Forecasting Carbon Prices Using Machine Learning: A Comparative Study of ARIMA, Random Forest, and LSTM Models

Erica Zhao

Preprint submitted to Elsevier *2024 Spring*

1. Introduction

The European Union Emissions Trading System (EU ETS) is a climate policy instrument designed to reduce greenhouse gas (GHG) emissions in a cost-effective manner. It was launched in 2005, and is the world's first and one of the largest carbon markets globally. The EU ETS requires polluters to pay for their GHG emissions, thereby creating a financial incentive for companies to cut emissions and invest in cleaner technologies. Covering emissions from electricity and heat generation, industrial manufacturing, aviation, and recently maritime transport, the EU ETS accounts for roughly 40% of the EU's total GHG emissions [3]. As the EU ETS has evolved, it has undergone four trading phases, each introducing significant reforms aimed at improving its efficiency and effectiveness in reducing emissions [12].

The EU ETS operates under a "cap-and-trade" principle, which involves setting a cap on the total amount of GHG emissions allowed from the sectors covered by the system. This cap is reduced annually to align with the EU's climate targets, ensuring that emissions decrease over time. Emission allowances, which permit companies to emit a certain amount of GHG, are either auctioned or allocated for free, depending on the industry. Companies that emit less than their allowance can trade the excess, while those that exceed their allocation must purchase additional allowances or face heavy fines [3]. This market-based approach not only supports emission reductions but also generates revenues to finance the green transition across the EU [4].

As the EU strives to achieve climate neutrality by 2050, understanding and accurately predicting carbon prices is vital for both policymakers and market participants. The EU ETS cap-and-trade mechanism makes carbon prices inherently volatile and influenced by various factors, such as regulatory changes, economic growth, and energy demand. Forecasting these prices helps stakeholders—including governments, businesses, and investors—make informed decisions about emissions management, risk mitigation, and strategic planning [14]. Reliable predictive models can enhance market stability by allowing for better anticipation of price movements, thus helping companies adapt their strategies to comply with emission regulations cost-effectively.

Moreover, the EU ETS has served as a blueprint for other emissions trading systems worldwide, making research into its market dynamics globally relevant [2]. Understanding how carbon pricing functions in the EU market can provide insights into designing or improving similar systems in other regions. By developing and comparing models such as ARIMA, Random Forest (RF), and Long Short-Term Memory (LSTM) networks, this research aims to identify which methodologies provide the most reliable predictions under varying market conditions. This comparative analysis not only contributes to the academic literature on carbon price forecasting but also offers practical insights that can enhance the efficacy of carbon trading systems,

supporting the overall goal of reducing global GHG emissions in a cost-effective manner.

2. Literature Review

The literature on carbon price forecasting has evolved significantly over the past two decades, reflecting the need for more accurate predictive models to cope with the growing complexity of carbon markets. Traditional econometric models, such as ARIMA, have been among the first tools used to forecast carbon prices due to their simplicity and statistical rigor. Zhu et al.[13] highlighted ARIMA's effectiveness in modeling linear components of carbon price time series, especially during the early phases of the EU ETS, when the carbon market exhibited relatively stable and predictable trends. However, ARIMA's limitations in capturing the nonlinear and volatile nature of carbon prices, particularly in the post-2017 EU ETS environment, have prompted researchers to explore more advanced methods [5, 8].

Machine learning models have become increasingly popular due to their ability to manage nonlinear relationships and complex patterns in data. Studies have shown that Random Forest (RF) and other machine learning models offer significant improvements in predictive accuracy by handling complex interactions among input variables [10]. RF, for instance, has demonstrated excellent predictive performance, particularly when integrated into hybrid frameworks with other machine learning models. Xu et al. [9] showed that RF consistently outperformed traditional econometric models such as ARMA in predicting carbon returns in China's carbon trading market, suggesting its advantage in managing the complexities of carbon price dynamics. Furthermore, Lu et al. [7] explored multiple machine learning approaches, including extreme gradient boosting and RF, to predict carbon trading volume and price in various Chinese carbon markets, emphasizing the adaptability of RF to different market characteristics.

Deep learning approaches, particularly LSTM networks, have also been extensively utilized due to their strength in modeling sequential data and learning long-term dependencies. Huang et al. [5] used a hybrid GARCH-LSTM model to address volatility in carbon prices under the EU ETS following market reforms in 2017. The hybrid model was capable of outperforming both standalone econometric and machine learning models, with significantly smaller forecasting errors. Zhang et al. [11] introduced a novel hybrid deep learning model, VMD-CNN-BILSTM-MLP, which leveraged variational mode decomposition to effectively handle noise in carbon price data, demonstrating superior performance compared to single models. Additionally, Li et al. [6] used Multivariate LSTM, taking into account multiple influencing fac-

tors, to improve the prediction of carbon prices in China, demonstrating that LSTM outperformed other deep learning models in terms of MAE, MSE, and RMSE metrics.

The importance of combining econometric models and machine learning techniques is emphasized by several hybrid approaches. For instance, Zhu et al.[14] proposed an empirical mode decomposition-based evolutionary least squares support vector regression (EMD-LSSVR) model, which leveraged both linear and nonlinear feature extraction for multiscale forecasting of carbon prices. Their results demonstrated improved statistical accuracy compared to ARIMA and standalone LSSVR models, highlighting the value of incorporating multiscale features for enhancing model robustness. Similarly, Yahcsi et al. [10] noted that the hybrid approach of RF and decision trees was particularly successful in managing the high dimensionality and nonlinear aspects of carbon pricing data.

Past literature has also pointed out the importance of selecting appropriate predictors in carbon price forecasting. For example, Abdi et al. [1] used Bayesian Networks to forecast carbon prices in the Western Climate Initiative market, demonstrating that energy prices and economic growth were the most influential drivers. The incorporation of external influencing factors such as energy markets, macroeconomic stability, and policy changes has been shown to enhance the predictive power of machine learning and deep learning models significantly [9, 11].

The contributions of this study are twofold. Firstly, it offers a rigorous comparative analysis of ARIMA, Random Forest (RF), and Long Short-Term Memory (LSTM) models, thereby enhancing the understanding of the performance of different forecasting methodologies within the complex dynamics of the carbon trading market. Secondly, this research elucidates the relative merits of traditional time series models in contrast to advanced machine learning approaches, providing valuable insights for both investors and policymakers. Specifically, the study aims to demonstrate that while ARIMA models may be suitable for relatively stable market conditions, machine learning and hybrid approaches provide superior predictive accuracy in the context of nonlinearities and heightened market volatility. This comparative evaluation serves to determine the optimal models for varying market scenarios and provides a basis for employing these models to effectively manage the uncertainties inherent in carbon price forecasting.

Past literature on carbon price forecasting has made significant advances by utilizing traditional statistical models, machine learning techniques, and hybrid approaches. However, much of the existing research has focused either on individual modeling techniques or lacked a direct, systematic comparison among various models under the same conditions. As a result, there has been limited insight into the specific contexts in which each approach might perform best. Furthermore, while hy-

brid models have shown improved accuracy, the question of which standalone model or hybrid combination is most suitable for specific market conditions has remained largely unexplored.

This study addresses these gaps by providing a comprehensive evaluation of three distinct types of models—ARIMA, RF, and LSTM—under identical market conditions. Unlike previous studies that predominantly concentrated on demonstrating the superiority of a single model or a specific hybrid framework, this research systematically contrasts the performance of traditional time series models, machine learning algorithms, and deep learning networks to offer a clear understanding of their relative strengths and weaknesses. By doing so, it clarifies the scenarios where simpler statistical models might suffice versus those where advanced machine learning models are essential for achieving high accuracy, thereby filling the gap in understanding the contextual effectiveness of different forecasting methods in the carbon trading market.

3. Methodology

3.1. ARIMA Model

ARIMA model has notable strengths for forecasting carbon prices, utilizing its robust capability to process time series data and capture the dynamic patterns in the volatile carbon prices. These features facilitate a comprehensive analysis, making ARIMA an ideal choice for short-term forecasting. However, we also need to acknowledge its potential limitations, such as the assumption of linearity and the need for data stationarity, which may cause potential issues.

Central to the ARIMA model is its ability to integrate autoregressive (AR) processes, differencing (I) to achieve stationarity, and moving average (MA) processes into a unified structure. The model is parameterized by three integers: p, d, and q, which denote the autoregressive order, the degree of differencing, and the moving average order, respectively. Mathematically, the ARIMA(p,d,q) model is expressed as:

$$y_t = u + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \ldots - \beta_q \varepsilon_{t-q} + \varepsilon_t \quad (1)$$

Here, y_t represents the forcasted carbon price at time t, and u is a constant term (which can be interpreted as the mean of the series if $d = 0$). The parameters $\alpha_1, \alpha_2, \ldots, \alpha_p$ represent the autoregressive (AR) terms, $\beta_1, \beta_2, \ldots, \beta_q$ represent the moving average (MA) terms, and $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{t-q}$ are the lagged forecast errors in the prediction equation.

Our methodology begins with a preliminary analysis to establish the presence of trends and autocorrelations in the carbon price data, guiding the selection of the ARIMA model's parameters. Diagnostic tools such as the autocorrelation function (ACF) and the partial autocorrelation function (PACF) assist in determining the optimal d value and refining the choices for p and q. This process ensures that the model adequately reflects the data's underlying patterns without overfitting. Parameter estimation is carried out using Maimum Likelihood Estimation (MLE), providing a robust framework for fitting the ARIMA model to the observed carbon prices. Model validation is then conducted by examining the residuals to confirm that they resemble white noise, indicating that the model captures all significant information in the data. The final phase involves applying the validated ARIMA model to forecast future carbon prices.

3.2. Random Forest

In this study, the Random Forest (RF) methodology was employed as the second ensemble-based predictive model for forecasting carbon prices. Random Forest is a form of the ensemble method known as Bagging, or Bootstrap Aggregating, which operates on the principle of creating multiple datasets from the original training set through bootstrap sampling. This sampling technique involves random selection with replacement, resulting in subsets where individual instances may repeat or be excluded.

Renowned for its robustness, the Random Forest is inherently adept at handling large, high-dimensional datasets and mitigating overfitting—a common pitfall in complex models. The ensemble nature of Random Forest, where multiple decision trees vote on the outcome, reduces variance and enhances performance stability. Nonetheless, it may incur considerable computational expense, particularly with a large number of trees, and can be less interpretable due to its ensemble complexity.

The RF algorithm constructs a multitude of decision trees $(T_1, T_2, \ldots, T_N)$ during the training process. For regression tasks, the RF employs bootstrap aggregating, where each tree T_i is trained on a bootstrap sample D_i, drawn with replacement from the complete training dataset D. During tree construction, a random subset of m features is selected at each node for splitting.
Given a new input x, each tree T_i yields a prediction y_i. The RF's final prediction $\hat{y}$ for input x is the average of these predictions:

$$\hat{y}(x) = \frac{1}{N} \sum_{i=1}^{N} y_i(x) \tag{2}$$

Within each tree, the learning objective is to minimize the prediction error, quantified by the mean squared error (MSE), following the least squares approach:

$$\mathrm{Err}(t) = \frac{1}{n_t} \sum_{D_t} (y_i - k_t)^2 \tag{3}$$

where n_t represents the number of instances in node t, and k_t is the average of instances in each node.

To implement the Random Forest model effectively in this study, we optimized several hyperparameters, including the number of trees (n_estimators), the maximum depth of each tree, and the number of features considered at each split. Each decision tree was meticulously trained on its respective bootstrap sample, and the collective output during the prediction phase was derived from averaging the results of these individual trees, providing a robust prediction of future carbon prices.

3.3. LSTM Model

Long Short-Term Memory (LSTM) networks, the third model integrated into our analysis, represent an evolution of recurrent neural networks (RNNs) specifically engineered to predict carbon prices. LSTMs excel in capturing and leveraging long-term dependencies within time-series data, a capability crucial for the accurate forecasting of dynamic economic indicators such as carbon prices. Their architecture is intricately designed to address the vanishing gradient problem—a significant challenge in traditional RNNs—thereby enhancing their learning capacity and predictive performance.

LSTM units include memory cells that preserve states over time, enabling sustained information retention throughout the sequence. The mathematical operations of an LSTM cell are governed by the following equations:

$$f_t = \sigma(W_f \cdot [h_{t-1}, y_t] + b_f) \tag{4}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, y_t] + b_i) \tag{5}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, y_t] + b_o) \tag{6}$$

$$c_t = \tanh(W_c \cdot [h_{t-1}, y_t] + b_c) \tag{7}$$

$$h_t = o_t * \tanh(c_t) \tag{8}$$

The forget gate f_t, input gate i_t, and output gate o_t within each cell dictate the preservation and incorporation of information, ensuring the model's adaptability to the intricacies of the data.

In our study, the LSTM model was meticulously configured and trained on historical carbon price data to harness its potential for deep temporal learning. The LSTM architecture was tailored to capture the complexity and nuances of the dataset, with hyperparameters optimized based on validation performance metrics. Training employed backpropagation through time, a technique allowing the model to refine its weights based on the temporal sequence of data, thereby enabling it to capture underlying temporal dynamics effectively. Following extensive training, the LSTM model was deployed to forecast future trends in carbon prices, leveraging its sophisticated ability to interpret and project observed historical patterns into future insights.

4. Data and Preprocessing

4.1. Descriptive Statistics

In the analysis of carbon pricing trends within the European Union's Emissions Trading System (EU ETS), this study draws upon a meticulously compiled dataset encompassing the period from January 1, 2021, to January 31, 2024. During this time frame, a total of 763 data points were collected, providing a granular view of carbon price movements. The carbon price data, sourced from Investing.com, is critical for understanding the economic implications of regulatory shifts and market dynamics under the EU's climate policy framework.

This timeframe is deliberately aligned with the initiation of Phase 4 of the European Union's Emissions Trading System (EU ETS), marking a significant transition from the system's previous phases. The focus on this specific period is due to the substantial regulatory shifts introduced in Phase 4, which differ markedly from the earlier phases in terms of emissions reduction targets and allowance allocation mechanisms.

Notably, in Phase 3 of the EU ETS (2013-2020), the Union-wide cap for stationary installations was subject to an annual decrease by a linear reduction factor of 1.74%, with the 2013 cap determined based on the average total quantity of allowances issued annually during 2008-2012. Conversely, Phase 4 (2021-2030) introduces a more aggressive approach to emissions reduction, with the cap on emissions decreasing annually at an increased linear reduction factor of 2.2%. This heightened ambition reflects the EU's commitment to achieving more substantial emissions reductions and necessitates an in-depth analysis of carbon pricing trends under the new regulatory framework. The study aims to isolate the impacts of these changes on carbon pricing, underscoring the importance of understanding the dynamics introduced by Phase 4's enhanced regulatory measures.

The descriptive statistics for the dataset, which includes carbon pricing trends within the European Union's Emissions Trading System (EU ETS) from January 1,

2021, to January 31, 2024, are presented in Table 1. These statistics provide insights into the behavior of the carbon market during Phase 4 of the EU ETS. The carbon price exhibits a mean of €72.98 and a standard deviation of €16.45, reflecting market volatility over the 763 observations in the dataset.

Table 1: Descriptive Statistics of the Dataset

Variable	Mean	Standard Deviation	Maximum	Minimum
Carbon	72.98	16.45	98.01	31.96
Coal	176.47	93.46	439.00	63.75
Crude Oil	83.82	14.26	127.98	51.09
Natural Gas	72.45	54.61	339.20	15.53
Fit_for_55	0.83	0.38	1	0
Law	0.22	0.41	1	0
Compromise	0.25	0.43	1	0
EPU	276.88	69.80	418.62	141.39
DAX	14980.43	1082.54	16972.34	11975.55
SPGTCLNT	1699.21	295.68	2720.79	1088.34
FTMIB	25692.05	2414.81	30744.24	20352.98
CAC40	6722.63	526.14	7677.47	5399.21
AEX	732.69	44.97	827.57	630.58
STOXX	448.01	22.92	494.35	382.89
Temperature	12.94	6.64	31.1	-3.2
Precipitation	1.53	4.05	36.03	0.0
Solar Energy	6.75	5.88	27.1	0.1
Wind Speed	11.39	3.82	30.7	3.8

The time series plot of the carbon price, depicted in Figure 1, emphasizes the fluctuations over the observed period. Furthermore, the study will explore the patterns in related energy commodities, policy indicators, European stock indices, and weather variables, all of which could influence emission levels and subsequently, carbon market dynamics.

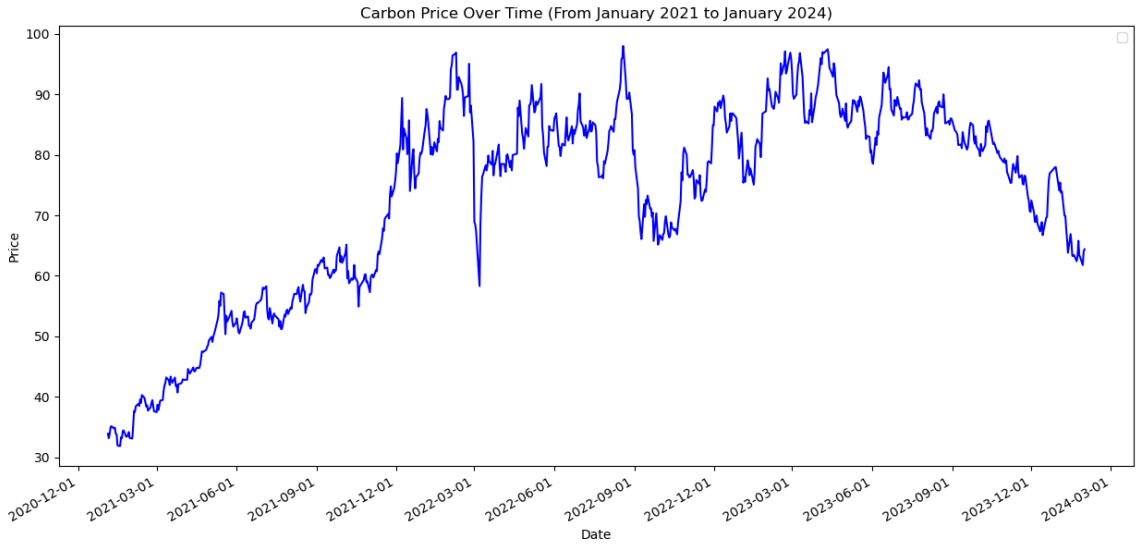

Figure 1: Time Series Plot of Carbon Price from January 2021 to January 2024.

4.2. Preprocessing

For the ARIMA model, addressing multicollinearity was a priority due to the nature of its statistical assumptions. Principal Component Analysis (PCA) was applied to two groups of features—stock indices (DAX, STOXX, FTMIB, CAC40, AEX, and SPGTCLNT) and energy prices (Crude Oil, Coal, and Natural Gas)—to reduce dimensionality and mitigate multicollinearity. Before applying PCA, these variables were standardized using a StandardScaler, ensuring that all features had a mean of zero and unit variance, thus preventing any single feature from dominating the principal components.

PCA resulted in two new components: PCA_{Stock_Index} and PCA_{Energy_Prices}, each capturing the majority of the variance within their respective feature sets. These components were subsequently included in the ARIMA model as inputs, thereby simplifying the feature set while retaining the relevant information from the original correlated variables. To ensure that the data was suitable for time series modeling, stationarity tests were conducted using the Augmented Dickey-Fuller (ADF) test. For any non-stationary variables, differencing was applied until the data achieved stationarity.

5. Performance and Results

5.1. Evaluation Criteria

Evaluating the accuracy and reliability of our predictive model is crucial for ensuring its practical applicability in forecasting carbon prices. In this study, we have utilized several statistical measures to assess model performance. The primary metric we considered is the Root Mean Squared Error (RMSE), which provides an estimate of the model's prediction error magnitude, offering an objective performance measure less sensitive to outliers. The Mean Absolute Percentage Error (MAPE) is used as a supplementary metric to understand the relative size of the forecast errors in percentage terms. These metrics are formally defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \tag{9}$$

$$\text{MAPE} = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \tag{10}$$

where y_t represents the actual observed value, $\hat{y}_t$ denotes the forecasted value, and n is the number of observations.

The RMSE serves as the principal measure due to its ability to capture the absolute errors in the model's predictions. Conversely, the MAPE provides additional context by illustrating the error magnitude relative to the actual values, which is particularly insightful when considering the proportional accuracy of the model across different price levels.

In addition to RMSE and MAPE, we also consider the Mean Absolute Error (MAE) and the Mean Square Error (MSE), defined as follows:

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \tag{11}$$

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2 \tag{12}$$

The MAE captures the average absolute difference between the observed and predicted values, while the MSE emphasizes larger errors due to its quadratic nature, giving more weight to significant deviations.

Finally, the model's goodness of fit is evaluated using the Coefficient of Determination, or R^2, which measures the proportion of variance in the dependent variable that is predictable from the independent variables. This metric is defined as:

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \tag{13}$$

where $\bar{y}$ is the mean of the actual values y_t. The R^2 value ranges between 0 and 1, with values closer to 1 indicating that the model explains a higher proportion of the variance in the data. In this study, the R^2 is calculated using the actual observed carbon prices and the predicted values, providing insight into how well the model fits the overall dataset.

Taken together, these evaluation metrics—RMSE, MAPE, MAE, MSE, and R^2—provide a comprehensive framework for assessing both the absolute and relative accuracy of the ARIMA model in forecasting carbon prices. The combined analysis offers a robust understanding of the model's performance across different dimensions of error and variability.

5.2. Model Performances

5.2.1. ARIMA Model

The ARIMA(1,1,1) model augmented with exogenous variables was employed to forecast carbon prices, utilizing PCA-transformed stock indices and energy prices as external regressors. To determine the appropriate order of the ARIMA model, a grid search was performed, exploring various combinations of autoregressive (p), differencing (d), and moving average (q) terms. The selection criterion was based on minimizing the Akaike Information Criterion (AIC), with ARIMA(1,1,1) emerging as the optimal configuration.

The model was trained using a rolling forecasting strategy, wherein it was updated daily with newly observed data from the test set. This approach ensured the model remained adaptive to real-time market fluctuations, enhancing its predictive accuracy. The dataset was split into an 80% training set and a 20% testing set for in-sample and out-of-sample evaluations, respectively.

Model performance was assessed through a comprehensive set of metrics. The results of the in-sample and out-of-sample performance evaluations are presented in Table 2.

Table 2: ARIMA(1,1,1) Model Performance Comparison

Metric	In-Sample	Out-of-Sample
RMSE	2.07	1.46
MAPE	2.08%	1.53%
MAE	1.50	1.19
MSE	4.28	2.13
R^2	0.986	0.967

The in-sample predictions exhibit a close alignment with the actual data, underscoring the model's strong fit to the training set. Similarly, the out-of-sample results demonstrate the model's ability to generalize effectively to new, unseen data. The rolling forecast strategy, in particular, contributed to maintaining a low error margin, even under more volatile market conditions. The results of the White test for heteroskedasticity returned a p-value of 0.17, indicating no significant presence of heteroskedasticity in the residuals, further confirming the adequacy of the model's fit. The figures below compare the actual and predicted carbon prices for both the in-sample and out-of-sample periods. As shown in Figures 2 and 3, the ARIMA model's predictions closely tracked the actual price movements, with only minor deviations in the out-of-sample predictions, which are typical for time series models. These deviations remained within acceptable error margins and did not significantly impact the model's overall performance.

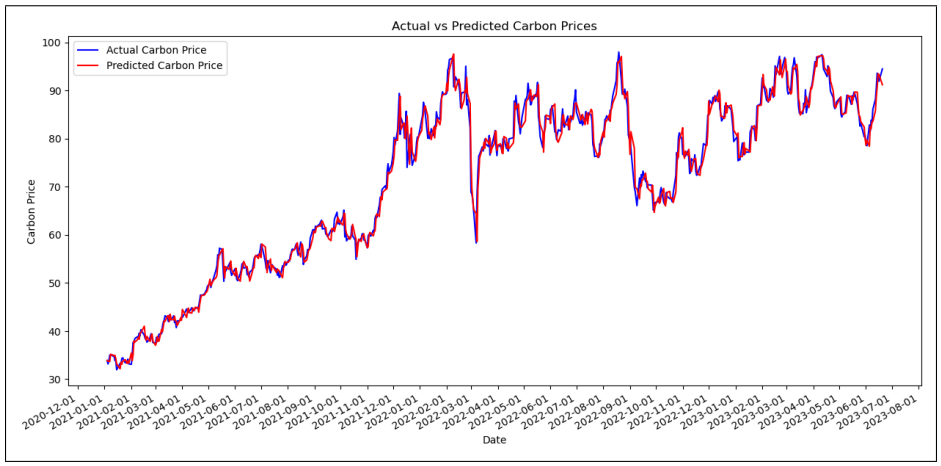

Figure 2: In-sample Actual vs Predicted Carbon Prices using ARIMA(1,1,1).

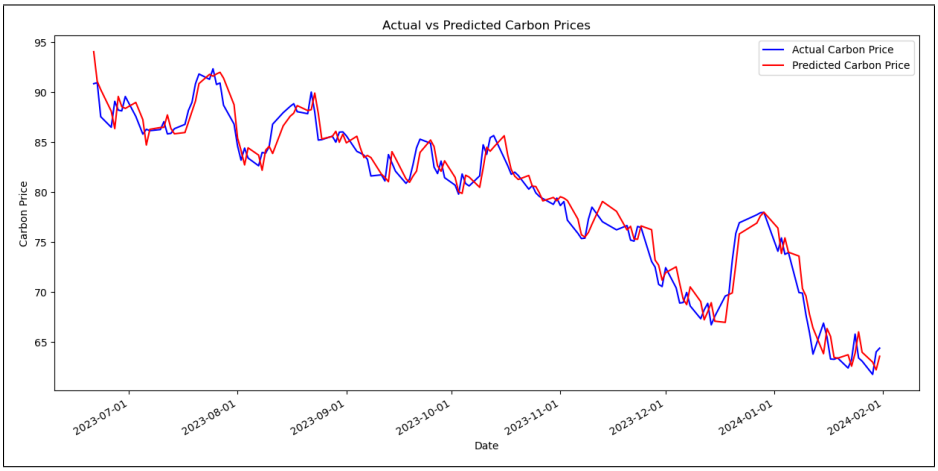

Figure 3: Out-of-sample Actual vs Predicted Carbon Prices using ARIMA(1,1,1).

The performance metrics indicate that the ARIMA(1,1,1) model, augmented with external variables, provides a robust and reliable framework for carbon price forecasting. The relatively low RMSE, MAPE, and MAE values, coupled with high R^2 scores across both in-sample and out-of-sample evaluations, suggest that the model captures the underlying market dynamics with considerable accuracy.

5.2.2. Random Forest Model

The Random Forest model was employed to forecast carbon prices, with two variations: one using the original undifferenced carbon price data and another using a differenced version of the series to ensure stationarity. Differencing was applied to the time series data to transform it into a stationary process, which is crucial for improving the model's performance when dealing with non-stationary data like carbon prices. The differencing technique involves subtracting the previous value of the time series from the current value, as shown in the following equation:

$$y'_t = y_t - y_{t-1} \tag{14}$$

where y'_t represents the differenced carbon price at time t, and y_t and y_{t-1} are the current and previous carbon prices, respectively. This operation eliminates trends and other non-stationary components, making the series more suitable for modeling with machine learning algorithms such as Random Forest.

To ensure robustness, we performed hyperparameter tuning using a grid search across several configurations, including `n_estimators`, `‘max_depth‘`, `‘min_samples_split‘`, `‘min_samples_leaf‘`, and `‘max_features‘`. The tuning process utilized time-series

cross-validation to capture the temporal structure of the data and minimize overfitting. The best hyperparameters for the model were selected based on minimizing the Mean Squared Error (MSE).

The Random Forest model was trained on 80% of the dataset, with the remaining 20% reserved for out-of-sample validation. For both the undifferenced and differenced datasets, model performance was evaluated using several statistical measures. The results for the undifferenced data are summarized in Table 3, while the differenced data results are presented in Table 4.

Table 3: Performance Summary for Random Forest (Without Differencing)

Measure	**In-Sample**	**Out-of-Sample**
RMSE	1.69	3.33
MAPE	1.58%	3.60%
MAE	1.17	2.72
MSE	2.85	11.11
R^2	0.991	0.826

Table 4: Performance Summary for Random Forest (With Differencing)

Measure	**In-Sample**	**Out-of-Sample**
RMSE	2.22	1.54
MAPE	2.17%	1.55%
MAE	1.56	1.21
MSE	4.93	2.37
R^2	0.984	0.963

The model trained without differencing performed well in the in-sample evaluation, achieving an R^2 value of 0.991, indicating that it captured the majority of the variance in the training data. However, when tested out-of-sample, the model's performance decreased, with a higher RMSE of 3.33 and an R^2 of 0.826, suggesting some overfitting. The error metrics indicate that while the undifferenced model could explain most of the variance in the training data, its generalization ability was limited in out-of-sample forecasts.

On the other hand, the differenced Random Forest model demonstrated strong generalization, with an out-of-sample RMSE of 1.54 and an R^2 of 0.963, reflecting a better ability to forecast carbon prices in previously unseen data. The in-sample performance remained robust, with an RMSE of 2.22 and an R^2 of 0.984. The relatively balanced performance across both in-sample and out-of-sample tests suggests

that differencing helped stabilize the model and prevent overfitting to the training data.

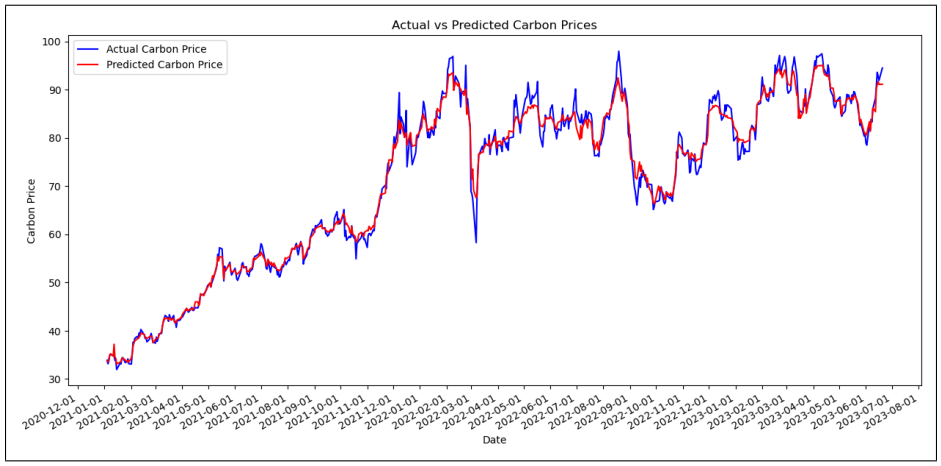

Figure 4: In-sample Actual vs Predicted Carbon Prices using Random Forest (Without Differencing).

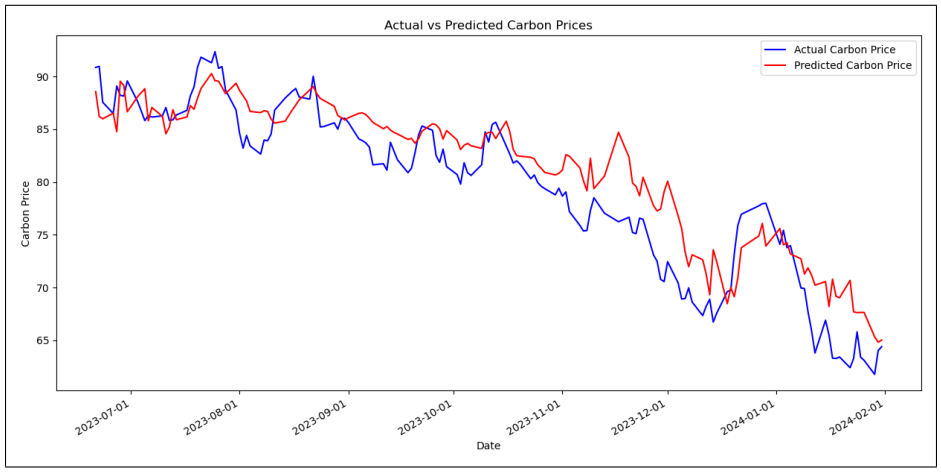

Figure 5: Out-of-sample Actual vs Predicted Carbon Prices using Random Forest (Without Differencing).

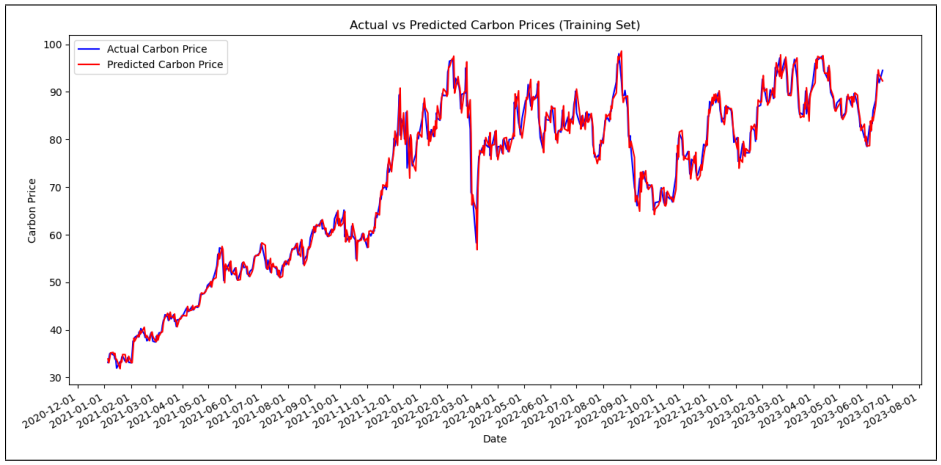

Figure 6: In-sample Actual vs Predicted Carbon Prices using Random Forest (With Differencing).

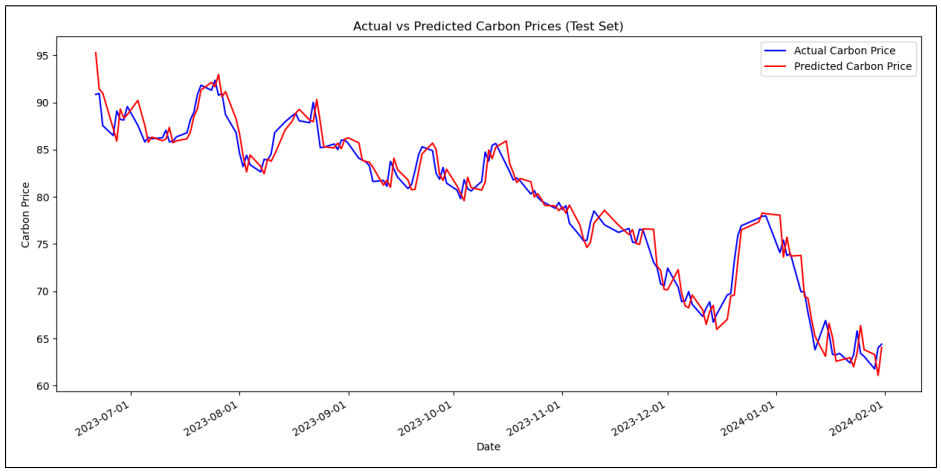

Figure 7: Out-of-sample Actual vs Predicted Carbon Prices using Random Forest (With Differencing).

Overall, the differenced model outperformed the undifferenced model in terms of generalization and error reduction in the test set. The findings indicate that the Random Forest model is more effective when applied to stationary data, as evidenced by the improved accuracy in both in-sample and out-of-sample evaluations. These results underline the importance of handling non-stationarity when forecasting volatile markets such as carbon prices, where structural changes and trends can significantly affect the model's predictions.

5.2.3. Long Short Term Memory Model

5.3. Discussion

Discuss the results, implications, and any observed trends or anomalies.

6. Conclusion

Summarize the findings, their implications for

References

[1] Abdollah Abdi and Sharareh Taghipour. Forecasting carbon price in the western climate initiative market using bayesian networks. *Carbon Management*, 10(3): 255–268, 2019.

[2] Thomas Aspinall, Adrian Gepp, Geoff Harris, Simone Kelly, Colette Southam, and Bruce Vanstone. Estimation of a term structure model of carbon prices through state space methods: The european union emissions trading scheme. *Accounting & Finance*, 61(2):3797–3819, 2021.

[3] European Commission. Climate action. `https://climate.ec.europa.eu`. Accessed: 2024-10-09.

[4] John Hua Fan and Neda Todorova. Dynamics of china's carbon prices in the pilot trading phase. *Applied Energy*, 208:1452–1467, 2017.

[5] Yumeng Huang, Xingyu Dai, Qunwei Wang, and Dequn Zhou. A hybrid model for carbon price forecasting using garch and long short-term memory network. *Applied Energy*, 285:116485, 2021.

[6] Houjian Li, Xinya Huang, Deheng Zhou, Andi Cao, Mengying Su, Yufeng Wang, and Lili Guo. Forecasting carbon price in china: A multimodel comparison. *International Journal of Environmental Research and Public Health*, 19(10):6217, 2022.

[7] Hongfang Lu, Xin Ma, Kun Huang, and Mohammadamin Azimi. Carbon trading volume and price forecasting in china using multiple machine learning models. *Journal of Cleaner Production*, 249:119386, 2020.

[8] Jujie Wang, Xin Sun, Qian Cheng, and Quan Cui. An innovative random forest-based nonlinear ensemble paradigm of improved feature extraction and deep learning for carbon price forecasting. *Science of the Total Environment*, 762:143099, 2021.

[9] Yingying Xu, Yifan Dai, Lingling Guo, and Jingjing Chen. Leveraging machine learning to forecast carbon returns: Factors from energy markets. *Applied Energy*, 357:122515, 2024.

[10] Mustafa Yahşi, Ethem Çanakoğlu, and Semra Ağralı. Carbon price forecasting models based on big data analytics. *Carbon Management*, 10(2):175–187, 2019.

[11] Xi Zhang, Kailing Yang, Qin Lu, Jingyu Wu, Liang Yu, and Yu Lin. Predicting carbon futures prices based on a new hybrid machine learning: Comparative study of carbon prices in different periods. *Journal of Environmental Management*, 346:118962, 2023.

[12] Xin Zhao, Meng Han, Lili Ding, and Wanglin Kang. Usefulness of economic and energy data at different frequencies for carbon price forecasting in the eu ets. *Applied Energy*, 216:132–141, 2018.

[13] Bangzhu Zhu, Julien Chevallier, Bangzhu Zhu, and Julien Chevallier. Carbon price forecasting with a hybrid arima and least squares support vector machines methodology. *Pricing and forecasting carbon markets: Models and empirical analyses*, pages 87–107, 2017.

[14] Bangzhu Zhu, Dong Han, Ping Wang, Zhanchi Wu, Tao Zhang, and Yi-Ming Wei. Forecasting carbon price using empirical mode decomposition and evolutionary least squares support vector regression. *Applied energy*, 191:521–530, 2017.