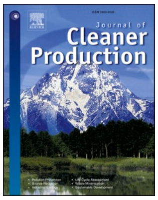
Journal of Cleaner Production 483 (2024) 144309

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro



A study on the carbon emission futures price prediction

Niteesh Kumar [a], Parthajit Kayal [a], Moinak Maiti [b,*]

[a] *Madras School of Economics (MSE), Gandhi Mandapam Road, Behind Government Data Centre, Kottur, Chennai, 600025, India*
[b] *Department of Finance, School of Economics and Finance, University of the Witwatersrand, Johannesburg, South Africa*

ARTICLE INFO

Handling Editor: Giorgio Besagni

Keywords:
Carbon trading
Machine learning
ARIMA model
Economic indicators
Environmental sustainability

ABSTRACT

This study focus on predicting prices of carbon emission futures using a range of methodologies, including traditional ARIMA models and machine learning algorithms. Our analysis uses data from 2005 to 2023 and encompass variables such as the Dow Jones Industrial Average, GDP, crude oil and natural gas futures, per capita carbon emissions, and industrial indices. We find a significant correlation between economic indicators and carbon futures prices. GDP and the Dow Jones have the most influence. Further, we observe that Machine learning models, especially Random Forest Regressor and Gradient Boosting Regressor, outperform traditional ARIMA models in predicting carbon futures prices. This highlights the effectiveness of modern approaches in understanding complex market dynamics. Additionally, the feature importance analysis emphasizes the critical role of economic variables in predicting carbon futures prices. Overall, the study provides valuable insights into the carbon emission futures market and offers implications for stakeholders in managing environmental risks and promoting sustainable development.

1. Introduction

Carbon emission futures are part of the wider market for carbon trading, designed to create economic incentives for companies and large production units to reduce greenhouse gas emissions. By enabling economic agents to purchase and sell their emission allowances, this market aims to encourage eco-friendly practices. Companies with an emission cap may use carbon emission futures to manage the financial risk associated with changes in carbon prices. Utilizing carbon future contracts, one can set the price for future carbon allowances, thereby helping to stabilize their costs.

> *"With over 18 years of operation, the EU ETS, covering about 36% of the European Union's total greenhouse gas emissions, has gained significant popularity among traditional and alternative asset managers. It has become an essential investment vehicle for managing asset risks"* ~ (Subramaniam et al., 2015).

This remains the largest Emissions Trading System (ETS) scheme globally, and with the planned expansion of this scheme and the projected growth in carbon trading elsewhere, it indicates a global market with a turnover in the region of $2 trillion by 2030. This is a significant increase compared to 2018 when the market for carbon credits was worth just $144 billion. Growing environmental policies compel firms to promote sustainable development. For example, under the Paris Agreement, countries have agreed to limit their emissions of greenhouse gases, with the upper limit for global warming being 2 °C; however, 1.5 °C is considered the more desirable goal. Carbon futures have become an indispensable asset for heavy energy-using factories, manufacturing industries, and large production units (Griffin et al., 2016). The management of the increasing risk of carbon futures prices is necessary for these large firms.

The role of Central banks is vital in promoting the transition of the economy towards a neutral-carbon state by providing guidance, setting standards, and, encouraging financial institutions to invest in sustainable projects (Chevallier, 2009). Multiple academic works have developed different models that quantify the carbon emissions cost by assessing individuals' willingness to pay to avoid environmental externalities. Primary goal of this approach is to evaluate the monetary value individuals are willing to pay/sacrifice to alleviate the impacts of carbon emissions. Popular economic valuation methods that researchers can use to estimate the economic costs borne by society due to the environmental consequences of carbon emissions are contingent valuation, choice experiments, and hedonic pricing models, etc. (Mendelsohn and Olmstead, 2009; Pizzol et al., 2015). These models also assist policymakers in formulating effective strategies and regulations to combat climate change and thereby bridging the gap between environmental

---

* Corresponding author.
E-mail addresses: fe23niteesh@mse.ac.in (N. Kumar), parthajit@mse.ac.in (P. Kayal), moinak.maiti@wits.ac.za, maitisoft@gmail.com (M. Maiti).

https://doi.org/10.1016/j.jclepro.2024.144309
Received 21 March 2024; Received in revised form 2 August 2024; Accepted 24 November 2024

conservation and economic growth on a large scale.

The level of economic activity is directly connected with the volume of greenhouse gas (GHG) emissions in the economy. A study by Reboredo (2014) highlights that "the state of macroeconomic indicators and financial markets influences carbon and energy prices." Consequently, the diversification potential of energy futures for carbon assets could yield substantial cost savings for firms. The introduction of the carbon emission futures market has facilitated new interactions within the broader domain of energy-related commodities, encompassing electricity generation, oil, coal, and natural gas production. Despite the challenges posed by significant disasters, the market holds immense potential for energy futures. As additional commodity pricing aligns with climate control targets, it compels trading within the EU emission trading system to evolve with multiple commodity classes, providing a price trend signal to guide investment decision-making. This necessitates a forecasting model capable of predicting future carbon prices. This study aims to leverage machine learning models alongside the traditional Autoregressive Integrated Moving Average (ARIMA) model to predict carbon future prices and compare the results.

This study focuses on exploring the dynamics of price prediction for carbon emission futures traded on the energy exchange. Employing a comprehensive approach, it utilizes both traditional and contemporary modelling techniques, such as the ARIMA model, and various machine learning algorithms including Random Forest Regressor, Gradient Boosting Regressor, K Nearest Neighbour, and Decision Tree Regressor. The dataset, collected on a monthly frequency, spans from June 2005 to December 2023. Key explanatory variables encompass the Dow Jones Industrial Average, United States (US) Gross Domestic Product (GDP), crude oil WTI futures, natural gas contracts, per capita carbon emissions in the US, and the US Industrial Index. These variables, along with GDP and per-capita carbon emissions, have been interpolated to align with a monthly timeframe. The primary objective of the study is to model, enhance predictive accuracy, and discern the factors influencing carbon emission futures prices. The Random Forest Regressor outperforms other models, including the traditional ARIMA model, as indicated by its high R^2 score and low mean squared error. Notably, GDP and the Dow Jones Industrial Average index exhibit the highest feature importance in determining carbon futures prices.

The notable contribution of this study lies in its utilization of an extensive set of explanatory variables such as the Dow Jones Industrial Average, US GDP, crude oil WTI futures, natural gas contracts, per capita carbon emissions in the US, Industrial Indices, manufacturing, crude processing, petroleum & coal products, chemicals, and plastics & rubber products, aiming to predict carbon futures prices. The study endeavours to discern the feature importance that accounts for a high level of carbon emissions or factors that significantly influence carbon futures prices. Notably, the carbon futures price is strongly influenced by GDP, the Dow Jones Industrial Average, and per-capita carbon emissions. Furthermore, the study highlights that with advancements in data science, machine learning models outperform conventional time-series models.

In this study, Section 2 elucidates the existing literature and underscores the importance of this research. Section 3 delineates the data and methodology utilized for the study, detailing the data collection process, pre-processing methods, and the methodologies employed. Section 4 delineates the results and discussions, juxtaposing them with findings from other research endeavours. Finally, Section 5 provides a summary of the research.

2. Literature review

In response to global climate change, carbon futures trading commenced in 2005 within the European Union (EU) ETS, proving to be an effective mechanism for reducing carbon emissions through the cap-and-trade approach. Since its inception, carbon assets have surged in popularity, prompting the need to forecast their prices, particularly among large production units exceeding the imposed cap. This underscores a significant disparity between the social cost estimates of carbon and the monetary value economists assign to carbon emissions, based on current willingness to pay for anticipated future damages. The escalating threats of global warming and the release of greenhouse gases impact various facets of life, primarily attributed to the pursuit of economic growth at the expense of dwindling natural resources (Maiti and Jadhav, 2021; Maiti, 2022a, 2022b; Murali et al., 2023). Zhu et al. (2017) underscored the critical importance of accurately predicting carbon futures prices for various stakeholders, including researchers, policymakers, and financial investors, given its significant impact on most human and economic activities.

This urgency is further compounded by the consensus reached by states in Copenhagen and Durban, aiming to limit the planet's average temperature increase to 2 °C. As per environment experts, this goal can be achieved by reducing greenhouse gas emissions by at least 50% by 2050 (Bunn and Fezzi, 2007). Long-term risk of global warming could adversely affect financial asset valuation (Alberola et al., 2009). Drawing on economic analysis, Christiansen et al. (2005) identified a few key determinants of prices in the EU ETS such as strategy and regulatory issues, market fundamentals, the role of fuel switching, weather patterns, and production levels. Empirical studies by Alberola et al. (2008) and Christiansen (2005) econometrically uncover the relationships between energy markets and carbon prices.

Over last two decades, numerous studies have utilized traditional and advanced time series models to forecast the price of carbon futures. For instance, study by Paolella and Taschini (2008) used a Generalized Autoregressive Conditional Heteroskedasticity (GARCH) structure approach to model the conditional dynamics of carbon EU Emission Allowance (EUA) carbon futures returns. Similarly, Byun and Cho (2013) applied various GARCH-type models (not just one GARCH model) to predict carbon future price dynamics in the European Climate Exchange (ECX) market. Their study compared the performances of those different GARCH-type models. Further, using ARMAX-GARCH[1] approach, Sanin et al. (2015) exhibited future price prediction in the EU ETS. The work by Benz and Trück (2009) employed a relatively advanced model like Markov switching and AR-GARCH to capture the various behaviours of EUA carbon futures contracts return volatility in the EU ETS. They also compared the performance of their model with traditional prediction models. In a recent study, using Vector Autoregressive (VAR) models, Gong et al. (2023) attempted to predict similar EUA futures contracts. Similarly, Koop & Tole (2013) forecasted future prices of carbon permits in the EU ETS using the dynamic averaging method. Their work also discussed the forecasting ability of fundamental and institutional factors of carbon. The studies predominantly utilized traditional statistical and econometric models like ARMA, ARIMA, GARCH, VAR, etc. for forecasting. Recent studies adopted advanced time series approaches like regime switching (Çanakoğlu et al., 2018), DCC-GARCH[2] (Hoque et al., 2023), GJR-GARCH-MIDAS[3] model (Niu and Liu, 2024), etc.

The dynamics of carbon futures prices are often characterized by non-linearity and irregularity. It has prompted the adoption of non-linear artificial intelligence statistical models in recent years. For instance, Fan et al. (2015) forecasted EUA carbon future price movements using artificial intelligence ensemble techniques. They achieve particularly improved performance through the multi-layered perception artificial neural network (MLP ANN) approach. The work by Zhu et al. (2015) utilized ensemble empirical mode decomposition-based least square support vector machines (LSSVM) in the similar context

---

[1] ARMAX is an acronym for AutoRegressive Moving Average model with eXogenous inputs.

[2] DCC refers to Dynamic Conditional Correlation.

[3] GJR-GARCH-MIDAS- The full form of GJR-GARCH-MIDAS Glosten-Jagannathan-Runkle Generalized Autoregressive Conditional Heteroskedasticity - Mixed Data Sampling.

while comparing their efficacy to conventional forecasting methods. The superiority of machine learning models in carbon future price predictions are validated in various studies (Huang et al., 2021; Shahzad et al., 2023; Zhu et al., 2022).

In recent work, various hybrid models (that combine traditional econometric models with machine learning models or joint multiple machine learning models) have been proposed. This is to enhance the models' forecasting capability in the context of carbon prices. For instance, Ji et al. (2019) united ARIMA with machine learning models such as convolutional neural networks (CNN) and long short-term memory networks (LSTM) for carbon futures price forecasting. Similarly, Huang et al. (2021) combined the GARCH and LSTM models for the same purpose. Chen et al. (2022) introduced a mixture method by incorporating fuzzy entropy and extreme learning machine model to decompose carbon futures prices into several intrinsic mode functions (IMFs) and one residue. Their attempt was to achieve superior performance in forecasting carbon futures prices. A recent work by Zhang et al. (2023a) attempted multi-step carbon price forecasting using a multivariate decomposition strategy and deep learning algorithms. Other few notable works that employ advanced mathematical techniques or machine learning algorithms in the context of carbon futures price prediction are Gong et al. (2023), Huang and He (2020), Qin et al. (2020), Zhao et al. (2024), etc.

Advancements in machine learning include the development of support vector machines (SVM) (Cortes and Vapnik, 1995). These machines are designed to overcome the problem of local optimal solutions commonly encountered in artificial neural networks (ANN). It is possible to achieve global optimal solutions by employing SVM. Overall, the use of machine learning models could significantly enhance the accuracy of carbon price predictions (Huang et al., 2021; Shahzad et al., 2023; Zhu et al., 2022).

3. Data & methodology

3.1. Data collection

3.1.1. Response variable

The carbon emission futures prices serve as the response variable or dependent variable in this study. Data collection is facilitated through investing.com, a renowned platform for financial asset price data. Carbon emission futures belong to the energy group of commodities and are traded on energy exchanges at 1000 tons per contract, with a tick size of 0.01. This indicates the increment by which the price of the commodity instrument changes.

3.1.2. Explanatory variable

The previous studies have shown that carbon future prices are related to the following variables namely Dow Jones Industrial Average (DJIA), GDP, crude oil and natural gas futures, per capita carbon emissions, and industrial indices. The Dow Jones Industrial Average and GDP serve as indicators of overall economic health and investor sentiment (Ball and French, 2021). During periods of economic prosperity and high consumer demand, industrial activity and production of goods tend to increase, impacting the necessity for carbon futures trading (Bielinskyi et al., 2021; Wang and Dong, 2023). Similarly, GDP growth rate is linked to economic growth and industrial productivity. However, studies show mixed results on relationship between GDP growth and carbon emissions (Zou, 2018; Lin and Jia, 2019). According to Zou (2018) GDP growth not leads to increase in carbon emissions. Whereas the study by Lin and Jia (2019) showed that carbon ETS is related to the GDP growth. Further, crude oil and natural gas are the key drivers of carbon prices as they can affect the energy consumption and production pattern (Van Ruijven and Van Vuuren, 2009; Chevallier, 2011; Liu et al., 2023). In the context of financial market, per capita carbon emissions and industrial indices are the crucial factors that affects carbon prices. A higher amount of per capita carbon emissions shows a greater carbon emissions and demand for carbon incentives (Alberola et al., 2008; Parker and Bhatti, 2020; Li et al., 2021).

The global refining carbon intensity varies between 13.9 and 62.1 kg of CO_2-equivalent (CO2e) per barrel at the country level. At the crude level, it ranges from 10.1 to 72.1 kg of CO2e per barrel (Brandt et al., 2020). Crude oil is a primary energy source used for transportation, industrial processes, and electricity generation (Maiti and Kayal, 2022). The extraction, refining, and consumption of crude oil involve energy-intensive processes often reliant on additional fossil fuel combustion, contributing to carbon emissions directly and indirectly throughout the oil supply chain. Plastics exhibit the most significant production growth among all bulk materials and already account for 4.5% of global greenhouse gas emissions. A circular economy approach, coupled with additional emphasis on the bioeconomy, can reduce resource consumption by 30% and achieve 10% greater emission reductions by 2050, while minimizing the potential for long-term negative emissions (Stegmann et al., 2022). The chemical industry impacts numerous sectors, with over 95% of all manufactured goods incorporating chemical products. Dependent heavily on fossil-based resources for feedstock and energy, the chemical industry contributes 5–6% of global GHG emissions, equivalent to 2.9 Gt CO_2 eq. annually (World Economic Forum, 2021).

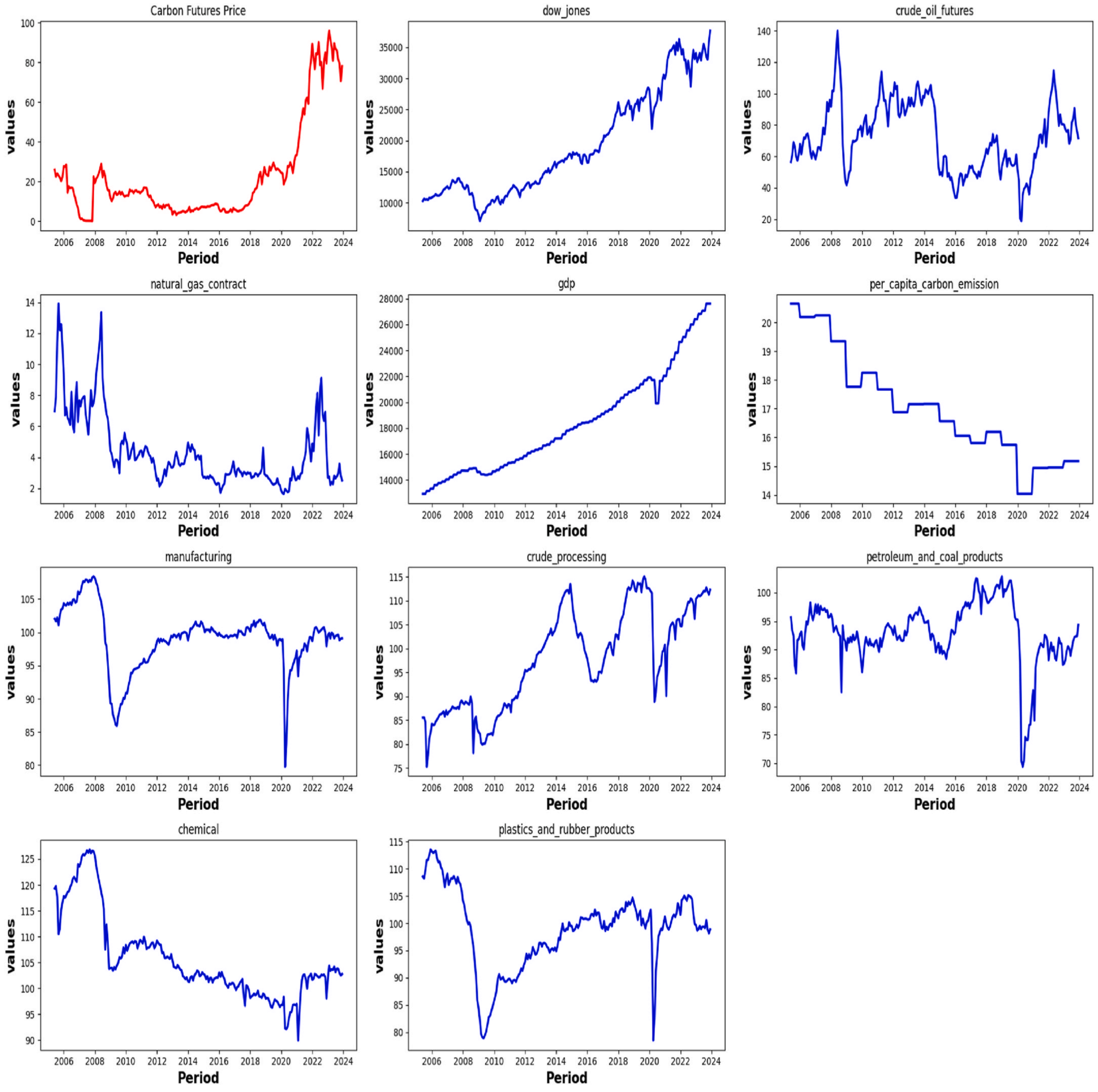
Based on the above information the present study also considers the following variables: manufacturing (An et al., 2021; Kim and Bae, 2022), crude processing (Abdul-Manan et al., 2017), petrol and coal products (Burgess, 1990), chemicals (Yadav et al., 2020), plastics, and rubber products (Hertwich, 2021). Thus, explanatory variables that are considered for this study include the Dow Jones Industrial Average, GDP (USA), crude oil WTI futures, natural gas contracts, per capita carbon emissions in the US, and the Industrial Index (USA), encompassing manufacturing, crude processing, petrol and coal products, chemicals, plastics, and rubber products. Data for these variables is sourced from various open-access platforms, including investing.com, federalreserve.gov, the Bureau of Economic Analysis, and ourworldindata.org, a reputable organization for data analysis. The dataset spans from June 2005 to December 2023, providing monthly data points. There appears to be a noticeable trend in all the prices in the dataset, and each variable is susceptible to seasonal variations (refer to Fig. 1). Specifically, variables such as the Dow Jones Industrial Average, GDP, crude processing, plastics and rubber products, and carbon futures exhibit a clear uptrend. Conversely, variables like natural gas contracts, per-capita carbon emissions, and chemicals demonstrate a distinct downtrend. Meanwhile, the remaining variables show consolidation patterns throughout the data period.

3.1.2.1. Feature engineering. The dataset includes missing values in two specific columns: GDP and per-capita carbon emissions. GDP data is only available on a quarterly basis, while per-capita carbon emissions data is only available annually. Therefore, linear interpolation is employed to address the missing values. This method operates under the assumption of a linear relationship between the available data points and fills in the empty cells accordingly. Following interpolation, no null variables remain in the dataset.

Although the data exhibits values in different ranges, standardization of the data is not conducted (refer to Fig. 2). The rationale behind this decision will become apparent from the selection of computational models.

3.2. Methodology

This section provides an overview of the models utilized for predicting carbon futures prices. We decided not to standardize the data or remove outliers because of the selected models are robust. These models are good at handling outliers and effectively interpret intricate data patterns.


Carbon Futures Price

dow_jones

crude_oil_futures

natural_gas_contract

gdp

per_capita_carbon_emission

manufacturing

crude_processing

petroleum_and_coal_products

chemical

plastics_and_rubber_products

Fig. 1. Line plot of variables in the data.

3.2.1. *Data-partitioning*

To achieve robust predictions from the machine learning models, we devised the dataset into two subsets: training and testing data. The partition size is set to 80% for training and 20% for testing, a standard practice achieved using scikit-learn. Additionally, for the ARIMA model, seasonality is extracted from the dataset, and the remaining data is employed for prediction purposes.

3.2.2. *Model selection*

Previous studies on carbon emission futures price prediction have employed various techniques, including advanced econometrics (Liu and Huang, 2021), artificial intelligence (Yahşi et al., 2019; Shahzad et al., 2023), hybrid models (Huang et al., 2021; Zhang et al., 2023b), and other innovative approaches (Li et al., 2022). The choice of models in these studies depended on the data characteristics and the specific context of the problem. Following a similar approach, this study selects different models based on criteria such as computational efficiency, data requirements, and simplicity and interpretability. The ARIMA model is chosen because it is widely used in finance for predictions (Maiti, 2021). Additionally, four supervised learning models are selected: Random Forest (Ensemble Learning), Gradient Boost (Ensemble Learning), K Nearest Neighbour (Instance-Based Learning), and Decision Tree (Tree-Based Model).

ARIMA The ARIMA model is a popular and effective technique for time series forecasting. It consists of three main components: autoregression, integration, and moving average. The autoregressive component uses past values of the time series to predict future values. The integration component incorporates past forecast errors into the model.

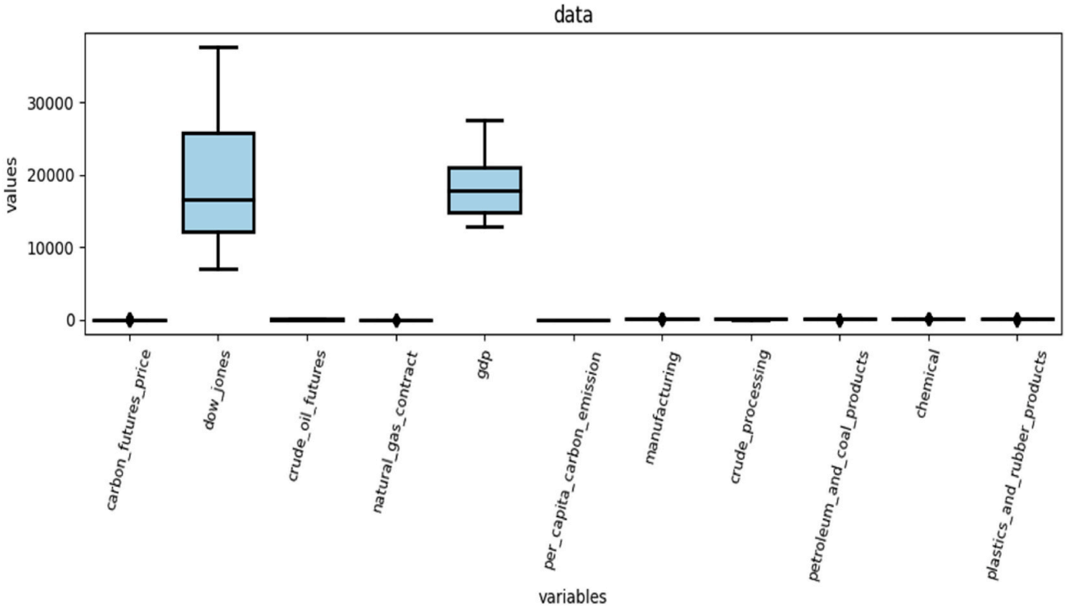


Fig. 2. Boxplot of dataset.

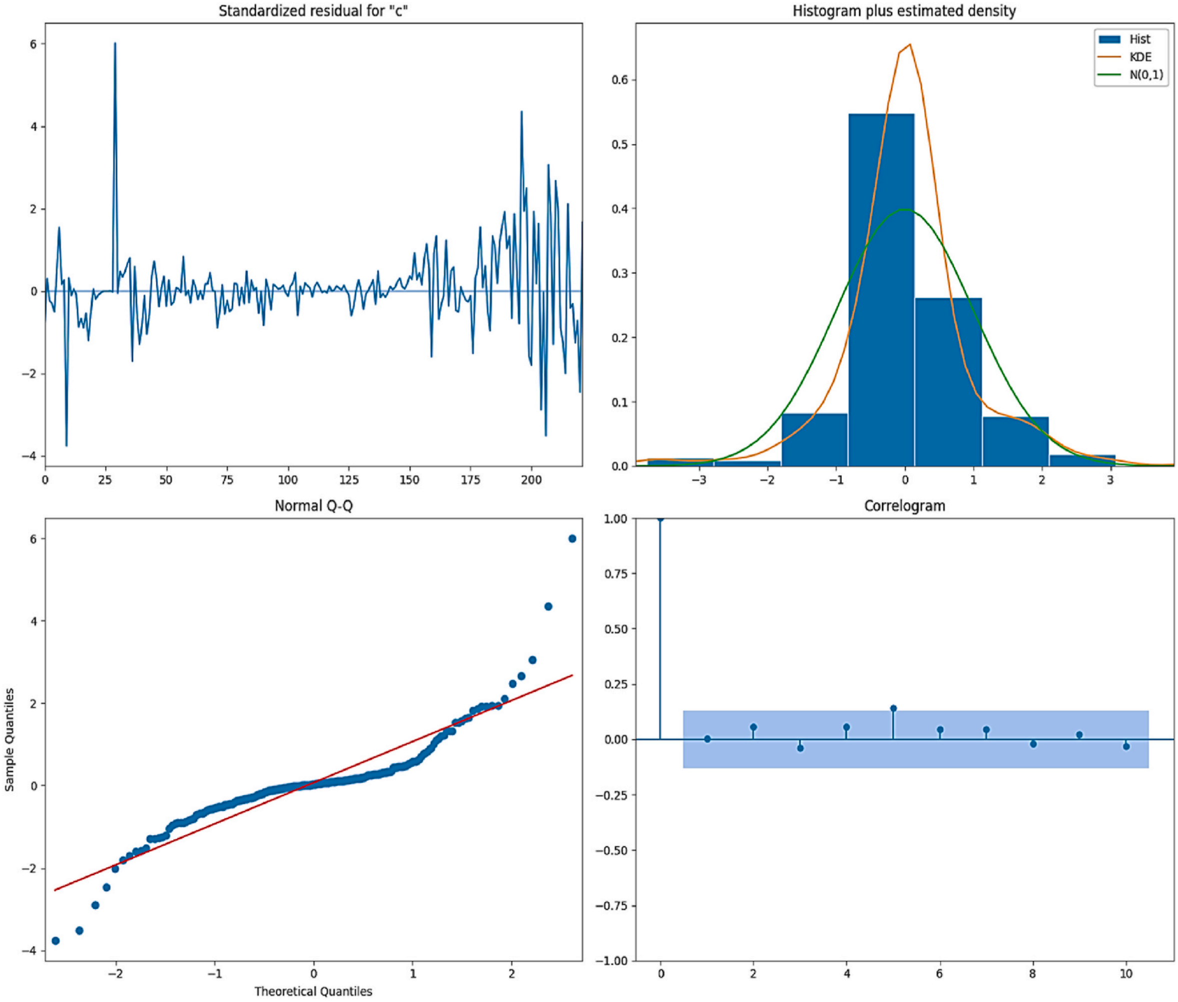


Fig. 3. Boxplot of dataset AIC test Plot.

It is to account for errors that may occur at different times. This helps in improving prediction accuracy (Ji et al., 2019). ARIMA model predicts future data points by analyzing its own historical data including lagged observations and past errors.

Before applying the ARIMA model, it is essential to ensure that the data is stationary. It means that statistical properties like mean and variance remain constant over time. This stationary property is verified to ensure that the model can accurately capture complex data relationships and provide precise predictions. We select the order of the ARIMA model based on the Akaike Information Criterion (AIC) test and test the stationarity using the Augmented Dickey-Fuller test. For this study, the chosen model is ARIMA (1, 1, 0). Detailed estimates are provided in Appendix. Fig. 3 indicates that there is less of a problem of heteroscedasticity. the standardised residual plot shows that variance remains relatively constant. The correlogram exhibit that all residuals fall within the significance level.

3.2.2.1. Random Forest. The Random Forest algorithm is a widely used machine learning method in the realm of supervised learning. It employs ensemble learning by combining multiple decision trees to analyze and solve complex data patterns (Ranjan et al., 2023; Vaidynathan et al., 2023). In a Random Forest regressor, the data is divided into various subsets. Each subsets represents individual trees or nodes. The predictions from these trees are then averaged to improve accuracy. A notable feature of Random Forest is its capability to randomly select subsets of features for each tree to optimize feature selection and thereby enhances prediction accuracy. In our model, we use 200 trees, each with a maximum depth of 20 and a sample split size of 2. Increasing the number of trees in the forest generally leads to improved prediction accuracy.

3.2.2.2. Gradient Boost. The Gradient Boosting algorithm (also known as Gradient Boost Machine or GBM) is renowned for its computational speed and predictive accuracy. This is important particularly when dealing with large and complex datasets (Ranjan et al., 2023; Vaidynathan et al., 2023). This algorithm focuses on minimizing prediction bias to enhance overall accuracy. As an ensemble technique, Gradient Boosting combines the predictions of multiple weak learners to refine the overall predictive performance. It does this by assigning weights to incorrect predictions made in previous iterations, gradually reducing prediction errors and improving accuracy over successive iterations. In our model, we employ 1000 estimators (weak learners) to predict the data.

3.2.2.3. K nearest neighbour. The K Nearest Neighbour (KNN) is known as the lazy learner algorithm. It is a supervised learning model. KNN retains all the provided data (training data) without performing any processing. It predicts the test data by comparing its similarity with the training data (Ranjan et al., 2023; Vaidynathan et al., 2023). The Euclidean distance metric is typically used in KNN to measure similarity and make predictions based on it. Specifically, KNN calculates the distance between the training dataset and the test data as one of the pieces of information to predict the outcome. In our model, we utilize a single nearest neighbour ($k = 1$) for predictions, ensuring simplicity and quick responsiveness.

3.2.2.4. Decision tree. A decision tree is a versatile supervised learning algorithm that organizes data into a hierarchical tree structure. Each internal node represents a feature or independent variable. Branches depict decision rules based on these features. The goal is to make prediction, provided at each leaf node (Ranjan et al., 2023; Vaidynathan et al., 2023). The model splits the data according to certain criteria, such as minimizing absolute error, to enhance prediction accuracy. One key advantages of decision trees over ARIMA models are their ability to handle non-linear relationships. Additionally, decision trees offer insights into the importance of different features to understand the factors driving the time series. Decision trees are more robust to outliers, as they focus on creating homogeneous groups rather than fitting a global model. They can handle both numerical and categorical data, adding to their versatility.

4. Results and discussion

It is apparent that there are both positive and negative correlations between variables and carbon futures prices. Notably, Dow Jones and GDP exhibit the strongest correlations with carbon futures prices, at 75.84% and 77.83%, respectively. The positive correlation can be attributed to overall economic expansion. A robust performance of the Dow Jones industrial average typically indicates a flourishing economy, which may spur heightened industrial activity, greater energy consumption, and increased demand for carbon-intensive products. Consequently, this surge in carbon emissions positively impacts carbon futures prices.

4.1. Correlation

Table 1 shows the correlation among the study variables. The estimates show that the carbon future prices are positively related to Dow jones, crude oil futures, natural gas contract, GDP, crude processing, and plastics & rubber products. Whereas carbon future prices are negatively related to per capita carbon emission, manufacturing, petroleum & coal products, and chemical. The negative correlation may be attributed to various factors such as the gradual reduction in per capita carbon emissions over time, stricter environmental policies leading to decreased production of petroleum and coal products, a shift towards sustainable energy sources, adoption of greener alternatives, advancements in technology, and changes in production methods. The correlations among the other variables are shown in Table 1, and are self-explanatory.

4.2. Descriptive statistics

Table 2 presents the descriptive statistics of the variables. The median of most variables, including crude oil futures, natural gas contracts, per capita carbon emissions, manufacturing output, crude processing, petroleum and coal products, and plastics and rubber products, is closely aligned with the mean. Only a few variables, such as the Dow Jones index, carbon futures prices, GDP, and chemicals, show slight deviations from their mean. Overall, the data indicate that most variables exhibit a symmetrical to slightly skewed distribution.

4.3. Empirical analysis

We present the empirical analysis and results for all the models employed in predicting carbon futures prices (Refer to Table 3).

4.3.1. R^2 score

R^2 is a widely used metric in machine learning for assessing the performance of a regression model. It is also known as the coefficient of determination. R^2 quantifies the proportion of variance in the dependent variable that is explained by the independent variables in the model. An R^2 value close to 1 means that the model accurately predicts the dependent variable by capturing most of the variability. Although, R^2 offers insights into the goodness of fit of the model, it may not effectively capture nonlinear relationships.

4.3.2. Mean squared error

Mean squared error (MSE) is another frequently used metric in machine learning. It measures the squared difference between the predicted values and the actual values in a regression problem. MSE asses the overall accuracy or performance of a prediction model for a given

Table 1
Correlation between carbon futures and explanatory variables.

	carbon futures price	dow jones	crude oil futures	natural gas contract	gdp	per capita carbon emission	manufacturing	crude processing	petroleum and coal products	chemical	plastics and rubber products
carbon futures price	1.0000										
dow jones	0.7584	1.0000									
crude oil futures	0.1420	−0.1655	1.0000								
natural gas contract	0.0475	−0.3485	0.3740	1.0000							
gdp	0.7783	0.9637	−0.1428	−0.4208	1.0000						
per capita carbon emission	−0.4593	−0.8032	0.2651	0.6944	−0.8494	1.0000					
manufacturing	−0.0353	0.0207	0.2100	0.4302	−0.0663	0.3732	1.0000				
crude processing	0.4051	0.7508	−0.0305	−0.4969	0.7822	−0.7124	0.1665	1.0000			
petroleum and coal products	−0.2882	−0.1393	0.0794	0.0099	−0.1381	0.2354	0.4891	0.1840	1.0000		
chemical	−0.2119	−0.5810	0.3457	0.7144	−0.6368	0.8807	0.5291	−0.6147	0.1841	1.0000	
plastics and rubber products	0.1640	0.2596	−0.0935	0.3920	0.1346	0.1824	0.8624	0.2065	0.2895	0.3214	1.0000

Table 2
Descriptive statistics.

	Mean	Std	Median	min	max
Carbon futures price	22.75951	24.58599	14.25	0.01	95.88
Dow jones	19125.76	8404.375	16580.84	7062.93	37689.54
Crude oil futures	72.41574	21.72156	70.61	18.84	140
Natural gas contract	4.411915	2.381960	3.653	1.64	13.9210
GDP	18349.78	3926.153	17804.20	12922.7	27610.10
Per capita carbon emission	17.01362	1.832459	16.87688	14.0341	20.65803
manufacturing	99.06393	4.618765	99.54150	79.7607	108.3983
Crude processing	97.74922	10.48170	98.61080	75.2859	115.0372
Petroleum and coal products	92.96686	5.618768	92.80320	69.4136	102.8569
Chemical	105.6860	8.214284	103.0104	89.8926	126.8536
Plastics and rubber products	98.71483	7.304498	99.65360	78.5025	113.5542

Table 3
Empirical results.

Model	R^2 Score	Mean squared error
Random Forest Regressor	0.99	6.54
Gradient Boosting Regressor	0.98	9.24
K Neighbours Regressor	0.97	13.61
ARIMA	0.97	16.74
Decision Tree Regressor	0.96	21.86

dataset. The unit of MSE is the square of the unit of the target variable.

4.3.3. Carbon futures price

In this analysis, we examine the dynamics of carbon futures prices. These prices are key indicators in the environmental commodities market and vital to the supply side. Carbon trading has significantly increased, rising from 7800 tons at $18.75 per ton in July 2005 to a peak of 739,470 tons at $84.43 in March 2022. The random forest regressor proved to be the most effective, achieving the highest R^2 value (0.99) and the lowest MSE (6.54).

Although the ARIMA model is well-regarded in time series analysis, machine learning models such as random forest regressor, gradient boosting regressors, and K-neighbours regressors outperformed the traditional ARIMA model. These machine learning models demonstrated higher R^2 value and lower MSE. Our results are quite consistent with recent studies like Zhang et al. (2017), Siami-Namini et al. (2018), Kontopoulou et al. (2023), and Ning et al. (2022). However, when we comparie the performance of different machine learning models, we observe that the decision tree regressor has performed relatively lower than the other models. This discrepancy can be attributed to the fact that models like random forests and gradient-boosting regressors excel by combining weak learners, whereas decision tree models operate independently. Additionally, as mentioned earlier, decision tree models tend to underperform on unstandardized data, which is evident from the results. Fig. 4 illustrates the actual and predicted carbon future prices. It is evident that all the models have accurately captured the actual and predicted prices, resulting in a close match. This alignment is due to all the models achieving more than 90 percent R^2 and a decent mean squared error.

4.3.4. Feature importance

Fig. 4 illustrates the actual versus predicted carbon futures prices. It is evident that the models have closely aligned the actual and predicted prices, as indicated by an R^2 value exceeding 90% and a low mean square error across all models. Describing feature importance (Fig. 5) in machine learning models involves communicating the significance of each feature's contribution to the model's predictions (Zien et al., 2009). In our model, GDP contributes the most to predicting carbon futures prices, with a weight of 32.14%. The Dow Jones index follows closely, accounting for 23.79% of the model's predictive power. Per capita carbon emissions contribute 22.92%, while crude processing accounts for 7.96%. Other features, including chemicals, crude oil futures, manufacturing, natural gas contracts, petroleum and coal products, and plastics and rubber products, each contribute less than 5% to the model's predictive weight.

5. Concluison

This study focus on predicting prices of carbon emission futures which is a crucial tool in the environmental commodity market. Carbon emission trading has gained popularity as a traditional investment and a modern hedge against climate change and global warming risks. This trend aligns with the European Union emission trading systems. Environmental policies and initiatives, such as the Paris Agreement, Kyoto Protocol, and COP meetings, have increased awareness of carbon futures' importance in the sustainable environmental market (Seo, 2017). Carbon emission futures are valuable assets for energy-intensive industries and large production units. They play a significant role in the growth of developing economies (Yuan and Zhao, 2016).

Our empirical analysis reveals that machine learning models, particularly random forest regressors and gradient boosting regressors, perform robustly. We observe that these models surpass the traditional ARIMA time series model in terms of forecasting ability. They achieve

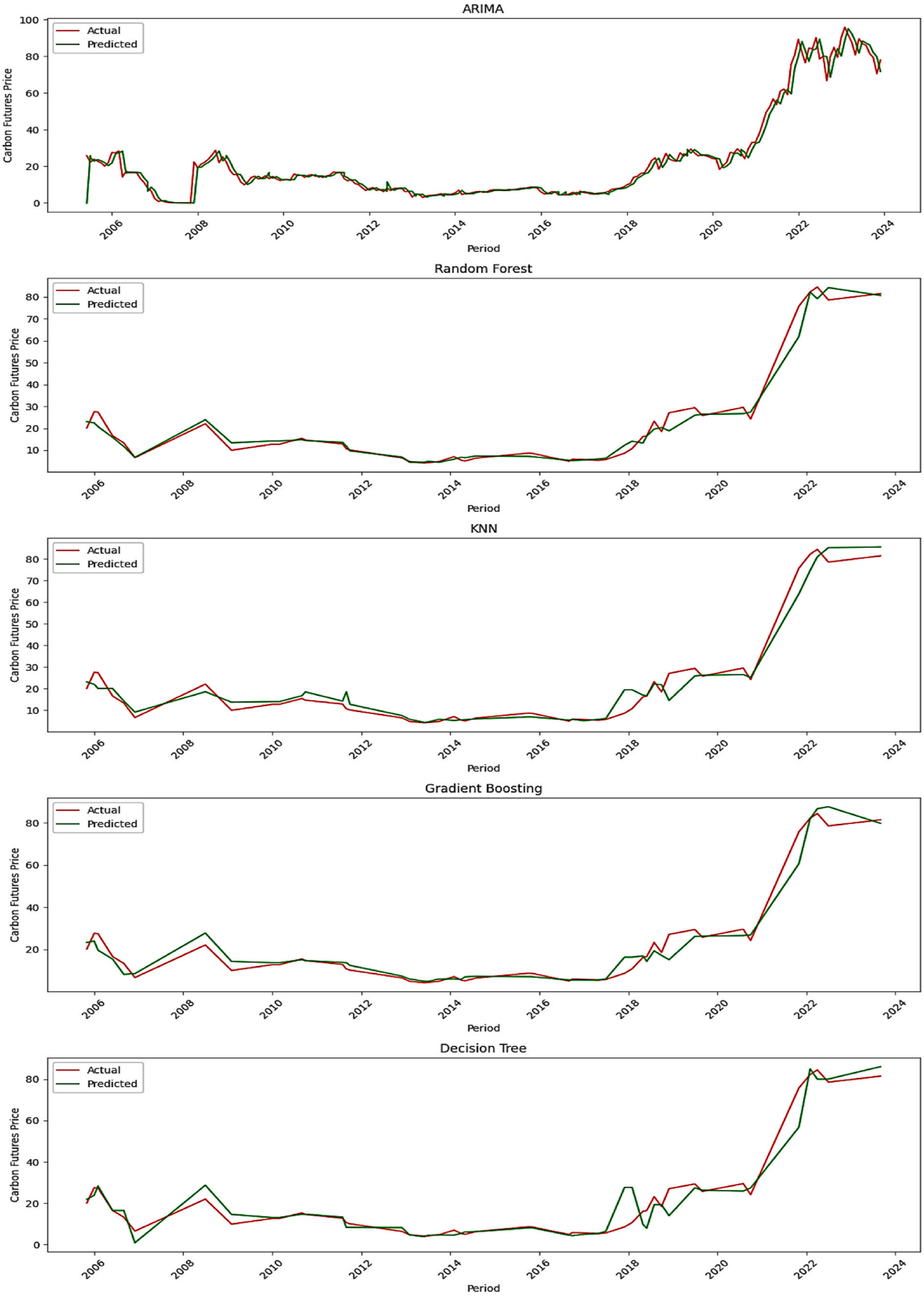


Fig. 4. Comparison of predictions of various models.

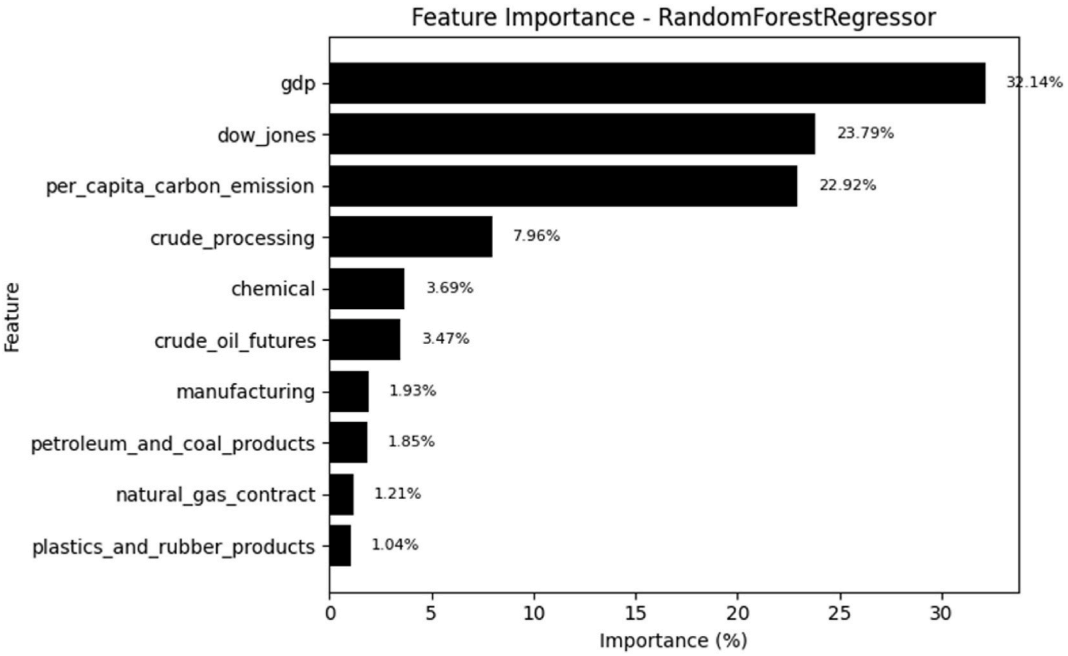


Fig. 5. Feature importance of independent variables.

higher R^2-scores and lower MSE. The study also highlights the significance of features in predicting carbon futures prices. GDP and Dow Jones are prominent features, followed by per capita carbon emissions.

CRediT authorship contribution statement

Niteesh Kumar: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Parthajit Kayal:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Formal analysis, Conceptualization. **Moinak Maiti:** Writing – review & editing, Writing – original draft, Supervision, Resources, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

There is no conflict of interest.

Appendix

Dickey-Fuller Test Results:

ADF Statistic:	−16.98400588
p-value:	0.00000000000000000000000000000089692
Critical Values:	
1%:	−3.460290639
5%:	−2.87470868
10%:	−2.573788599

P < Significance level.
H_0 rejected, the data is stationary.

AIC Test:

Tested ARIMA (0, 1, 0)	AIC: 1220.321667261503
Tested ARIMA (0, 1, 1)	AIC: 1218.6131359748888
Tested ARIMA (0, 1, 2)	AIC: 1219.4863465476535
Tested ARIMA (0, 1, 3)	AIC: 1220.8601782666806
Tested ARIMA (0, 1, 4)	AIC: 1221.0804291515851
Tested ARIMA (1, 1, 0)	AIC: 1218.0701459600823
Tested ARIMA (1, 1, 1)	AIC: 1218.9491046932349
Tested ARIMA (1, 1, 2)	AIC: 1220.817020038106
Tested ARIMA (1, 1, 3)	AIC: 1221.7668593106464
Tested ARIMA (1, 1, 4)	AIC: 1218.9855630571533
Tested ARIMA (2, 1, 0)	AIC: 1219.0215953569063
Tested ARIMA (2, 1, 1)	AIC: 1220.842193200638
Tested ARIMA (2, 1, 2)	AIC: 1221.5687865618634

(*continued on next page*)

(*continued*)

Tested ARIMA (0, 1, 0)	AIC: 1220.321667261503
Tested ARIMA (2, 1, 3)	AIC: 1223.3832229555792
Tested ARIMA (2, 1, 4)	AIC: 1219.4083853961456
Tested ARIMA (3, 1, 0)	AIC: 1220.795154418359
Tested ARIMA (3, 1, 1)	AIC: 1222.765101108996
Tested ARIMA (3, 1, 2)	AIC: 1219.2733739964442
Tested ARIMA (3, 1, 3)	AIC: 1221.5018505732637
Tested ARIMA (3, 1, 4)	AIC: 1221.2752464932728
Tested ARIMA (4, 1, 0)	AIC: 1222.4937446049198
Tested ARIMA (4, 1, 1)	AIC: 1221.1523997990612
Tested ARIMA (4, 1, 2)	AIC: 1219.4097515885999
Tested ARIMA (4, 1, 3)	AIC: 1221.4006625258799
Tested ARIMA (4, 1, 4)	AIC: 1222.5637097098502
Best order: (1, 1, 0)	**Best AIC: 1218.0701459600823**

Data availability

Data will be made available on request.

References

Abdul-Manan, A.F., Arfaj, A., Babiker, H., 2017. Oil refining in a CO2 constrained world: effects of carbon pricing on refineries globally. Energy 121, 264–275.

Alberola, E., Chevallier, J., Chèze, B., 2008. Price drivers and structural breaks in European carbon prices 2005–2007. Energy Pol. 36 (2), 787–797.

Alberola, E., Chevallier, J., Chèze, B., 2009. Emissions compliances and carbon prices under the EU ETS: a country specific analysis of industrial sectors. J. Pol. Model. 31 (3), 446–462.

An, Y., Zhou, D., Yu, J., Shi, X., Wang, Q., 2021. Carbon emission reduction characteristics for China's manufacturing firms: implications for formulating carbon policies. J. Environ. Manag. 284, 112055.

Ball, C., French, J., 2021. Exploring what stock markets tell us about GDP in theory and practice. Res. Econ. 75 (4), 330–344.

Benz, E., Trück, S., 2009. Modeling the price dynamics of CO2 emission allowances. Energy Econ. 31 (1), 4–15.

Bielinskyi, A.O., Matviychuk, A.V., Serdyuk, O.A., Semerikov, S.O., Solovieva, V.V., Soloviev, V.N., 2021. Correlational and non-extensive nature of carbon dioxide pricing market. In: International Conference on Information and Communication Technologies in Education, Research, and Industrial Applications. Springer International Publishing, Cham, pp. 183–199.

Brandt, P., Yesuf, G., Herold, M., Rufino, M.C., 2020. Intensification of dairy production can increase the GHG mitigation potential of the land use sector in East Africa. Global Change Biol. 26 (2), 568–585.

Bunn, D.W., Fezzi, C., 2007. Interaction of European Carbon Trading and Energy Prices.

Burgess, J.C., 1990. The contribution of efficient energy pricing to reducing carbon dioxide emissions. Energy Pol. 18 (5), 449–455.

Byun, S.J., Cho, H., 2013. Forecasting carbon futures volatility using GARCH models with energy volatilities. Energy Econ. 40, 207–221.

Çanakoğlu, E., Adıyeke, E., Ağralı, S., 2018. Modeling of carbon credit prices using regime switching approach. J. Renew. Sustain. Energy 10 (3).

Chen, P., Vivian, A., Ye, C., 2022. Forecasting carbon futures price: a hybrid method incorporating fuzzy entropy and extreme learning machine. Ann. Oper. Res. 313 (1), 559–601.

Chevallier, J., 2009. Carbon futures and macroeconomic risk factors: a view from the EU ETS. Energy Econ. 31 (4), 614–625.

Chevallier, J., 2011. A model of carbon price interactions with macroeconomic and energy dynamics. Energy Econ. 33 (6), 1295–1312.

Christiansen, A.C., Arvanitakis, A., Tangen, K., Hasselknippe, H., 2005. Price determinants in the EU emissions trading scheme. Clim. Pol. 5 (1), 15–30.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297.

Fan, X., Li, S., Tian, L., 2015. Chaotic characteristic identification for carbon price and an multi-layer perceptron network prediction model. Expert Syst. Appl. 42 (8), 3945–3952.

Gong, X., Li, M., Guan, K., Sun, C., 2023. Climate change attention and carbon futures return prediction. J. Futures Mark. 43 (9), 1261–1288.

Griffin, P.W., Hammond, G.P., Norman, J.B., 2016. Industrial energy use and carbon emissions reduction: a UK perspective. Wiley Interdisciplinary Reviews: Energy Environ. 5 (6), 684–714.

Hertwich, E.G., 2021. Increased carbon footprint of materials production driven by rise in investments. Nat. Geosci. 14 (3), 151–155.

Hoque, M.E., Bilgili, F., Batabyal, S., 2023. What do we know about spillover between the climate change futures market and the carbon futures market? Climatic Change 176 (12), 166.

Huang, Y., Dai, X., Wang, Q., Zhou, D., 2021. A hybrid model for carbon price forecasting using GARCH and long short-term memory network. Appl. Energy 285, 116485.

Huang, Y., He, Z., 2020. Carbon price forecasting with optimization prediction method based on unstructured combination. Sci. Total Environ. 725, 138350.

Ji, L., Zou, Y., He, K., Zhu, B., 2019. Carbon futures price forecasting based with ARIMA-CNN-LSTM model. Procedia Computer Science 162, 33–38.

Kim, P., Bae, H., 2022. Do firms respond differently to the carbon pricing by industrial sector? How and why? A comparison between manufacturing and electricity generation sectors using firm-level panel data in Korea. Energy Pol. 162, 112773.

Kontopoulou, V.I., Panagopoulos, A.D., Kakkos, I., Matsopoulos, G.K., 2023. A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. Future Internet 15 (8), 255.

Koop, G., Tole, L., 2013. Forecasting the European carbon market. J. Roy. Stat. Soc. Stat. Soc. 176 (3), 723–741.

Li, G., Ning, Z., Yang, H., Gao, L., 2022. A new carbon price prediction model. Energy 239, 122324.

Li, R., Wang, Q., Liu, Y., Jiang, R., 2021. Per-capita carbon emissions in 147 countries: the effect of economic, energy, social, and trade structural changes. Sustain. Prod. Consum. 27, 1149–1164.

Lin, B., Jia, Z., 2019. Impacts of carbon price level in carbon emission trading market. Appl. Energy 239, 157–170.

Liu, H., Pata, U.K., Zafar, M.W., Kartal, M.T., Karlilar, S., Caglar, A.E., 2023. Do oil and natural gas prices affect carbon efficiency? Daily evidence from China by wavelet transform-based approaches. Resour. Pol. 85, 104039.

Liu, Z., Huang, S., 2021. Carbon option price forecasting based on modified fractional Brownian motion optimized by GARCH model in carbon emission trading. N. Am. J. Econ. Finance 55, 101307.

Maiti, M., 2021. Risk analysis. In: Applied Financial Econometrics. Palgrave Macmillan, Singapore. https://doi.org/10.1007/978-981-16-4063-6_6.

Maiti, M., 2022a. Does development in venture capital investments influence green growth? Technol. Forecast. Soc. Change 182, 121878.

Maiti, M., 2022b. Does improvement in green growth influence the development of environmental related technology? Innovation and Green Development 1 (2), 100008.

Maiti, M., Jadhav, P., 2021. Impact of pollution level, death rate and illness on economic growth: evidence from the global economy. SN Business & Economics 1 (9), 109.

Maiti, M., Kayal, P., 2022. Asymmetric information flow between exchange rate, oil, and gold: new evidence from transfer entropy approach. J. Risk Financ. Manag. 16 (1), 2.

Mendelsohn, R., Olmstead, S., 2009. The economic valuation of environmental amenities and disamenities: methods and applications. Annu. Rev. Environ. Resour. 34, 325–347.

Murali, M., Kayal, P., Maiti, M., 2023. Should you invest in the companies that promote the Circular Economy idea? Manag. Environ. Qual. Int. J.

Ning, Y., Kazemi, H., Tahmasebi, P., 2022. A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet. Comput. Geosci. 164, 105126.

Niu, H., Liu, T., 2024. Forecasting the volatility of European Union allowance futures with macroeconomic variables using the GJR-GARCH-MIDAS model. Empir. Econ. 1–22.

Paolella, M.S., Taschini, L., 2008. An econometric analysis of emission allowance prices. J. Bank. Finance 32 (10), 2022–2032.

Parker, S., Bhatti, M.I., 2020. Dynamics and drivers of per capita CO2 emissions in Asia. Energy Econ. 89, 104798.

Pizzol, M., Weidema, B., Brandão, M., Osset, P., 2015. Monetary valuation in life cycle assessment: a review. J. Clean. Prod. 86, 170–179.

Qin, Q., He, H., Li, L., He, L.Y., 2020. A novel decomposition-ensemble based carbon price forecasting model integrated with local polynomial prediction. Comput. Econ. 55, 1249–1273.

Ranjan, S., Kayal, P., Saraf, M., 2023. Bitcoin price prediction: a machine learning sample dimension approach. Comput. Econ. 61 (4), 1617–1636.

Reboredo, J.C., 2014. Volatility spillovers between the oil market and the European Union carbon emission market. Econ. Modell. 36, 229–234.

Sanin, M.E., Violante, F., Mansanet-Bataller, M., 2015. Understanding volatility dynamics in the EU-ETS market. Energy Pol. 82, 321–331.

Seo, S.N., 2017. Beyond the Paris Agreement: climate change policy negotiations and future directions. Regional Science Policy & Practice 9 (2), 121–141.

Shahzad, U., Sengupta, T., Rao, A., Cui, L., 2023. Forecasting carbon emissions future prices using the machine learning methods. Ann. Oper. Res. 1.

Siami-Namini, S., Tavakoli, N., Namin, A.S., 2018. A comparison of ARIMA and LSTM in forecasting time series. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, pp. 1394–1401.

Stegmann, P., Daioglou, V., Londo, M., et al., 2022. Plastic futures and their CO2 emissions. Nature 612, 272–276. https://doi.org/10.1038/s41586-022-05422-5.

Subramaniam, N., Wahyuni, D., Cooper, B.J., Leung, P., Wines, G., 2015. Integration of carbon risks and opportunities in enterprise risk management systems: evidence from Australian firms. J. Clean. Prod. 96, 407–417.

Vaidynathan, D., Kayal, P., Maiti, M., 2023. Effects of economic factors on median list and selling prices in the US housing market. Data Science and Management 6 (4), 199–207.

Van Ruijven, B., Van Vuuren, D.P., 2009. Oil and natural gas prices and greenhouse gas emission mitigation. Energy Pol. 37 (11), 4797–4808.

Wang, Z., Dong, Z., 2023. How does the time-varying network structure evolve between the EU carbon futures prices and industrial and energy-related indices? A study based on a time-varying T-copula. Int. J. Energy Res. 2023 (1), 6696059.

World Economic Forum, 2021. Global Chemical Companies Collaborate in Pivotal Move to Net-Zero. https://www.weforum.org/press/2021/10/global-chemical-companies-collaborate-in-pivotal-move-to-net-zero/.

Yadav, V.G., Yadav, G.D., Patankar, S.C., 2020. The production of fuels and chemicals in the new world: critical analysis of the choice between crude oil and biomass vis-à-vis sustainability and the environment. Clean Technol. Environ. Policy 22, 1757–1774.

Yahşi, M., Çanakoğlu, E., Ağralı, S., 2019. Carbon price forecasting models based on big data analytics. Carbon Manag. 10 (2), 175–187.

Yuan, R., Zhao, T., 2016. Changes in CO2 emissions from China's energy-intensive industries: a subsystem input–output decomposition analysis. J. Clean. Prod. 117, 98–109.

Zhang, N., Lin, A., Shang, P., 2017. Multidimensional k-nearest neighbor model based on EEMD for financial time series forecasting. Phys. Stat. Mech. Appl. 477, 161–173.

Zhang, X., Yang, K., Lu, Q., Wu, J., Yu, L., Lin, Y., 2023b. Predicting carbon futures prices based on a new hybrid machine learning: comparative study of carbon prices in different periods. J. Environ. Manag. 346, 118962.

Zhang, K., Yang, X., Wang, T., Thé, J., Tan, Z., Yu, H., 2023a. Multi-step carbon price forecasting using a hybrid model based on multivariate decomposition strategy and deep learning algorithms. J. Clean. Prod. 405, 136959.

Zhao, Y., Huang, Y., Wang, Z., Liu, X., 2024. Carbon futures price forecasting based on feature selection. Eng. Appl. Artif. Intell. 135, 108646.

Zhu, B., Chevallier, J., Zhu, B., Chevallier, J., 2017. Carbon price forecasting with a hybrid Arima and least squares support vector machines methodology. Pricing and forecasting carbon markets: Models and empirical analyses 87–107.

Zhu, B., Wang, P., Chevallier, J., Wei, Y., 2015. Carbon price analysis using empirical mode decomposition. Comput. Econ. 45, 195–206.

Zhu, B., Ye, S., Wang, P., Chevallier, J., Wei, Y.M., 2022. Forecasting carbon price using a multi-objective least squares support vector machine with mixture kernels. J. Forecast. 41 (1), 100–117.

Zien, A., Krämer, N., Sonnenburg, S., Rätsch, G., 2009. The feature importance ranking measure. Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2009, Bled, Slovenia, September 7-11, 2009, Proceedings, Part II 20, 694–709. Springer Berlin Heidelberg.

Zou, X., 2018. An analysis of the effect of carbon emission, GDP and international crude oil prices based on synthesis integration model. Int. J. Energy Sect. Manag. 12 (4), 641–655.