Science and Technology Indonesia
e-ISSN:2580-4391 p-ISSN:2580-4405
Vol. 10, No. 3, July 2025

Research Paper

Carbon Price Prediction by Incorporating Fossil Fuel Prices Using Long Short-Term Memory with Temporal Pattern Attention (TPA-LSTM)

Edo Priyo Utomo Putro Mujiono[1], Imam Mukhlash[1*], Yan Aditya Pradana[1], Endah R.M. Putri[1], Mohammad Isa Irawan[1]

[1]Department of Mathematics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember Surabaya, 60111, Indonesia
*Corresponding author: imamm@matematika.its.ac.id

Abstract
Reliable carbon price prediction can help to stabilize the carbon market, minimize financial risks for investors, and encourage innovation in green industries. The forecasts have a crucial role in reaching advanced goals for reducing emissions, aiding the shift toward an economy with reduced carbon emissions, and lessening the adverse effects of climate change overall. This paper proposes applications of LSTM with Temporal Pattern Attention (TPA-LSTM) to predict carbon price fluctuations. The prediction of carbon price fluctuations not only draws on its own historical information but also from its main predictors, including fossil fuel prices from 2018 to 2023. The TPA-LSTM method is a combined method that uses the LSTM layer as the initial input of the model. Furthermore, the output generated by the LSTM layer serves as the input to the TPA layer, which is then used to predict the carbon price for the following day. The model is tested by predicting the test data and calculating the evaluation results. The experimental results indicate that TPA-LSTM has surpassed other models in accuracy by showing better performance based on MSE, RMSE, MAE, and MAPE metrics.

Keywords
Greenhouse Gas, Carbon Price, Carbon Market, Long Short-Term Memory, Temporal Pattern Attention

Received: 25 October 2024, Accepted: 20 May 2025
https://doi.org/10.26554/sti.2025.10.3.856-865

1. INTRODUCTION

Data from the National Disaster Management Agency (BNPB) shows that natural disasters in Indonesia have become more frequent from 2015 to 2020. This is related to climate change caused by ongoing environmental pollution, one of which is air pollution by carbon gas emissions. Increased carbon emissions in the Earth's atmosphere can result in a greenhouse gas effect, an event where heat from sunlight that should be reflected into space is instead trapped by gases in the Earth's atmosphere (Yani et al., 2021). The main effect of greenhouse gases in accelerating global warming is a negative impact on the balance of natural ecosystems, economic development, and even the survival of humans and other creatures (Ullah et al., 2024). To address global climate change, carbon emissions trading markets have become an efficient tool for capping emissions while minimizing economic costs (Zhu et al., 2017).

The European Union initiated the carbon trading process when it set up the European Union Emissions Trading System, a system that began in 2005 and is currently the largest in the world based on the cap-and-trade principle. So far, the EU ETS has managed to cut carbon emissions at low costs due to a pricing mechanism for carbon emission allowances. Accordingly, under the EU's carbon importers' emissions regulations, the importers of high-emission goods have to pay fees based on what producers have paid (Green, 2021). This is an incentive mechanism to change their energy structure toward low-carbon energy sources for the regulated installations (Zhao et al., 2018). The EU Emissions Trading System puts the price of carbon at 62.45 Euros a ton of pollution emitted, and it keeps growing. Carbon prices are bound to fluctuate both in the form of outright increases or decreases (Stavins, 2020). A number of factors influence the high and low price of carbon: government policies, market demand, technology, environmental conditions, and so on (Yang et al., 2018). The main factors influencing carbon price include the price of fossil fuels, especially oil and coal (Chen and Liu, 2020). Fossil fuels remain the dominant source of energy supply in the world and the leading contributor to carbon emissions; thus, fossil fuel prices influences the demand for and supply of carbon prices (Li and Wang, 2019). Following different transmission channels, such as substitution effects, income effects, cost effects, and expectation effects, the crude oil prices and coal is able to drive the variation in carbon price (Nourry and Seetanah, 2020).

Carbon pricing forecasting is helpful for establishing carbon pricing mechanisms and provides actionable insights for decision making in production, operations, and investments (Narassimhan et al., 2018). Accurate predictions of price fluctuations in emerging commodity markets from carbon prices are essential for a wide range of stakeholders (Dama and Sinoquet, 2021). Carbon price levels reveal marginal reduction cost information, precise predictions help improve carbon market structures and shaping climate policies that support effective emission reductions (Duan et al., 2021).

Many forecasting methods have been developed in recent years to achieve accurate carbon price prediction results. Wang developed the boosting-ARMA model to address the problem of model selection and examined the applicability of the ARMA method for forecasting carbon prices (Wang, 2017). Ren and Lo (2017) used a generalized autoregressive conditional heteroscedasticity (GARCH) model to analyze the dynamics of carbon prices. Considering various carbon price predictors, especially those fossil energy and economic aspects, Han et al. (2019) then used a Back Propagation (BP)-based prediction termed the MIDAS-BP combination model to foresee carbon price fluctuations. Yang et al. (2018) develop a mixed Daubechies wavelet-genetic-algorithm-radial basis function neural network model and explored its application for forecasting prices in the European carbon emission market. Zhang and Yang (2016) proposed a multi-frequency combination prediction model for carbon market prices using polar symmetrical mode decomposition methods, nonlinear autoregression, neural networks, and machine support vectors. Pradana et al. (2023) used two algorithms to perform carbon price predictions in time series predictions, Deep Multilayer Perceptron and Long Short-Term Memory, based on carbon price data from Europe's EEX carbon market. Duan et al. (2023) applied the Attention mechanism-based Long Short-Term Memory, the so-called AttLSTM, to predict carbon price fluctuations on the EU emission trading market. On the other sides, Wei et al. (2023) applied TPA-LSTM to predict passenger flow, TPA-LSTM has shown better performance compared to other prediction methods such as CNN, LSTM, and CNN-LSTM. By utilizing the temporal pattern attention mechanism, TPA-LSTM can focus on important information in historical passenger data and capture significant temporal patterns, thereby being able to predict passenger flows more accurately, especially in identifying peak flows during peak hours on weekdays and weekends.

The contribution of this current research work to the world of predicting the price of carbon involves creating an LSTM-TPA-LSTM model. Thus, it works on improving a more robust method for the forecast, which includes choosing appropriate input variables the prices of carbon, crude oil, and coal that will necessarily help in achieving the precision of the forecast. The proposed TPA-LSTM not only mastered the temporal pattern of both carbon price data and variables but also gave a more precise weight in different parts of the time sequence, hence leading to more accurate predictions. This paper is going to fill the gaps in earlier methods of prediction and provide a more reliable tool for the understanding of future carbon price fluctuations.

Furthermore, this paper is structured as follows: part 2 experimental section, part 3 discusses implementation and results. It ends with part 4, which is the conclusion.

2. EXPERIMENTAL SECTION

This section discusses the architecture proposed in this paper. The TPA-LSTM model is a hybrid model, which combines LSTM and Temporal Pattern Attention models. The TPA-LSTM mechanism can process multi-variable time series with a focus on identifying relevant temporal patterns in the data. This mechanism combines the power of LSTM and CNN with attention mechanisms to improve prediction accuracy. The research methodology includes several stages, which are shown in Figure 1.

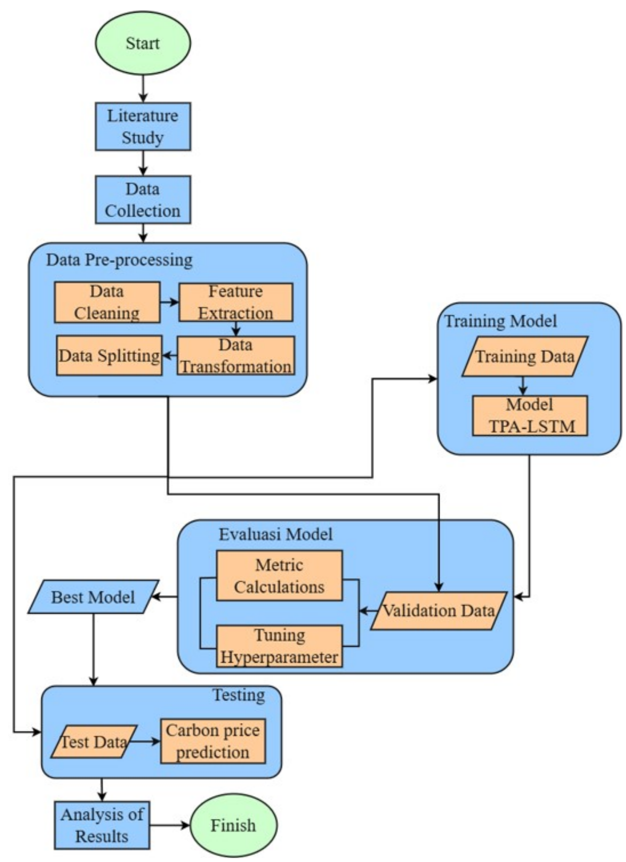

Figure 1. Research Flow Diagram

2.1 Data

Carbon price data is collected from investing.com. Investing.com is well renowned among professional investors worldwide. In October 2018 it ranked third among financial information websites worldwide, surpassing Bloomberg (Shen et al., 2024; Sayed et al., 2024). This data consists of the daily carbon

Table 1. Carbon Price Sample

Date	Price	Open	High	Low	Vol.	Change %
12/29/2023	77.98	78.00	78.61	77.28	NaN	0.05 %
12/28/2023	77.94	77.69	78.02	76.98	0.94K	0.26%
12/27/2023	77.74	77.27	78.57	77.08	0.96K	1.04%
12/22/2023	76.94	75.88	77.08	75.00	0.45K	1.40%
12/21/2023	75.88	73.50	76.42	73.50	0.82K	3.66%
...	...	...	...	...	...	...
01/08/2018	8.07	8.07	8.07	8.07	14.25K	-1.47 %
01/05/2018	8.19	8.19	8.19	8.19	5.56K	0.12%
01/04/2018	8.18	8.18	8.18	8.18	10.38K	-0.73 %
01/03/2018	8.24	8.24	8.24	8.24	7.96K	0.24%
01/02/2018	8.22	8.22	8.22	8.22	9.63K	-4.42 %

Table 2. Crude Oil Price Sample

Date	Price	Open	High	Low	Vol.	Change %
12/29/2023	71.65	71.99	72.62	71.25	214.49K	-0.17 %
12/28/2023	71.77	73.80	74.40	71.72	262.75K	-3.16 %
12/27/2023	74.11	75.32	75.66	73.77	253.32K	-1.93 %
12/26/2023	75.57	73.56	76.18	73.13	208.72K	2.41%
12/25/2023	73.79	73.49	73.94	73.48	NaN	0.31%
...	...	...	...	...	...	...
01/05/2018	61.44	61.90	62.04	61.09	563.04K	-0.92 %
01/04/2018	62.01	61.96	62.21	61.59	654.36K	0.62%
01/03/2018	61.63	60.39	61.97	60.28	673.86K	2.09%
01/02/2018	60.37	60.20	60.74	60.10	510.31K	0.22%
01/01/2018	60.24	60.26	60.28	60.15	NaN	-0.30 %

price of ICE shown in Table 1, the price of crude oil from real time capital.com shown in Table 2 and the price of coal from the NYSE from 2018 to 2023 shown in Table 3. In this study, the price variable from each of the data was used.

Table 3. Coal Price Sample

Date	Price	Open	High	Low	Vol.	Change %
12/29/2023	60.97	61.46	61.60	60.78	301.36K	-0.78 %
12/28/2023	61.45	61.80	62.08	60.95	365.04K	-1.33 %
12/27/2023	62.28	61.90	62.88	61.71	278.01K	1.02%
12/26/2023	61.65	61.81	62.36	61.39	279.29K	-0.24 %
12/22/2023	61.80	61.79	62.41	61.35	455.84K	0.64%
...	...	...	...	...	...	...
01/08/2018	18.69	18.35	18.80	18.23	947.87K	2.11%
01/05/2018	18.30	18.44	18.53	18.24	427.32K	-0.63 %
01/04/2018	18.42	18.24	18.53	18.07	572.13K	1.34%
01/03/2018	18.17	18.30	18.31	17.71	880.45K	0.11%
01/02/2018	18.15	17.24	18.17	17.04	906.12K	6.32%

2.2 Data Exploration

At this subsection, the initial analysis of the data is carried out to understand the relationship between the variables. Correlation matrices are used to identify linear relationships between feature variables in a dataset. This helps in understanding whether there are highly correlated features that might provide redundant or collinear information (Gong et al., 2024). This dataset revealed the link among crude oil, coal, and carbon prices.

2.3 Data Pre-processing

Before the data is used in forming the model, data pre-processing is carried out. The data pre-processing used includes handling for missing values, data cleaning, feature extraction, data transformation, and retrieval of a specified time range for test, validation, and training data (Pradana et al., 2023; Tawakuli et al., 2024).

2.3.1 Missing Value Handling

To handle the missing values in the datasets used by adjusting the number of records between the three carbon, crude oil, and coal datasets. Because of the daily data used to record dates on each dataset, the datasets have different record numbers, which is handled by unifying all datasets by taking the Date and Price columns.

2.3.2 Outlier Handling

In the combined carbon, crude oil and coal price dataset, there is an outlier on the crude oil prices in the form of negative values. A price cannot have a negative value so that it can be interpolated so that the data does not display a negative value.

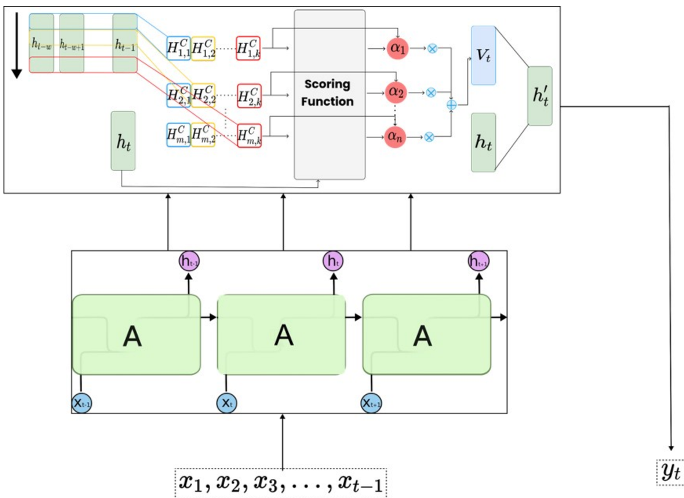

Figure 2. TPA-LSTM Model

2.3.3 Data Normalization

The dataset is normalized by changing the data scale in the range of 0 to 1 using MinMaxScaler so that the computational process can be performed efficiently. After normalization, the data was devided into three data, namely test, validation, and training data. The majority of the data was used for training, with smaller portions reserved for validation and testing, respectively. Here is the formula used Equation (1).

$$X_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

2.4 Model LSTM with Temporal Pattern Attention (TPA-LSTM)

The TPA-LSTM model is a hybrid model that integrates LSTM models with Temporal Pattern Attention shown in Figure 2. The TPA-LSTM mechanism can process multi-variable time series, putting attention on finding relevant temporal patterns in data (Shih et al., 2019). This mechanism encompasses the strengths of both LSTM and CNN, along with the attention mechanisms, for improving prediction accuracy.

The initial layer of TPA-LSTM is the LSTM. Remembering the order of information $x_1, x_2, \cdots, x_t$, where $x_i \in R^n$ with $i = [1, \cdots, t]$, RNNs generally define repeated, F, and compute functions $h_t \in R^m$ for every step of the time, t, shown in Equation (2).

$$h_t = F(h_{t-1}, x_t) \tag{2}$$

The F function's implementation varies depending on the type of RNN cell utilized. LSTMs, which are commonly employed, have slightly different recurrent functions:

$$h_t, c_t = F(h_{t-1}, c_{t-1}, x_t) \tag{3}$$

Equation (3) is defined by the following equation:

$$h_t = o_t \odot tanh(c_t) \tag{4}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot tanh(x_t W_{xg} + h_{t-1} W_{hg}) \tag{5}$$

$$o_t = \sigma(x_t W_{xo} + h_{t-1} W_{ho}) \tag{6}$$

$$f_t = \sigma(x_t W_{xf} + h_{t-1} W_{hf}) \tag{7}$$

$$i_t = \sigma(x_t W_{xi} + h_{t-1} W_{hi}) \tag{8}$$

where i_t, f_t, o_t dan $c_t \in R^m$, $W_{xi}, W_{xf}, W_{xo}, W_{xg} \in R^{mxn}$ dan $W_{hi}, W_{hf}, W_{ho}, W_{hg} \in R^{mxm}$, dan $\odot$ is wise product, and σ is a sigmoid function.

LSTM mechanism uses multiple gates to regulate the flow of information in the cell's memory. Input gate (it) determine what new information to store Equation (8). Forget gate (ft) determine which information to discard Equation (7). Output gate (ot) specifies the portion of the cell's memory to be ejected Equation (6). Cell state (ct) Equation (5) updated using hidden state (ht) Equation (4) generated from the current cell state and gate output. Here, xt is the input at the time t, h_{t-1} is the previous hidden state, and $Wxi, Wxf, Wxo, Wxg, Whi, Whf, Who, Whg$ is a weight matrix that connects the input and hidden state with the gate. Sigmoid function (σ) and tanh (*tanh*) are used as activation function to regulate the output of the gate and candidate cell state. The hidden state that has been processed by the LSTM becomes the input of the Temporal Pattern Attention.

Given hidden state from the previous LSTM $H = h_1, h_2, \cdots, h_{t-1}$, and initial input matrix $X = x_1, x_2, \cdots, x_t$, the short-term signal patterns and relationships among the three variables are extracted using Equation (9).

$$H^c_{i,j} = \sum_{l=1}^{w} H_{i(t-w-1+1)} \times C_{j(T-w+1)} \tag{9}$$

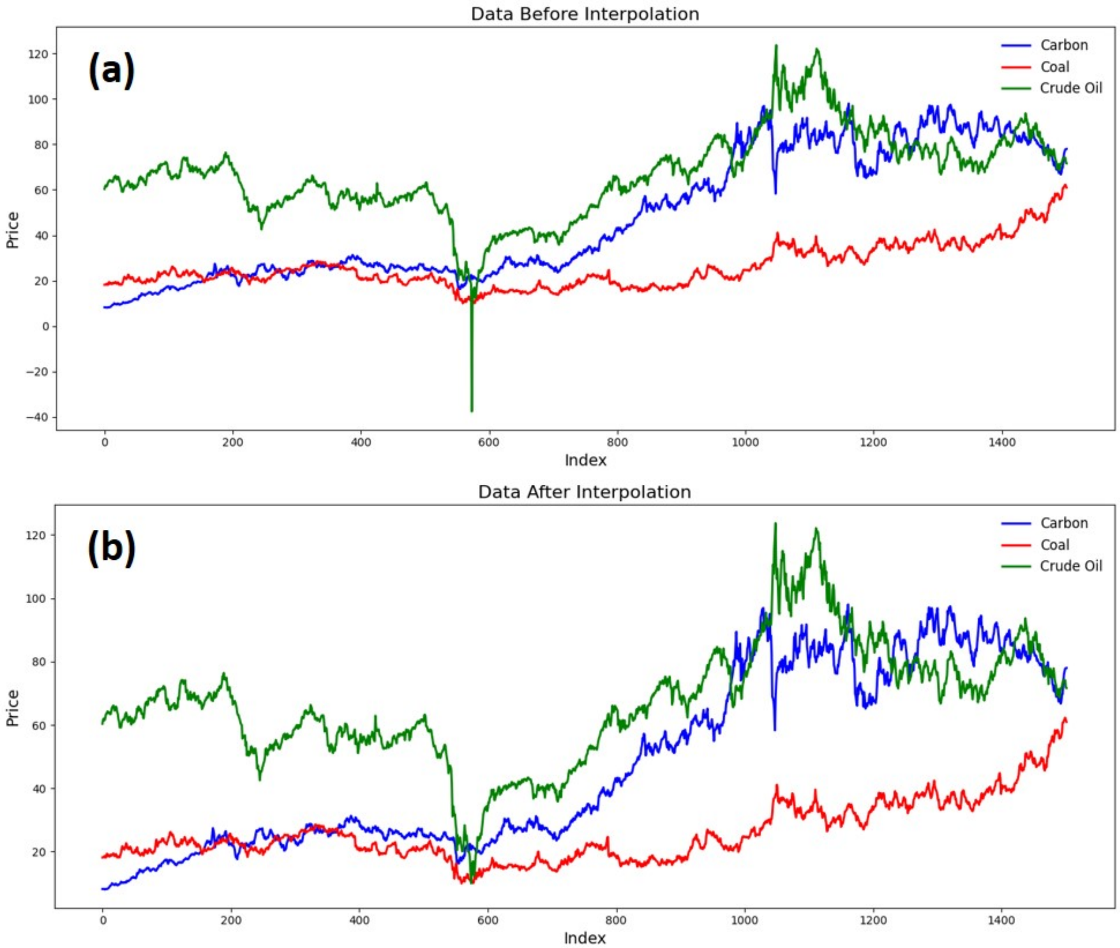

Figure 3. Carbon, Crude Oil and Coal Price Chart: (a) Before Interpolation; (b) After Interpolation

To assess the relevance, a computation is performed between the convolution of the i-th vector and the j-th filter. This value is denoted as $H^C_{(i,j)}$. By calculating v_t as a weighted number of rows vector from $H^C \in R^{nxk}$ defined in Equation (10) are the assessment functions f : $R^k \times R^m \to R^{kxm}$. To evaluate relevance.

$$f(H^c_i, h_t) = (H^c_i)^{\intercal} W_a h_t \tag{10}$$

where H^C_i is line i of H^C, dan $W_a \in R^{(km)}$. The weight of attention α_i obtained is shown in Equation (11).

$$\alpha_i = Sigmoid(f(H^c_i, h_t)) \tag{11}$$

Expect more than one variable to be valuable for forecasting so utilize the sigmoid activation function instead of softmax. Finishing process, line vector H^C, weighed by α_i to get the context vector $v_t \in R^k$shown in Equation (12).

$$v_t = \sum_{i=1}^{n} H^c_i \alpha_t \tag{12}$$

Then we integrate *vt* and *ht* to generate the final prediction shown in Equation (13) and Equation (14).

$$h'_t = W_h h_t + W_v v_t \tag{13}$$

$$y_{t+1 \leftarrow A} = W'_h h'_t \tag{14}$$

where $h_t, h'_t \in R^m, W_h \in R^(mm), W_v \in R^{mxk}$, dan $W_{h'} \in R^{nxm}$ and $y_{t-1+\Delta} \in R^n$.

In MTS forecasting, MTS is given, $X = x_1, x_2, \ldots, x_t$, where $x_i \in R^n$ shows the value seen at the time step i. The forecasting task aims to predict the value of $x_{t-1+\Delta}$, where Δ is a fixed horizon depending on the specific task. The predicted value is denoted as as $y_{t-1+\Delta}$, while the truth is represented as

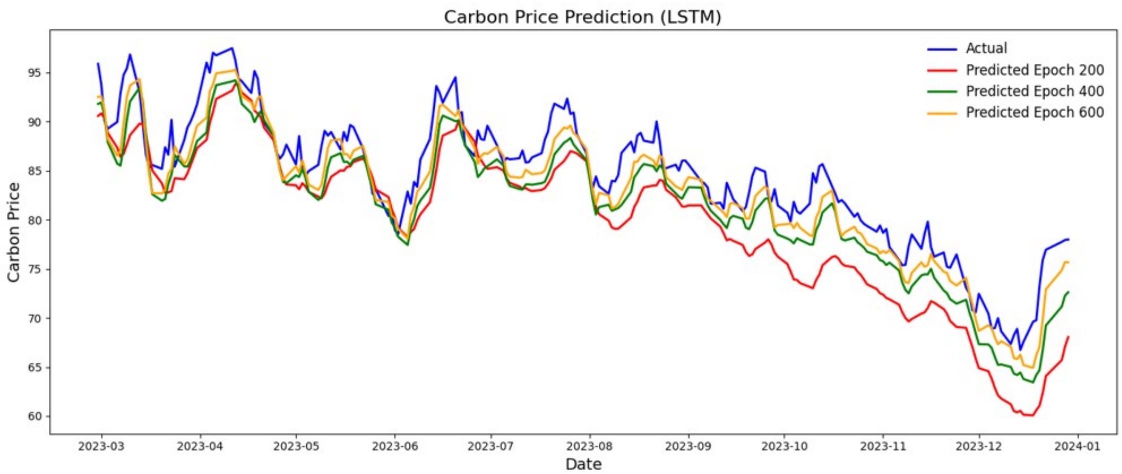

Figure 4. Carbon Price Prediction Graph using LSTM

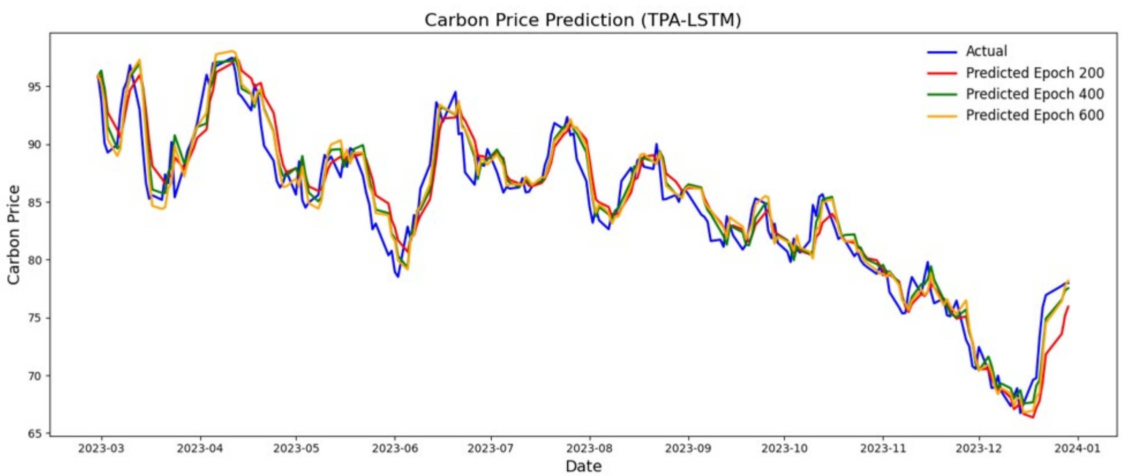

Figure 5. Carbon Price Prediction Graph using TPA-LSTM

$y_{t-1+\Delta}$=$x_{t-1+\Delta}$. To forecast $x_{t-1+\Delta}$, the model utilizes a sequence of past observations $x_{t-w}, x_{t-w+1}, \cdots, x_{t-1}$, where w is window size.

2.5 Performance of Metrics

The prediction outputs from the TPA-LSTM model were analyzed using Mean Squared Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Percentage Error (MAPE). The mathematical formulation of the error metrics is presented in the following equation.

2.5.1 Mean Squared Error (MSE)

MSE represents the average of the squared differences between the actual values and the predicted values. A low or almost zero MSE shows that the prediction results depend on actual data and are dependable for further forecasts (Pradana et al., 2023). Here is the formula used Equation (15).

$$MSE = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n} \tag{15}$$

where :
n represents the total count of samples,
y_i represents the actual value for the ith sample,
$y_{\hat{i}}$ represents the predicted value for the ith sample.

2.5.2 Root Mean Square Error (RMSE)

Calculating the RMSE requires square rooting the MSE. A low RMSE value signifies higher accuracy in the error estimation method (He et al., 2023). Here is the formula used Equation (16).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \tag{16}$$

where :
n represents the total count of samples,
y_i represents the actual value for the ith sample,
$y_{\hat{i}}$ represents the predicted value for the ith sample.

2.5.3 Mean Absolute Error (MAE)

MAE is the average of the absolute values of the errors, reflecting the overall magnitude of the deviations without considering their direction (Huang et al., 2021). Here is the formula used Equation (17).

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y_i| \tag{17}$$

where :
n represents the total count of samples,
y_i represents the actual value for the ith sample,
$y_{\hat{i}}$ represents the predicted value for the ith sample.

2.5.4 Mean Percentage Error (MAPE)

MAPE presents errors as percentages, so enhancing the perception of the model's correctness. Here is the formula used Equation (18).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{y_i - \bar{y}_i}{y_i}\right| \tag{18}$$

where :
n represents the total count of samples,
y_i represents the actual value for the ith sample,
$y_{\hat{i}}$ represents the predicted value for the ith sample.

3. RESULTS AND DISCUSSION

3.1 Data Collection

We use the latest data from investing.com, which is daily historical data on carbon, crude oil and coal prices from 2018 to 2023. The dataset downloaded has a time range from January 1, 2018 - December 31, 2023. This carbon price dataset consists of 1543 records with a number of 7 columns. In the crude oil price dataset, there are 1585 records with a number of 7 columns. The Coal price data contains 1508 records with a number of 7 columns. In this paper, the utilized attribute is the price, representing the closing price of each dataset.

3.2 Data Pre-processing

In the combined carbon, crude oil and coal price dataset, there is an outlier on the price of crude oil in the form of negative values shown in Figure 3a. The price cannot have negative values and needs to be interpolated to eliminate the negative values. The interpolated data is shown in Figure 3b.

3.3 Carbon Price Prediction Results

The model was compiled using Adaptive Moment Estimation (Adam) optimization and the loss function Mean Square Error (MSE). After conducting several training experiments to obtain optimal parameters, the model training parameters were obtained, including window size of 16, batch size of 16, number of filters of 81, hidden size of 81, learning rate of 0.0001, and number of epochs of 600.

Table 4. MAE Trial Results with Several Variations

Epoch	LSTM			TPA-LSTM		
	1 day	5 days	10 days	1 day	5 days	10 days
200	3.983	1.881	3.439	0.432	2.272	2.451
400	3.787	2.157	3.316	0.156	1.746	1.965
600	3.686	2.086	3.150	0.667	1.450	1.826

Table 5. MSE Trial Results with Several Variations

Epoch	LSTM			TPA-LSTM		
	1 day	5 days	10 days	1 day	5 days	10 days
200	15.867	4.951	16.482	0.187	14.048	8.597
400	14.343	6.312	15.821	0.024	5.260	5.342
600	13.585	6.009	13.591	0.004	3.518	5.193

Table 6. RMSE Trial Results with Several Variations

Epoch	LSTM			TPA-LSTM		
	1 day	5 days	10 days	1 day	5 days	10 days
200	3.98	2.225	4.059	0.432	3.748	2.932
400	3.787	2.512	3.977	0.156	2.293	2.311
600	3.686	2.451	3.686	0.067	1.875	2.278

Table 7. MAPE Trial Results with Several Variations

Epoch	LSTM			TPA-LSTM		
	1 day	5 days	10 days	1 day	5 days	10 days
200	4.15%	2.02%	3.64%	0.45%	3.61%	2.66%
400	3.95%	2.33%	3.52%	0.16%	1.92%	2.12%
600	3.84%	2.26%	3.36%	0.07%	1.59%	1.97%

After the model has been trained, it then makes predictions on the test dataset. The prediction values are stored in a variable which is then converted back to the original scale to be able to calculate the evaluation value using MAE, MSE, RMSE, and MAPE. The trial was carried out by making variations on the number of epochs carried out for data training and np subsequent data prediction. The variation in the number of epochs includes: 200, 400, 600. The carbon price prediction graph shows that both the LSTM (Figure 4) and TPA-LSTM (Figure 5) models show an increasing trend in accuracy as the number of training epochs increases for 200, 400, and 600 epochs. This is indicated by the increasing approach of the prediction line to the actual line, especially in the TPA-LSTM model. The higher the number of epochs, the model has more opportunities to learn historical data patterns so that the prediction results become more precise.

Table 8. MAE Trial Results with Several Variations

Epoch	1 input			3 input		
	1 day	5 days	10 days	1 day	5 days	10 days
200	1.953	1.573	2.656	0.432	3.272	2.451
400	1.413	2.751	2.423	0.156	1.746	1.965
600	0.442	1.724	1.982	0.067	1.450	1.826

Table 9. MSE Trial Results with Several Variations

Epoch	1 input			3 input		
	1 day	5 days	10 days	1 day	5 days	10 days
200	3.814	3.174	9.476	0.187	14.048	8.597
400	1.997	10.554	9.313	0.024	5.260	5.342
600	0.196	4.579	5.811	0.004	3.518	5.193

3.4 Result Analysis

The trial was carried out by making variations on the number of epochs carried out for data training and np subsequent data prediction. The variation in the number of epochs includes: 200, 400, 600. As for the number of predictions for the next data, namely the next 1, 5, 10 days. The results of the calculation of the evaluation metrics of MAE shown in Table 4, MSE shown in Table 5, RMSE shown in Table 6 and MAPE shown in Table 7.

The best results were obtained in the TPA-LSTM model trial with the number of epochs of 600 and the number of data predictions in the next 1 day with MAE values of 0.067, MSE of 0.004, RMSE of 0.067, and MAPE of 0.07%. Overall, in the trial, all scenarios of the TPA-LSTM model succeeded in capturing the temporal pattern from the data shown by the prediction graph that followed the pattern or graph from the original data. This is because temporal pattern attention is used to capture temporal patterns in carbon, crude oil, and coal price data and then give different weights to different parts of the time sequence. Temporal pattern attention helps the model to concentrate on the most significant time steps.

Besides conducting experiments using multiple inputs and multiple outputs, we also perform experiments with a single input and output, focusing solely on the carbon price. This experiment is carried out to prove that carbon prices are influenced not only by their own historical value but also by other prices of fossil fuels, particularly the prices of coal and oil (Duan et al., 2023). The experiments are carried out by adjusting the number of training epochs to 200, 400, and 600. Additionally, predictions were made for 1, 5, and 10 days ahead. The experimental finding indicates that predicting carbon prices using 3 attributes is better than using 1 attribute. The results of the calculation of the evaluation metrics of MAE shown in Table 8, MSE shown in Table 9, RMSE shown in Table 10 and MAPE shown in Table 11.

Table 10. RMSE Trial Results with Several Variations

Epoch	1 input			3 input		
	1 day	5 days	10 days	1 day	5 days	10 days
200	1.953	1.781	3.078	0.432	3.748	2.932
400	1.413	3.248	3.051	0.156	2.293	2.311
600	0.442	2.140	2.410	0.067	1.875	2.278

Table 11. MAPE Trial Results with Several Variations

Epoch	1 input			3 input		
	1 day	5 days	10 days	1 day	5 days	10 days
200	2.03%	1.72%	2.83%	0.45%	3.61%	2.66%
400	1.47%	3.01%	2.62%	0.16%	1.92%	2.12%
600	0.46%	1.89%	2.14%	0.07%	1.59%	1.97%

Finally, all results have shown that TPA-LSTM performs better in carbon price forecasting while using a multi-input-multi-output framework. In such a setting, the model's predictive capabilities are significantly improved with the inclusion of key inputs such as carbon price, crude oil price, and coal price. In the multi-input setup, these additional input variables help the model capture more dimensions of variation responsible for driving carbon price fluctuations and, hence, generate better forecasts. This contrasts with the single input-single-output experiments, which were conducted on carbon prices alone, where the method TPA-LSTM is doing well but without the enriched precision observed in the multi-input setup. Including relevant energy commodities, such as crude oil and coal, therefore complements the understanding of the market fluctuations at play and allows the TPA-LSTM to carry out multi-input-multi-output carbon price predictions more effectively.

4. CONCLUSIONS

In this paper, the TPA-LSTM method has been implemented to predict carbon prices based on carbon, crude oil, and coal price data starting from 2018 to 2023. The TPA-LSTM model succeeded in recognizing temporal patterns in the data to produce good predictions for the data. In the trial of several scenarios from the TPA-LSTM model, it was shown that in the hyperparameter tuning test scenario with a window size of 16, a batch size of 16, a number of filters of 81, a hidden size of 81, a learning rate of 0.0001, and a number of epochs 600 results in the most accurate prediction. In the test, the LSTM model was also used; however, it was less effective at capturing temporal patterns compared to the TPA-LSTM model, resulting in lower prediction accuracy. This is because LSTM has limitations in retaining information from very long temporal dependencies and often has difficulty in focusing on the most relevant information. In contrast, the attention mechanism in TPA-LSTM allows it to concentrate on the most relevant parts of the input sequence, leading to improved prediction accuracy. The TPA-LSTM approach demonstrates superior accuracy in forecasting next-day carbon prices relative to the LSTM method, there is an increase in accuracy with MAE values of 0.067, MSE of 0.004, RMSE of 0.067, and MAPE of 0.07%.

5. ACKNOWLEDGEMENT

The research received funding from the Ministry of Education, Culture, Research, and Technology through the Postgraduate Research Scheme under contract number 1707/PKS/ITS/2024.

REFERENCES

Chen, Y. and Z. Liu (2020). Key Factors Affecting Carbon Prices from a Time-Varying Perspective. *Energy Economics*, **92**; 104956

Dama, F. and C. Sinoquet (2021). Time Series Analysis and Modeling to Forecast: A Survey

Duan, K., X. Ren, Y. Shi, T. Mishra, and C. Yan (2021). The Marginal Impacts of Energy Prices on Carbon Price Variations: Evidence from a Quantile-on-Quantile Approach. *Energy Economics*, **95**; 105131

Duan, K., R. Wang, S. Chen, and L. Ge (2023). Exploring the Predictability of Attention Mechanism with LSTM: Evidence from EU Carbon Futures Prices. *Research in International Business and Finance*, **59**; 101314

Gong, H., Y. Li, J. Zhang, B. Zhang, and X. Wang (2024). A New Filter Feature Selection Algorithm for Classification Task by Ensembling Pearson Correlation Coefficient and Mutual Information. *Engineering Applications of Artificial Intelligence*, **131**; 107865

Green, J. F. (2021). Does Carbon Pricing Reduce Emissions? A Review of Ex-Post Analyses. *Environmental Research Letters*, **16**(4); 043004

Han, M., L. Ding, X. Zhao, and W. Kang (2019). Forecasting Carbon Prices in the Shenzhen Market, China: The Role of Mixed-Frequency Factors. *Energy*, **171**; 69–76

He, K., Q. Yang, L. Ji, J. Pan, and Y. Zou (2023). Financial Time Series Forecasting with the Deep Learning Ensemble Model. *Mathematics*, **11**(4); 1054

Huang, Y., X. Dai, Q. Wang, and D. Zhou (2021). A Hybrid Model for Carbon Price Forecasting Using GARCH and Long Short-Term Memory Network. *Applied Energy*, **285**; 116485

Li, J. and X. Wang (2019). Examination of the Transmission Mechanism of Energy Prices and Carbon Prices Based on the VAR Model. *Energy Reports*, **5**; 1398–1406

Narassimhan, E., K. S. Gallagher, S. Koester, and J. R. Alejo (2018). Carbon Pricing in Practice: A Review of Existing Emissions Trading Systems. *Climate Policy*, **18**(8); 967–991

Nourry, M. and B. Seetanah (2020). Testing the Impact of Renewable Energy and Oil Price on Carbon Emission Intensity in the Transportation Sector. *Renewable Energy*, **151**; 1212–1220

Pradana, Y. A., I. Mukhlash, M. I. Irawan, and E. R. Putri (2023). Carbon Price Prediction in the European Market Using Deep Learning. In *2023 10th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*. pages 307–314

Ren, C. and A. Lo (2017). Emission Trading and Carbon Market Performance in Shenzhen, China. *Applied Energy*, **193**; 414–425

Sayed, G. I., E. I. Abd El-Latif, A. Darwish, V. Snasel, and A. E. Hassanien (2024). An Optimized and Interpretable Carbon Price Prediction: Explainable Deep Learning Model. *Chaos, Solitons & Fractals*, **188**; 115533

Shen, D., Z. Tong, and J. W. Goodell (2024). Do Online Message Boards Convey Cryptocurrency-Specific Information? *International Review of Financial Analysis*, **91**; 102950

Shih, S.-Y., F.-K. Sun, and H.-Y. Lee (2019). Temporal Pattern Attention for Multivariate Time Series Forecasting. *Machine Learning*, **108**; 1421–1441

Stavins, R. N. (2020). The Future of US Carbon-Pricing Policy. *Environmental and Energy Policy and the Economy*, **1**(1); 8–64

Tawakuli, A., B. Havers, V. Gulisano, D. Kaiser, and T. Engel (2024). Survey:Time-series data preprocessing: A survey and an empirical analysis. *Journal of Engineering Research*

Ullah, A., M. Dogan, A. Pervaiz, A. A. A. Bukhari, H. T. Akkus, and H. Dogan (2024). The Impact of Digitalization, Technological and Financial Innovation on Environmental Quality in OECD Countries: Investigation of N-Shaped EKC Hypothesis. *Technology in Society*, **77**; 102484

Wang, N. (2017). Forecasting of Carbon Price Based on Boosting-ARMA Model. *Statistical Information Forum*, **32**; 28–34

Wei, L., D. Guo, Z. Chen, J. Yang, and T. Feng (2023). Forecasting Short-Term Passenger Flow of Subway Stations Based on the Temporal Pattern Attention Mechanism and the Long Short-Term Memory Network. *ISPRS International Journal of Geo-Information*, **12**; 25

Yang, X., J. Liang, J. Jiang, and M. Jun-Long (2018). Price Forecasting Algorithm of Carbon Emission Rights Under

Multiscale Fractal Characteristics. *Control Theory & Applications*, **35**; 224–231

Yani, I., Y. Resti, F. Burlian, and A. Yani (2021). Prediction of Plastic-Type for Sorting System Using Fisher Discriminant Analysis. *Science and Technology Indonesia*, **6**(4); 313–318

Zhang, C. and X. Yang (2016). Price Prediction of China's Regional Carbon Market Based on Multi-Frequency Combination Model. *System Engineering Theory & Practice*, **36**; 3017–3025

Zhao, X., M. Han, L. Ding, and W. Kang (2018). Usefulness of Economic and Energy Data at Different Frequencies for Carbon Price Forecasting in the EU ETS. *Applied Energy*, **216**; 132–141

Zhu, B., D. Han, P. Wang, Z. Wu, T. Zhang, and Y. M. Wei (2017). Forecasting Carbon Price Using Empirical Mode Decomposition and Evolutionary Least Squares Support Vector Regression. *Applied Energy*, **191**; 521–530

