Received: 23 December 2022, Accepted: 22 December 2023, Published: 29 December 2023, Publisher: UTP Press, Creative Commons: CC BY 4.0

FORECASTING COAL PRICES USING DECISION TREE APPROACH

R. Murugesan, Umesh Shinde

Department of Humanities and Social Sciences, National Institute of Technology, India

*E-mail: rmurugan@nitt.edu

ABSTRACT

Coal is an important fuel in the world's energy systems. Today's world depends heavily on coal. It contributes to 41% of the world's power and acts as a key ingredient in the production process of 90% of cement and 70% of the world's steel. An accurate prediction model for forecasting future coal prices would provide critical information and a timely caution to governments to ensure a steady energy supply. Studies on forecasting coal prices regarding related commodities such as crude oil, iron, natural gas, and paper are rare. Hence, this research aims to forecast coal prices using historical data for the 2013-2016 period. The study uses a decision tree model to anticipate the coal prices, and the impulse response function and variance decomposition are used to show the dynamic connections of coal prices with various variables. The model in predicting coal prices performs well with mean absolute percentage errors of around 5%. The study results will likely enable new insights in providing valuable inputs for all the stakeholders.

Keywords: *Coal price forecasting, decision tree, prediction*

INTRODUCTION

Coal is one of the main energy sources worldwide, known as the "black gold" and "industrial food" (Wu, Xia, & Liu, 2023). Notwithstanding the increasing appeal of renewable energy sources, such as wind and solar, coal is anticipated to remain a mainstream energy in the next few decades. As a crucial bulk commodity, the fluctuation of coal prices profoundly affects the development of coal enterprises, upstream and downstream industries, and even the stability of the global economy (Zhang, Cao, Thé, & Yu, 2022).

Several industries use coal, including those that produce electricity, iron and steel, cement, and liquid fuel. Most of the coal is either used to produce iron and steel (coking coal) or to generate electricity (using lignite or steam coal) (Zhao et al., 2012). Power generation relies heavily on coal, and this dependence is expected to persist. 41% of the world's electricity is now produced by coal, and over the next 30 years, the anticipated percentage will not change.

Given the importance of coal as a strategic energy source, forecasting coal prices is critical for investors and policymakers (Xu & Zhang, 2022a). The modelling and predicting volatility is a significant problem in risk management, asset allocation, commodity pricing, and policymaking. Furthermore, predicting volatility is a critical task in economic markets, which has long piqued the interest of practitioners and academicians. The price instability of mining ventures has had a significant impact on recent mining ventures and related activities. (such as those involving gold, coal, and iron ore) (Alameer et al., 2020). Accurate coal price forecasting (Zhu et al., 2017) will aid in the decision-making of energy policymakers engaged in both the supply and demand of coals (Wang, 2017; Zhang et al., 2022; Hintermann, 2010; Wei et al., 2010; Seifert et al., 2008, Keppler & Mansanet 2010). Therefore, investigating a coal price prediction model in a market economy is useful (Mei et al., 2014).

Studies related to forecasting coal prices in terms of other related commodities such as crude oil, iron, natural gas, and paper are very rare. Hence, this research aims to forecast coal prices using historical data for the 2013-2016 period. Since precise and consistent coal price forecast is of huge consequence to augment the steadiness of the coal market, many forecasting methods are developed to predict the price of Coal, including conventional econometrics models and machine learning approaches (Tsi et al., 2013; Zhu et al. 2017; Mei et al. 2014; Ballings et al. 2015; Gumus et al., 2017).

Literature shows that machine learning can provide more accurate predictions than econometrics models. Recently, research has explored the possibility of using a decision tree-based machine-learning algorithm to predict coal prices. Decision tree learning (Breiman et al., 1984) is a popular data mining methodology that was used to forecast stock prices, oil prices, and property prices (Fan et al., 2006; Tiwari et al., 2015; & Bhar et al., 2015), in addition to assessing business performance (Delen et al., 2013). A decision tree algorithm separates a data set via recursive partitioning in order to train a model, which is then used to forecast the value of a dependent variable based on the independent variables (Breiman et al., 1984). As a popular machine learning approach, decision tree models have great predictive power in some studies. Moreover, unlike other machine learning models that are considered 'black boxes' due to difficulties in interpretations, decision tree models are also interpretable in theory (Loh, 2014). A decision tree is an attractive machine-learning method due to its efficiency, robustness, and relatively simple structure (Quinlan, 1986).

Since oil, coal, paper, and natural gas still account for a significant share of the globalenergy package, these are regarded as the central industries within the energy stock markets. Accordingly, almost all global economies have traditionally relied on these fossil fuels to meet their respective energy. Thus, predicting the trends in energy market-related returns is deemed important from the points of view of both energy producers and consumers (Alam et al., 2023). A major research gap in coal price prediction exists after reviewing previous studies extensively (Zhang et al., 2022). Hence, this research aims to employ a decision tree machine learning algorithm for the prediction task of coal prices with other independent variables such as crude oil, iron, natural gas, and paper. These independent variables are considered influencing variables with regard to coal prices. The study has utilized the impulse response function and variance decomposition to show the dynamic connections of coal prices with other variables under investigation.

The results of the study show that the decision tree-based technique is capable of forecasting coal prices both in the short and long term. This approach is versatile and general in nature. The forecasting effect of the decision tree model is better than the benchmark model in terms of stability and accuracy (Wu et al., 2023). The experimental results of the distinctive coal price model establish that the proposed strategies are all valuable for improving forecasting performance (Zhao et al., 2021). Moreover, the proposed model outperforms all state-of-the-art baseline models in terms of both model accuracy and stability, which can be used for short-term forecasting of coal prices. The model also shows that additional variables can be included in the training. In this study, four extraneous factors that influence coal prices were considered for the study. The study enables new insights in providing valuable inputs for all the stakeholders

RELATED STUDIES

Zhao et al. (2022) applied a hybrid multi-step model using the decomposition method and deep learning algorithms to predict coal prices. Xu and Zhang (2022a) builds a non-linear autoregressive neural network model for the prediction of coal prices using the data of closing prices documented daily of coal traded in China Zhengzhou commodity exchange. The analysis made on the coal price forecast has been considered good for policymaking purposes. When doing policy research that entails creating perspectives on price patterns, the prediction results presented here could serve as independent technical predictions and be shared with other projections.

In order to assess transmission between exchange rate, coal price, palm oil price, and CPI-consumer price index, Chandrarin et al. (2022) used dynamic. Connectedness under the time-varying parameter vector autoregression approach contributed by Antonakakis and Gabauer (2017). The study took into account the degree of overall interconnectedness between the many factors that were being examined. Additionally, their analysis demonstrated the value

of research in developing policies to improve the advantages of coal and palm oil price rents.

A hybrid deep-learning model was used by Alameer et al. (2020) to predict future monthly fluctuations in coal prices. The accuracy of the following forecasters for predicting changes in the price of coal is supported by this study: prices for copper, natural gas, iron, silver, crude oil, and gold, in addition to exchange rates for AUD, IDR, and RMB. This study emphasized how acquiring coal mine assets from abroad could significantly reduce China's expenditure on coal imports by developing an accurate forecast model for predicting coal prices multiple steps in advance. Matyjaszek et al. (2019) looked at how well traditional time series models, ROBUST models, ARIMA models, GRNNs, and MLFNs predicted coking coal prices by analyzing and considering the full-time series and taking into account the transgenic time series. The latter serves as the foundation for a theory that discusses the existence of cycles in the price of raw materials and the occurrence of anomalous phenomena brought on by these cycles, and it removes these cycles from the time series in order to increase forecasting accuracy.

Ferrari et al. (2021) focused on the forecasting of quarterly nominal global energy prices of commodities, including oil, gas, and coal, using the Global VAR dataset suggested by Mohaddes and Raissi (2018) in their research, "Forecasting Energy Commodity Prices: A Large Global Dataset Sparse Approach." This dataset contains several trimestral macroeconomic variables for the 33 largest economies that may be enlightening. This study applied dynamic factor models, which are based on the personalized maximum likelihood method, allowing shrink parameters to 0 and evaluating sparse factor loadings. The study exhibited the significance of estimating coal price forecasting as one of the key findings. In another survey executed by Li et al. (2022), the presence and factors determining the multiple bubbles in the steam coal market since the eradication of the double-tract system of coal and electricity in 2012 were verified using the Generalized Supremum Augmented Dickey-Fuller. The test confirmed the explosive bubbles commenced, which were caused because of policy changes with reference to coal prices and environmental regulations. The research insisted that Producers, consumers, and investors are required to pay consideration to the spillover of price bubbles among regions, especially in major coal-producing areas. In yet another study done by Kong, Dong, and Jiang (2019) on the research topic entitled Forecasting the development of China's coal-to-liquid industry under security, economic, and environmental constraints emphasized the importance of forecasting coal prices for the sake of developing Coal-to-Liquid technology as a replacement to assuage oil deficiency. Wang et al. (2020), in their empirical study, used the difference-in-differences model to quantitatively appraise the action consequence of coal de-capacity policies on coal prices where they concluded that coal price prediction could solve the problem of decline in coal prices causing overcapacity, which questions the survival of coal industries that eventually affects economic growth, energy security, and industrial raw materials supply.

An accurate forecasting model for future coal price fluctuations provides critical information and early warning for the government and policymakers to provide a stable supply of energy, which is considered the main concern for China's policymakers. Alameer et al. (2020) proposed a deep-learning model for accurately forecasting monthly coal price fluctuations at different horizons. The proposed LSTM–DNN model combined long-term memory (LSTM) and deep neural network (DNN) to demonstrate how the LSTM– DNN model leads to improvements in predictive accuracy. The study compared the results of the LSTM–DNN model to other competitive models, which include multilayer perceptron neural network (MLP) and support vector machine (SVM). Experimental results of the study showed the superiority of the hybrid LSTM–DNN model over the other competitive models and its capability to forecast multiple steps accurately and flexibly.

Many researchers have paid close attention to a diverse variety of time series techniques for price forecasts. The models commonly found in the literature include autoregressive moving average models, vector error correction, and many variations of these models. For coal, these models have also been employed to forecast its consumption and investment in addition to prices (Matyjaszek et al., 2019; Zhang et al., 2020) and price probability densities (Ding et al., 2021). Recently, machine-learning approaches have shown great potential for forecasting consumption and prices (Gao & Wang, 2021).

Machine learning techniques often seen in the literature include the neural network, support vector regression, random forest, boosting (Ding et al., 2021; Wang & Kang, 2021), and Gaussian process regression (Martinho et al., 2020). In particular, previous studies have suggested neural networks' great potential in the area of time series forecasts for economic data, many of which are found to be rather chaotic. Previous research has demonstrated that neural networks can generate decently performance in terms of accuracy across various forecast settings. These results might benefit from the decision tree's self-learning. Capability (Xu & Zhang, 2022b) and design for characterizing potential non-linearities (Altan et al., 2021) in many time series. A decision tree's advantage over other non-linear methods is that a large number of functions can be approximated well through different multilayer designs. The present work will examine a decision tree for thermal coal price forecasts. It is investigated via the non-linear auto-regressive neural network. By examining model settings over different algorithms, delays, hidden neurons, and data splitting ratios, the model leading to generally decent accuracy and stability is constructed. One might use results here for standalone technical forecasts or combine the results with other forecasts when conducting policy analysis that needs price trend perspectives. The framework of forecasts here might also be able to be generalized to related forecast problems in other different economic sectors. We contribute to the literature in the following ways. First, to our knowledge, this is the first study examining the price forecast problem, keeping other influencing variables such as iron, natural gas, paper, and oil, which are all important financial indexes with significant social impacts.

In this work, we demonstrated that the decision tree learning-based system can effectively forecast future coal prices with an error rate of less than 5%. Additionally, this study discovered that the method's forecasts are accurate across various time ranges, ranging from days to years, and the resultant precision was sustained even with longer predictions. This research strongly believes that this methodology will be beneficial in numerous applications for predicting commodity prices due to its simplicity and accuracy. The following are some ways that this study adds to the body of literature. Firstly, to the best of knowledge, this is the first study to look at the coal price forecasting problem in relation to other variables, which is a significant financial index with a big social impact. The Indian market was examined in the earlier studies on coal price forecasts (reviewed above), but they focused on a different series than the one taken into account in the present study (Ding et al., 2021). Secondly, decision-making benefited from the investigation of daily prices in this context, which should be crucial for forecast users like traders, hedgers, and speculators who frequently use price forecasts. Except for a few studies (Zhang et al., 2022) that look at daily data, the majority of the prior research on coal price forecasts has focused on annual, monthly, and weekly data.

METHODOLOGY

A variety of variables may cause a change in coal prices. For instance, historical coal prices are likely to influence the price of coal at some point in the future. In addition, various metal prices on the market may be related to coal prices as alternatives. For instance, the price of coal is projected to decrease in line with other energy products. Additionally, because of the link to coal production, the price of coal is likely to be influenced by energy costs, as well as the cost of both crude oil and natural gas. Additionally, industrial demand and consumption are subsequently impacted by the broader economic environment and may be tied to the price of coal (Gargano & Timmermann, 2014). Finally, it is possible that the price of coal can be affected by unanticipated external factors. For instance, coal and paper prices may coincide unexpectedly. We gathered the prices of coal, crude oil, natural gas, iron, and paper from October 22, 2010, to December 31, 2016, based on the justification and subject to the availability of data. The data used in this study was taken from www.investing.com. Weekends were not included in this analysis because there was no information about them in the raw data. The holiday prices are represented using the averaged values of the previous day and the next day; however, holidays are sporadic and rare, then,

$$P_i = 0.5 \times (P_{i-1} + P_{i+1}) \quad (1)$$

If a holiday falls on the i-th day. The final processed data set included prices going back 1543 days. According to Quinlan (1992), the prominent advantage of the decision tree is that it allows for interpretation easily after making the prediction. The raw plot of all the under consideration has been represented in Figure 1.

As shown in Figure 1, all the respective variables under consideration show greater variation, which indicates the presence of volatility, random fluctuations, and uncertainties. Hence, normality distribution tests for the variables are required to be done. The results of the summary statistics (skewness, kurtosis) and Jarque-Bera test are shown in Table 1.

Table 1 contains the data's summary statistics. The price series and its first differences series are not observed to follow normal distributions in Table 1, which is probably not surprising for economic data (Xu, 2022a). Table 1 specifically demonstrates that the price series' skewness and kurtosis are, respectively, 1.2887 and 4.5221 (for coal), 1.2987 and 4.4221 (for natural gas), 1.1891 and 3.9998 (for iron), and 0.9987 and 3.8989 (for paper) and that the Jarque-Bera test's p-value is well below 0.001, rejecting the null hypothesis of a normal distribution. The first differences in the price series have skewness and kurtosis of 1.7697 and 21.4107 for coal, -1.6891 and 20.5012 for natural gas, -.7561 and 19.2356 for iron, and -1.6563 and 19.1256 for paper, respectively.

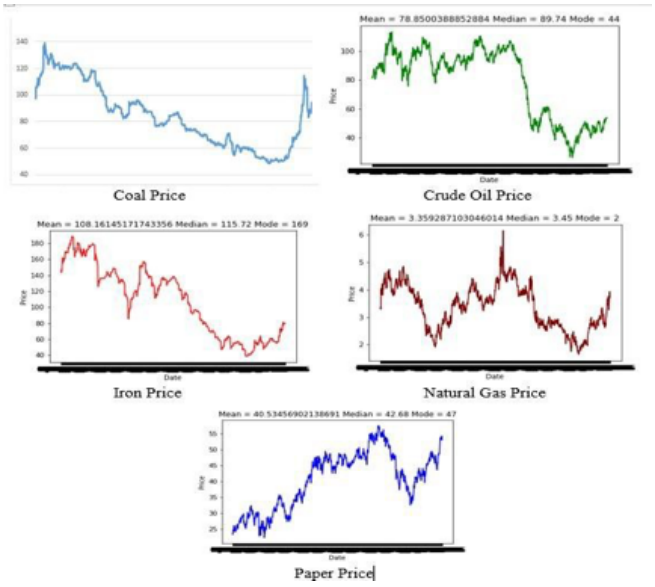


Figure 1 The actual data of coal price and other impacting factors from October 22, 2010, to December 31, 2016

The null hypothesis that a normal distribution exists is rejected because the p-value for the Jarque-Bera test is similarly less than 0. 001. The first differences series' kurtosis surpasses a normal distribution's kurtosis (i.e., 3) and is leptokurtic, even if the price series' skewness and first differences series' skewness are both very close to a normal distribution's skewness (i.e., 0).

This study also considers the Shapiro-Wilk, Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling tests for the price series and its first differences series. The p-values of all of these tests are found to be significantly less than the cutoff of 0.001, which disproves the null hypothesis of a normal distribution.

Table 1 Summary statistics: Skewness, Kurtosis, and Jarque-Bera p-value for all the data series

Variables	Position	Skewness	Kurtosis	Jarque-Bera p-value
Coal	At level	-1.2887	4.5221	0.001
	At I difference	-1.7697	21.4107	0.001
Natural Gas	At level	-1.2987	4.4221	0.001
	At I difference	-1.6891	20.5012	0.001
Iron	At level	-1.1891	3.9998	0.001
	At I difference	-0.7561	19.2356	0.001
Paper	At level	0.9987	3.8989	0.001
	At I difference	-1.6563	19.1256	0.001

RESULTS AND DISCUSSIONS

Correlation of Coal Prices with the Selected Variables

The correlation between the coal price and the other factors we chose was first verified. As anticipated, we have seen that the price of coal rises together with the price of crude oil, natural gas, and iron (Figure 2).

To measure the observed qualitative correlations between coal and the chosen other factors, we generated Pearson cross-correlation coefficients. The Pearson cross-correlation coefficient, which ranges from -1 to +1, measures the degree of linear dependency and linear correlation between two variables (Pearson 1895). Pearson's coefficient of correlation is +1 for all positive correlations, 0 for no correlation, and 1 for all negative correlations. Pearson's correlation coefficient (Pearson 1895) is the covariance of two variables divided by the sum of their standard deviations.

The correlation coefficient for the coal price ρ_c, and iron price ρ_I as define as:

$$\rho\,(C,\,I) = \frac{cov\,(P_C,\,P_I)}{\sigma P_C \cdot\, \sigma P_I} \qquad (2)$$

Where σP_C and σP_I are the standard deviations of the coal and iron prices, respectively, and the covariance between them was:

$$COV_{P_C - P_I} = \frac{\Sigma (P_{C_i} - P_I)(P_{I_i} - P_I)}{N - 1} \quad (3)$$

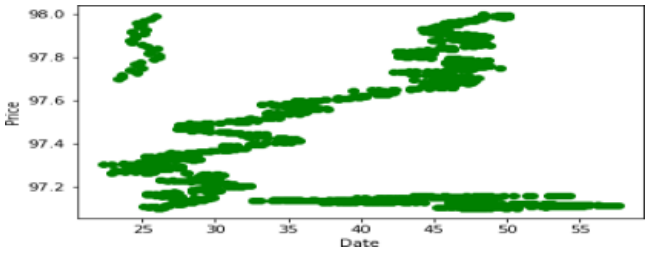


Figure 2 Scatter plots of coal prices and other factors such as crude oil, iron, natural gas, and paper

We discovered that all of the positive Pearson cross-correlation coefficients between coal and the chosen additional variables ranged from 0.01 to 0.7. The price of coal is substantially associated with the prices of crude oil, natural gas, and iron, as shown by Pearson's correlation values of 0.64, 0.47, and 0.41, respectively (Figure 3). These strong positive relationships can be simply explained by the fact that coal is used to produce iron and is dependent on energy sources like crude oil.

The correlation between the coal price and the price of natural gas (0.47) is not as high as the correlation between coal price and crude oil (0.64). The low correlation coefficient for natural gas would suggest that it is less crucial than crude oil in the processing of coal. We also found that there is virtually little association between the price of coal and the price of paper (0.012). The fundamental cause is outside the purview of this article.

Coal Price Prediction Using Decision Tree Learning

A popular technique in data mining is decision tree regression. A decision tree algorithm builds a model by a recursive partitioning procedure by dividing a training data set, X = X1, X2..., XN, into Y = Y1, Y2..., YN. The 'Coal Price' column, which is the label, is the only column from the dataset that isn't present in the X variable in this case. The X variable has the attribute set, and the y variable holds the labels that contains values from the 'Coal Price' column.

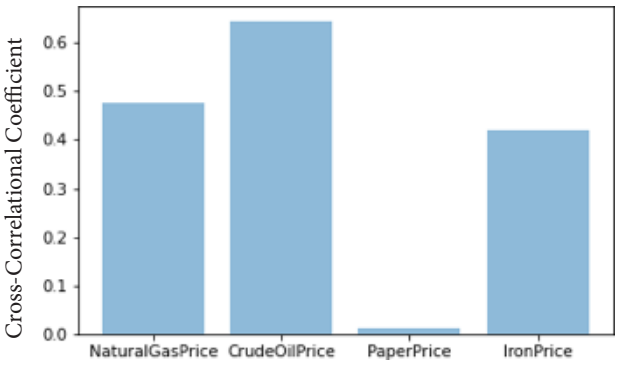


Figure 3 Pearson's coefficient of correlation between the price of coal and the other selected factors affecting the price of coal (crude oil, natural gas, iron, and paper)

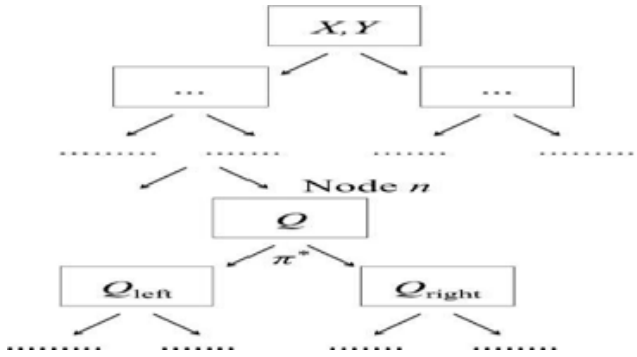


Figure 4 Schematic representation of a decision tree model

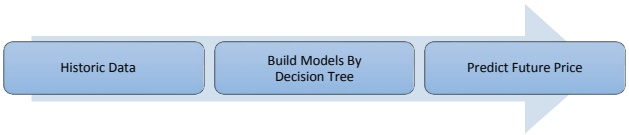


Figure 5 Illustration of the decision tree-based model

Training and Making Predictions

We created a training data set utilizing historical data for the purpose of training the decision-tree model for precise forecasting of future prices. A list of Explanatory variables (x_n) and a Dependent variable (y_n), where (y_n) is the price of coal on the ninth day, make up the training data set (counted from October 22, 2010).

The decision-tree model (Figure 4) is then trained using all of the historical data up to this point. The price of coal is then predicted using the learned model. The whole collection of 309 observations from October 22, 2010, to December 31, 2016, was used to create the training data set.

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{607} \end{bmatrix} \rightarrow \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{607} \end{bmatrix}$$

Where y_1 is the coal price, x_1 is the range of prices of all the other independent variables. The predicted values are compared with the actual values to check the accuracy.

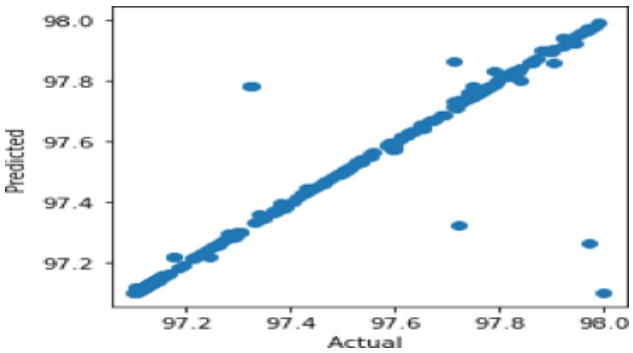


Figure 6 Actual price vs. predicted price of test data

Evaluating the Algorithm

MAE, MSE, and RMSE are often used metrics to assess the performance of the regression process. Function from the Scikit-Learn package have been used for calculating these values. We calculated the mean absolute error to validate the prediction errors of our methodology.

Mean Absolute Error:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - x| \quad (4)$$

Root Mean Square Error:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{N}} \quad (5)$$

Mean Square Error:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2 \quad (6)$$

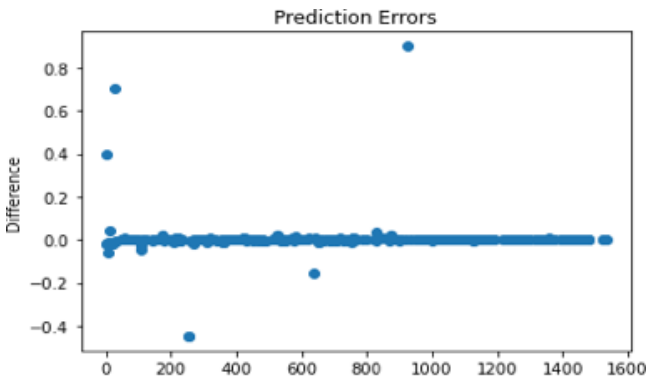


Figure 7 Difference between the actual and predicted values (i.e., the prediction errors)

CONCLUSION

Prediction of prices of thermal coal is a momentous topic for investors and policymakers considering its position as a strategic energy source (Xu & Zhang, 2022). In the present study, this forecasting problem is investigated in a dataset of daily prices of thermal coal prices. When the full-time series is used for the forecast, the decision tree model provides a more accurate prediction. In forecasting future coal prices, the decision tree method was applied. The prices of iron, crude oil, natural gas, paper, and previous coal prices was used as independent variables for predicting future coal prices. With MAPE below 5%, this decision tree-based technique is capable of forecasting coal prices both in the short and long term. The approach used was versatile and general in nature. However, additional factors and variables can be included in training the decision-tree model. In this study, four extraneous factors that influence coal prices were used to generate the independent variables. The decision tree-based method has several advantages over conventional regression models and other approaches. The decision tree-based method has several advantages over conventional regression models and other approaches. Firstly, there is no need for assumptions. Secondly, this strategy may be used to solve classification as well as regression problems. Thirdly, the model created using decision tree learning is simple to understand. Lastly, this approach is effective for examining the connection between the input data and the target variable and for determining which independent factors have the most impact on forecasting the target variable. Consequently, it is helpful to add to and augment conventional regression techniques. The longest forecast horizon in the current

study is 4 years (or 5-6 years when weekend dates are taken into consideration), subject to data availability. However, if historical data are provided, it may be easily extended to 10 or 20 years. Additionally, the present method's assumption-free nature reduces the risk brought on by human judgment and bias. Hence, this approach expected could be applied widely.

NOMENCLATURE

GRNN-Generalised Regression Neural Networks	MAE- Mean Absolute Error
MLFNMulti-Layer Feedforward Networks	MSE-Mean Square Error
MAPE -Mean Absolute Percentage Error	RMSE-Root Mean Square Error

REFERENCES

Alam, M. S., Murshed, M., Manigandan, P., Pachiyappan, D., & Abduvaxitovna, S. Z. (2023). Forecasting oil, coal, and natural gas prices in the pre-and post-COVID scenarios: Contextual evidence from India using time series forecasting tools. *Resources Policy*, *81*, 103342. https://doi.org/10.1016/j.resourpol.2023.103342 .

Altan, A., Karasu, S., & Zio, E. (2021). A new hybrid model for wind speed forecasting combining long short-term memory neural network, decomposition meth-ods, and grey wolf optimizer. *Applied Soft Computing*, *100*,106996. https://doi.org/10.1016/j.asoc.2020.106996

Alameer, Z., Fathalla, A., Li, K., Ye, H., & Jianhua, Z. (2020). Multistep-ahead forecasting of coal prices using a hybrid deep learning model. *Resources Policy*, *65*, 101588.

Bonita, O., & Muflikhah, L. (2018, November). Comparison of Gaussian and ANOVA Kernel in Support Vector Regression for Predicting Coal Price. In *2018 International Conference on Sustainable Information Engineering and Technology (SIET)* (pp. 147- 150). IEEE.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., (1984). *Classification and Regression Trees*.

Ballings M., Van den Poel D., Hespeels N., Gryp R., (2015) Evaluating multiple classifiers for stock price direction prediction. *Expert Syst Appl*, *42*,7046–7056.

Chandrarin, G., Sohag, K., Cahyaningsih, D. S., Yuniawan, D., & Herdhayinta, H. (2022). The response of exchange rate to coal price, palm oil price, and inflation in Indonesia: Tail dependence analysis. *Resources Policy*, *77*, 102750.

Ding, L., Zhao, Z., & Han, M. (2021). Probability density forecasts for steam coal prices in China: The role of high-frequency factors. *Energy*, *220*, 119758.

Fan, G.-Z., Ong, S.E., Koh, H.C., (2006). Determinants of house price: a decision tree approach. *Urban Stud*. *43*(12), 2301–2315.

Fan JH, Todorova N (2017) Dynamics of China's carbon prices in the pilot trading phase. *Appl Energy*, *208*, 1452–1467.

Ferrari, D., Ravazzolo, F., & Vespignani, J. (2021). Forecasting energy commodity prices: A large global dataset sparse approach. *Energy Economics*, *98*, 105268.

Gargano, A., Timmermann, A., (2014). Forecasting commodity price indexes using macroeconomic and financial predictors. *Int. J. Forecast*. 30 (3), 825–843.

Gao, L., & Wang, G. (2021). Research on the improved neural network of coal price forecast based on co-integration theory. *IOP conference series: Earth and environmental science, 769* (p. 042028). IOP Publishing. https://doi.org/10.1088/1755- 1315/769/4/042028

Gumus, M., Kiran M.S., (2017) Crude oil price forecasting using XGBoost. In: *2017 International Conference on Computer Science and Engineering (UBMK)*. IEEE, pp 1100–1103

Hintermann, B. (2010) Allowance price drivers in the first phase of the EU ETS. J Environ Econ Manage, *59*, 43–56

Keppler, J.H., Mansanet-Bataller M., (2010) Causalities between CO2, electricity, and other energy variables during phase I and phase II of the EU ETS. *Energy Policy*, 38, 3329-3341.

Kong, Z., Dong, X., & Jiang, Q. (2019). Forecasting the development of China's coal-to-liquid industry under security, economic, and environmental constraints. *Energy Economics*, *80*, 253–266.

Liu, Z.Y., & Sun, Z.D. (2017). The carbon trading price and trading volume forecast in Shanghai City by BP neural network, Urban Civil Eng., *11*(3), 595–601

Li, Z.-Z., Su, C.-W., Chang, T., & Lobonţ, O.-R. (2022). Policy-driven or market-driven? Evidence from steam coal price bubbles in China. Resources Policy, 78, 102878.

Loh, W. (2014). Fifty Years of Classification and Regression Trees. *International Statistical Review*, *82*(3), 329-348.

Matyjaszek, M., Fernández, P. R., Krzemie N, A., Wodarski, K., & Valverde, G. F. (2020). Forecasting coking coal prices by means of Arima models and neural networks, considering the transgenic series theory. *Resources Policy, 61*, 283–292. https://doi.org/10.1016/j.resourpol.2019.02.017

Mei, J., He, D., Harley, R., Habetler, T., Qu, G. (2014). A random forest method for real-time price forecasting in the New York electricity market. In *2014 IEEE Power and Enegery Society General Meeting Conference.*

Matyjaszek, M., Riesgo Fernández, P., Krzemień, A., Wodarski, K., & Fidalgo Valverde, G. (2019). Forecasting coking coal prices by means of Arima models and neural networks, considering the transgenic time series theory. Resources Policy, *61*, 283– 292.

Pearson, K., (1895) Notes on regression and inheritance in the case of two parents. *In: Proceedings of the Royal Society of London*, *58*, 240–242.

Quinlan, J.R. (1992). *Learning with Continuous Classes*. Proceedings of Australian Joint Conference on Artificial Intelligence, Hobart, 343-348.

Seifert, J., Uhrig-Homburg, M., Wagner, M. (2008) Dynamic behavior of CO_2 spot prices. *J Environ Econ Manage*, 56,180–194

Tsai, M.T., Kuo, Y.T.A (2013) Forecasting System of Carbon Price in the Carbon Trading Markets Using Artificial Neural Network. *International Journal of Environmental Science & Development*, *4*(2),163–167.

Wang, N. (2017), Economics S O, University X. Forecasting of Carbon Price Based on Boosting-ARMA Model.

Wang, X., Liu, C., Chen, S., Chen, L., Li, K., & Liu, N. (2020). Impact of coal sector's de-capacity policy on coal price. *Applied Energy*, *265*, 114802.

Wei, Y.M., Wang, K., Feng, Z.H., Cong, R.G. (2010). *Carbon Finance and Carbon Market: Models and Empirical Analysis,* Science Press, Beijing.

Wei, S., Chongchong, Z., & Cuiping, S. (2018). Carbon pricing prediction based on wavelet transform and K-ELM optimized by bat optimization algorithm in China ETS: The case of Shanghai and Hubei carbon markets. *Carbon Management,* 9(6), 605-617

World Coal Institute. (2005, May) Sumber Daya Batu Bara: Tinjauan lengkap mengenai batu bara. https://www.worldcoal.org

Wu, S., Xia, G., & Liu, L. (2023). A novel decomposition integration model for power coal price forecasting. *Resources Policy*, *80*, 103259. https://doi.org/10.1016/j.resourpol.2022.103259

Xu, X., & Zhang, Y. (2022a). *Thermal coal price forecasting via the neural network*. Intelligent Systems with Applications, 14, 200084. https://doi.org/10.1016/j.iswa.2022.200084

Xu, X., & Zhang, Y. (2022b). *Network analysis of housing price comovements of a hundred Chinese cities*. National Institute Economic Review. https://doi.org/101017/nie.2021.34.

Zhang, F., & Wen, N. (2022). Carbon price forecasting: a novel deep learning approach. *Environmental Science and Pollution Research*, 1-14.

Zhang, K., Cao, H., Thé, J., & Yu, H. (2022). A hybrid model for multi-step coal price forecasting using decomposition technique and deep learning algorithms. *Applied Energy*, 306, 118011. https://doi.org/10.1016/j.apenergy.2021.118011

Zhao, G., Che, K., & Zhao, Y. (2012). A study on predicting coal market price in China based on time sequence models. *International Journal of Business and Social Science*, *3*(4).

Zhao, M., Li, L., & Xu, Z. (2021). Study on hotel selection method based on integrating online ratings and reviews from multi-websites. *Information Sciences*, *572*, 460–481. https://doi.org/10.1016/j.ins.2021.05.042

Zhu, B., Han, D., Wang, P., Wu, Z., Zhang, T., & Wei, Y.-M. (2017). Forecasting carbon price using empirical mode decomposition and evolutionary least squares support vector regression. Applied Energy, *191*, 521–530