component analysis to analyze influencing factors enhances prediction accuracy significantly. (c) The improved chicken swarm algorithm proves effective in optimizing SVM, contributing to the practicality and effectiveness of the ICSO-SVM model. (d) Compared to alternative methods, the newly established ICSO-SVM model demonstrates practicality and promise in forecasting CO2 emissions.Despite the impressive results achieved by this study, several issues remain to be addressed in future research. For instance, there may be additional influencing factors yet to be uncovered, and efforts should be made to minimize information loss during the dimensional reduction process. Additionally, further investigation into prediction models employing more advanced intelligent algorithms is warranted to improve forecasting precision.

**Table 1**

*Literature Survey Summary of factors influencing Carbon Credits.*

| **Authors** | **Goals** | **Algorithms** | **Approach & Outcome** | **Limitations** |
| --- | --- | --- | --- | --- |
| Al Shatri & Hussain (2023) | Data-driven approach of external factors influencing carbon credit prices: A systematic literature review. | Theoretical survey. | A systematic review of the literature to indicate 20 factors influencing carbon credit prices. | Focuses only on papers and not on experiments. |
| Guðbrandsdóttir (2011) | Predicting the price of EU ETS carbon credits: A correlation, principal component,t and latent root approach. | Correlation analysis and principal component analysis (PCA).  Multiple linear regression. | Focuses on the British energy market, global equity indices, and relevant currencies. Dimension reduction techniques, particularly correlation analysis and principal component analysis (PCA), are employed to identify key variables.  Correlation analysis identifies relationships between variables such as Carbon Emission Reductions (CERs), equity indices, and EU Allowances (EUAs).  PCA reveals clusters of variables representing different aspects of energy sources and market indices.  Multiple linear regression models using same-day correlations show promise in predicting EUA price development. | The study aims to examine market relationships and their impact on carbon prices within the EU ETS.  Lagging the data by one business day reduces the predictive power of the model. |
| Lu et al. (2020) | Carbon trading volume and price forecasting in China using multiple machine learning models. | CEEMDAN-RBFNN:98.40% and CEMDAN-GWO-KNEA : 97.89%, | To provide accurate forecasts of carbon prices and trading volumes in China's carbon markets by the implementation of 6 ML model extreme gradient boosting, random forest,a kernel-based nonlinear extension of the Arps decline model optimized by grey wolf optimizer (GWO-KNEA), support vector machine optimized by particle swarm optimizer, support vector machine optimized by fruit fly optimizer and simulated annealing algorithm, and radial basis function neural network (RBFNN). | NA |
| Mao & Yu (2024) | A hybrid forecasting approach for China's national carbon emission allowance prices with balanced accuracy and interpretability. | FS-CEEMDAN-VMD-GWO-LightGBM model  quadratic decomposition technique ,with  **MAE:0.0926**  **RMSE:0.1140**  **MAPE:0.1643**  **R2:0.9928** | Break down carbon price series into high-frequency, medium-frequency, and low-frequency components, reducing complexity and noise interference.  identify critical features and reduce computational complexity while improving model interpretability. | The study needs to consider dynamic social emergencies and external factors such as post-pandemic impacts and geopolitical events like the Russo-Ukrainian War. |
| Liu et al. (2015) | China's carbon-emissions trading: Overview, challenges, and future. | Theoretical study of Carbon emission on a country's GDP. | Reduction potential will be improved mainlyby utilizing technology, such as extensive utilization of nuclear power and renewable energy in the power sector, and the share of coal for [power generation](https://www.sciencedirect.com/topics/engineering/power-generation) reduced to 34% by 2030. | Imperfect trading mechanisms.  Lagging legislation and regulatory systems. |
| Kunda & Phiri (2017) | An approach for predicting CO2 emissions using data mining techniques. | SMOreg algorithm | Average percentage contribution of CO2  emission was 50.9% for manufacturing and construction,  31.7% for transport, 6.7% for electricity and heat production,  7.1% for residential, commercial, and public services, and  3.4% for other sectors in Zambia. In the year 2000, the transport sector surpassed manufacturing and construction as the leading Contributor to CO2 emissions and reached a peak of 51.5% in 2007. | NA |
| Mardani et al. (2020) | A multi-stage method to predict carbon dioxide emissions using dimensionality reduction, clustering, and machine learning techniques. | Adaptive neuro-fuzzy inference system (ANFIS).Artificial neural network (ANN). The singular value decomposition-self-organizing map-adaptive neuro-fuzzy inference system method that achieves an MAE accuracy of 0.065, outperforming multiple linear regression. | The proposed multi-stage approach is evaluated using real-world data from Group 20 nations.  The combination of SOM clustering, SVD dimensionality reduction, and ANN provides a superior accuracy compared to other methods. | NA |
| Wang & Ye (2017) | Forecasting Chinese carbon emissions from fossil energy consumption using non-linear grey multivariable models. | Non-linear grey multivariable model (NGM).  The NGM(1, N) model significantly outperforms traditional grey. | The study proposes a non-linear grey multivariable model to capture the non-linear effects of gross domestic product (GDP) on carbon emissions from fossil energy consumption.  The empirical results demonstrate that the proposed nonlinear grey multivariable model accurately reflects the mechanism of non-linear effects of GDP on carbon emissions from fossil energy (1953 to 2013) consumption.  The model outperforms traditional grey models and autoregressive integrated moving average (ARIMA) models in terms of forecast accuracy. | NA |
| Boateng et al. (2020) | Predicting building-related carbon emissions: A test of machine learning models. | Random Forest (RF): R^2 of 99.88%  K-Nearest Neighbor (KNN):  R^2 of 99.87%  Extreme Gradient Boosting (XGBoost): R^2 of 99.77%  Decision Tree (DT): with R^2 of 99.63%  Adaptive Boosting (AdaBoost):​​  R^2 of 99.56%  Support Vector Regression (SVR): R^2 of 97.67%. | Random forest was the top-performing model for accurately forecasting building-related emissions.  KNN also produced accurate predictions quickly and promptly |  |
| Li et al. (2022) | Forecasting carbon price in China: a multi-model comparison. | Multivariate LSTM(Long Short-Term Memory):  MAE:0.617,  MSE:0.957,  RMSE:0.978  SVR(Support Vector Regression):  MAE:2.776,  MSE:9.381,  RMSE:3.063  MLP(Multilayer Perceptron):  MAE:1.124,  MSE: 1.493,  RMSE: 1.222  RNN (Recurrent Neural Network):  MAE: 0.685,  MSE:1.119,  RMSE: 1.058 | The study used historical data on carbon prices and three energy prices (coal, natural gas, and petroleum) in these provinces from May 2014 to July 2021. The findings revealed that Multivariate LSTM performed best with minimum values in evaluation metrics for both HBEA and GDEA provinces. | NA |
| Wu et al. (2015) | Modelling and forecasting CO2 emissions in the BRICS (Brazil, Russia, India, China, and South Africa) countries using a novel multivariable grey model. | Multivariable grey model called GOM (1, N) | The existing GM(1, N) model uses ﬁrst-order accumulated generating operation sequence. To enhance prediction accuracy, the grey model must focus on new information, and therefore the opposite-direction accumulated generating operator is introduced into the GM(1, N) model.  Growth in the economy has led to more carbon emissions in countries like China, India, and South Africa but less in Brazil and Russia. | NA |
| AcheamPong & Boateng (2019) | Modelling carbon emission intensity: Application of artificial neural network. | Artificial neural network (ANN) | forecast each nation's(Australia, Brazil, China, India, and the USA) level of carbon emissions. Carbon emission intensity measures the amount of Carbon dioxide released per one unit of GDP. | NA |
| Zhu et al. (2023) | Prediction of Carbon Emission Right Price Based on XGBoost Algorithm. | Original-XGBoost:  RMSE:48.42, MAE:5.15, MAPE:0.02  Bayes-XGBoost:  RMSE:1.27, MAE: 0.55, MAPE:0.02 | Forecast the domestic carbon price from 2013 to 2021. The data was then standardized and the Pearson correlation coefficient was calculated to identify and remove highly correlated features via Principal Component Analysis. | NA |
| Bhatt et al. (2023) | Forecasting and mitigation of global environmental carbon dioxide emission using machine learning techniques. | Linear Regression: RMSE:0.183, KNN, Decision Tree: RMSE:0.129, Random Forest Regression, SVM:  RMSE:0.263 | Historical US dataset with 38 features related to emissions in the US, pre-processed the data and applied PCA to reduce it to 17 input parameters. Multiple machine learning models were tested, with the Decision Tree Regressor having the best prediction accuracy at 99%. |  |
| Lei & Yang (2020) | Influencing factors analysis and forecasting of residential energy-related CO2 emissions utilizing optimized support vector machine. | 4 principal factors (4-ICSO-SVM)  MAPE(%):1.21%  RMSE (million tons):0.4346  **CSO-SVM**  MAPE(%):2.37%  RMSE (million tons):0.8282  **PSO-SVM**  MAPE(%):2.77%  RMSE (million tons):0.8322  **GA-SVM**  MAPE(%):3.76%  RMSE (million tons):1.4040  **SVM**  MAPE(%):4.17%  RMSE (million tons):01.3082 | Based on grey relational analysis and principal component analysis, the ICSO-SVM model can highly improve the forecasting accuracy. | The study runs for residential energy-related CO2 emissions at city level only in China. |

This project aligns with the reviewed research Zhu et al., (2023); Lei & Yang, (2020); Guðbrandsdóttir( 2011), which highlights the importance of data preparation. Our proposed study incorporates similar techniques like normalization, standardization,grey scale analysis, and dimensionality reduction using PCA. These methods ensure our models work with clean and efficient data, enhancing their performance. Zhu et al., (2023); Mao & Yu, (2024); Lu et al., (2020); etc., prioritize achieving high accuracy in carbon credit prediction. This aligns with the critical need in this study which is reliable forecasting in the carbon market.Utilizing a combination of powerful models for CO2 emission prediction, proposed by Zhu et al. (2023) described as Bayes Optimized XGBoost, Hybrid ICSO-SVM, inspired by Lei & Yang (2020), well-established Artificial Neural Network (ANN) model, employed by Acheampong & Boateng (2019). Decision Tree Regressors by Bhatt et al.(2023).

The proposed study introduces a novel approach by combining several advanced techniques, a new hybrid model of Convolutional Neural Networks with Iterative Bayesian Filter Adaptation (CNN-IBFA) for capturing complex relationships within the data. Including Bayes Optimized XGBoost, and Hybrid ICSO-SVM, for improving prediction accuracy. While previous studies have predominantly focused on individual models such as XGboost, ANN, or ICSO-SVM for carbon emission prediction, this research stands out by proposing a unique amalgamation of these methodologies. Moreover, unlike studies that concentrate solely on specific sectors like residential or buildings, the project takes a holistic approach by incorporating data from diverse sectors including commercial, electric, industrial, residential, and transportation. This comprehensive analysis enables a deeper understanding of the factors influencing carbon credit values, potentially leading to more robust and accurate predictions.

# **References**

Acheampong, A. O., & Boateng, E. B. (2019). Modeling carbon emission intensity: Application of artificial neural network. *Journal of Cleaner Production*, *225*(225), 833–856. https://doi.org/10.1016/j.jclepro.2019.03.352

Alshatri, N., & Hussain, F. K. (2023, October 19). *Data driven approach of external factors influencing carbon credit prices: A systematic literature review*. Social Science Research Network. https://doi.org/10.2139/ssrn.4602609

Anthony, L. F. W., Kanding, B., & Selvan, R. (2020). Carbontracker: Tracking and predicting the carbon footprint of training deep learning models. *ArXiv:2007.03051 [Cs, Eess, Stat]*. https://arxiv.org/abs/2007.03051

Bhatt, H., Davawala, M., Joshi, T., Shah, M., & Unnarkat, A. (2023). Forecasting and mitigation of global environmental carbon dioxide emission using machine learning techniques. *Cleaner Chemical Engineering*, *5*, 100095.

Boateng, E. B., Twumasi, E. A., Darko, A., Tetteh, M. O., & Chan, A. P. C. (2020). Predicting building-related carbon emissions: A test of machine learning models. *Enabling AI Applications in Data Science*, 247–266. https://doi.org/10.1007/978-3-030-52067-0\_11

Carbon Credits. (2022, January 26). *The ultimate guide to understanding carbon credits*. Carbon Credits. https://carboncredits.com/the-ultimate-guide-to-understanding-carbon-credits/

Chi, Y. J., Pan, H.-Y., Cheng, F.-J., Chang, Y.-I., & Chuang, P.-C. (2022). Experience of carbon monoxide poisoning and the outcome predicting score: A multicenter retrospective study. *The American Journal of Emergency Medicine*, *58*, 73–78. https://doi.org/10.1016/j.ajem.2022.05.012

Guðbrandsdóttir, H. (2011). *Predicting the price of EU ETS carbon credits: A correlation, principal component and latent root approach*. <https://citeseerx.ist.psu.edu/document>?

Hu, B., & Cheng, Y. (2023). Predicting regional carbon price in China based on multi-factor HKELM by combining secondary decomposition and ensemble learning. *PLOS ONE*, *18*(12), e0285311–e0285311. https://doi.org/10.1371/journal.pone.0285311

Huang, Y., & He, Z. (2020). Carbon price forecasting with optimization prediction method based on unstructured combination. *Science of the Total Environment*, *725*(138350), 138350. https://doi.org/10.1016/j.scitotenv.2020.138350

Kenton, W. (2023, November 17). *Carbon credit*. Investopedia. https://www.investopedia.com/terms/c/carbon\_credit.asp

Kunda, D., & Phiri, H. (2017). An approach for predicting CO2 emissions using data mining techniques. *International Journal of Computer Applications*, *172*(13), 7-10.

Li, H., Huang, X., Zhou, D., Cao, A., Su, M., Wang, Y., & Guo, L. (2022). Forecasting carbon price in China: a multi model comparison. *International Journal of Environmental Research and Public Health*, *19*(10), 6217.

Liu, L., Chen, C., Zhao, Y., & Zhao, E. (2015). China׳s carbon-emissions trading: Overview, challenges and future. *Renewable and Sustainable Energy Reviews*, *49*, 254–266. https://doi.org/10.1016/j.rser.2015.04.076

Lu, H., Ma, X., Huang, K., & Azimi, M. (2020). Carbon trading volume and price forecasting in China using multiple machine learning models. *Journal of Cleaner Production*, *249*, 119386. https://doi.org/10.1016/j.jclepro.2019.119386

Mao, Y., & Yu, X. (2024). A hybrid forecasting approach for China's national carbon emission allowance prices with balanced accuracy and interpretability. *Journal of Environmental Management*, *351*(119873), 119873. https://doi.org/10.1016/j.jenvman.2023.119873

Mardani, A., Liao, H., Nilashi, M., Alrasheedi, M., & Cavallaro, F. (2020). A multi-stage method to predict carbon dioxide emissions using dimensionality reduction, clustering, and machine learning techniques. *Journal of Cleaner Production*, *275*, 122942. https://doi.org/10.1016/j.jclepro.2020.122942

Peterdy, K. (n.d.). *Carbon credits*. Corporate Finance Institute. https://corporatefinanceinstitute.com/resources/esg/carbon-credit/

Wang, Z.-X., & Ye, D.-J. (2017). Forecasting Chinese carbon emissions from fossil energy consumption using non-linear grey multivariable models. *Journal of Cleaner Production*, *142*, 600–612. https://doi.org/10.1016/j.jclepro.2016.08.067

Wen, L., & Cao, Y. (2020). Influencing factors analysis and forecasting of residential energy-related CO2 emissions utilizing optimized support vector machines. *Journal of Cleaner Production*, *250*, 119492. https://doi.org/10.1016/j.jclepro.2019.119492

*What are carbon markets?* (n.d.). Www.youtube.com. https://www.youtube.com/watch?v=dHj-H1giztA

Wu, L., Liu, S., Liu, D., Fang, Z., & Xu, H. (2015). Modeling and forecasting CO2 emissions in the BRICS (Brazil, Russia, India, China, and South Africa) countries using a novel multi-variable grey model. *Energy*, *79*, 489-495.

Youcef Djenouri, Asma Belhadi, Srivastava, G., & Lin, J. (2022). Hybrid graph convolution neural network and branch-and-bound optimization for traffic flow forecasting. *Hybrid Graph Convolution Neural Network and Branch-And-Bound Optimization for Traffic Flow Forecasting*, *139*, 100–108. https://doi.org/10.1016/j.future.2022.09.018

Zhao, T., & Liu, Z. (2022). Drivers of CO2 Emissions: A Debt Perspective. *International Journal of Environmental Research and Public Health*, *19*(3), 1847. https://doi.org/10.3390/ijerph19031847

Zhao, Y., Liu, L., Wang, A., & Liu, M. (2023). A novel deep learning based forecasting model for carbon emissions trading: A comparative analysis of regional markets. *Solar Energy*, *262*, 111863. https://doi.org/10.1016/j.solener.2023.111863

Zhou, W., Zeng, B., Wang, J., Luo, X., & Liu, X. (2021). Forecasting Chinese carbon emissions using a novel grey rolling prediction model. *Chaos Solitons & Fractals*, *147*, 110968–110968. https://doi.org/10.1016/j.chaos.2021.110968

Zhu, F., Liu, P., & Hu, P. (2023). Prediction of Carbon Emission Right Price Based on XGBoost Algorithm. *Frontiers in Business, Economics and Management*, *7*(1), 61-67.