E-ISSN 2586-9027
Homepage : *https://tci-thaijo.org/index.php/SciTechAsia* Science & Technology Asia
Vol.30 No.4 October - December 2025 Page: [82-89]
Original research article

Prediction of Indonesian Coal Prices Using Support Vector Regression Method with Grid Search Optimization

Joana Kriskinantyas Rahayu, Nariza Wanti Wulan Sari*, Desi Yuniarti, Siti Mahmuda, M. Fathurahman

Statistics Program Study, Faculty of Mathematics and Natural Sciences, Mulawarman University, Samarinda 75119, Indonesia

Received 17 May 2025; Received in revised form 29 July 2025
Accepted 5 September 2025; Available online 17 December 2025

ABSTRACT

Coal prices are highly volatile due to fluctuations in global supply and demand, government policies, and economic and political factors. Therefore, accurate price prediction is essential and can be achieved using Support Vector Regression (SVR). This study aimed to identify the most accurate SVR model for forecasting Indonesian coal prices and analyze its prediction performance. The Radial Basis Function (RBF) kernel was used with hyperparameter ranges of $10^{-3} \le C \le 10^2$, $10^{-3} \le \gamma \le 10^2$, and $\varepsilon = 0.01, 0.02, 0.03$, tested at data splits of 70:30, 80:20, and 90:10. Optimal hyperparameters were determined using the grid search algorithm, and model performance was evaluated using the Mean Absolute Percentage Error (MAPE). Results showed that the best prediction accuracy was achieved with a MAPE of 5.459%, indicating excellent model performance. The optimal configuration was the SVR with RBF kernel at a 90:10 ratio, where $C = 12.5$, $\gamma = 0.1$, and $\varepsilon = 0.025$. The resulting model effectively approximated the actual coal price data, confirming its reliability for forecasting Indonesian coal prices.

Keywords: Coal price; Grid search; Machine learning; Prediction; Support vector regression

1. Introduction

Along with the development of advanced technology, the term Artificial Intelligence (AI) is increasingly common. AI is a field of computer science that focuses on the automation of intelligent behavior through knowledge representation data structures, knowledge implementation

***Corresponding author:** nariza@fmipa.unmul.ac.id **doi:** 10.14456/scitechasia.2025.71

algorithms, and related programming languages and techniques [1]. There are several AI models such as Expert System (ES), Fuzzy Inference System (FIS), Deep Learning (DL), and Machine Learning (ML) now commonly used [2]. Researchers have begun to use ML as an important method for analyzing various relationships between variables, with the main goal of predicting the future and generating new knowledge [3].

Machine Learning (ML) is a computational method that automatically trains a system from input data to achieve a goal without explicit programming. ML algorithms operate like “soft code”, by adapting through experience to improve performance and generalization to new data [4]. There are three ML training methods, i.e., supervised learning, unsupervised learning, and reinforcement learning.

Supervised learning methods operate by providing the input data and output labels to be learned by the model. One of the most commonly used models in supervised learning is the Support Vector Machine (SVM) [2]. SVM operates by transforming the training data through a nonlinear mapping into a higher dimensional space. In this new feature space, the SVM identifies an optimal linear separating hyperplane. The optimality of the hyperplane is achieved through the utilization of support vectors and maximal margins [5].

Support Vector Regression (SVR) is an adaptation of SVM for the case of regression, which excels at capturing non-linear data trends and overcoming overfitting [2]. The performance of SVR is highly dependent on hyperparameter selection. The idea of automatic hyperparameter optimization, such as using a grid search algorithm with cross-validation is important to find the best combination of parameters efficiently [6]. Previous research has shown the success of SVR optimization using grid search as in research conducted by Purwoko, et al [7]. The research compared Fourier series and SVR with grid search optimization in predicting Indonesian non-oil and gas export prices, showing that the SVR model with grid search produced more accurate predictions. This is proven by the Mean Absolute Percentage Error (MAPE) value of SVR of 9.29% which is lower than the MAPE value of the Fourier series model 15.26%.

Indonesia’s non-oil and gas exports have great potential, especially coal where the price is influenced by various economic factors and global policies [8]. Coal is a crucial element in the operations of large industrial companies, so regular purchases from trusted producers are a must to maintain a quality supply [9]. According to [10] and [11] coal prices are affected by various factors such as supply and demand, transportation costs, quality, government policies, and crude oil prices, as well as other factors such as importing countries’ electricity needs and global mining policies. Predicting coal prices is important for industries to plan their budgets accordingly. The prediction of coal prices in this study was based on various economic factors that may affect coal prices such as natural gas prices, crude oil prices, export-import values, rupiah exchange rates against the dollar, and inflation.

Based on the previous explanation, this research aims to identify the best prediction accuracy that can be achieved by the SVR model in predicting Indonesian coal prices, as well as knowing the results of predicting Indonesian coal prices based on the best SVR model.

2. Methods

2.1 Research design and research variables

This research used ex post facto comparative causal design and was non-experimental research because the data collected was not through direct research. The data collection technique used in this research is purposive sampling. Researchers considered used coal price data because it is very rare for research to use coal price data as a case study that needed to researched; by predicting Indonesian coal prices it is hoped that it could help industries in determining the right budget.

The data used is secondary data from Indonesian coal price data from January 2014 to December 2024 obtained from the official website of the Ministry of Energy and Mineral Resources which was accessed through the page www.minerba.esdm.went.id. This study was conducted by recapitulating Indonesian coal price data (Y), historical natural gas futures prices (X_1), historical West Texas Intermediate (WTI) crude oil futures prices (X_2), Indonesian exports values (X_3), Indonesian import values (X_4), dollar exchange rates against the rupiah (X_5), and Indonesian inflation data (X_6) from January 2014 to December 2024.

2.2 Data analysis technique

The analytical stages in this research are as follows:

1. Prepare data of coal prices, historical natural gas futures prices, historical WTI crude oil futures prices, exports values, import values, dollar exchange rates against the rupiah, and Indonesian inflation from January 2014 to December 2024. The total data used is 132 data for each variable.

2. Analyze descriptive statistics used R software based on the variables studied.

3. Divide the data into training and testing data with the proportions used was 70:30, 80:20 and 90:10.

4. Normalize data using the z-score normalization method used Eq. (2.1).

$$v^* = \frac{v - \bar{A}}{\sigma_A}, \tag{2.1}$$

   where v is the observation object, $\bar{A}$ is the mean and σ_A is the standard deviation of the variable A.

5. Determine the kernel function to use in the research.

   Kernel methods generate algorithms by replacing the inner product with a suitable positive definite function, which implicitly performs a nonlinear mapping of the training data into a high-dimensional input space [12]. The kernel function that will be used in this study is the Radial Basis Function (RBF) in Eq. (2.2).

$$k(x, x') = \exp(-\gamma \|x - x'\|)^2. \tag{2.2}$$

6. Optimize the hyperparameters C, γ, and ε using the grid search algorithm for each proportion used.

   The grid search optimization method tests each combination of hyperparameter values in turn and then compares the smallest error value resulting from the combination of hyperparameter values [13]. Grid search is not able to utilize zones that perform well directly. Therefore, certain procedures must be performed manually to determine the optimal hyperparameter values. This procedure starts with a large search space, and then shortens the search space based on the previous well-performing hyperparameter configuration results; these configuration steps are repeated several times to achieve optimal results [14].

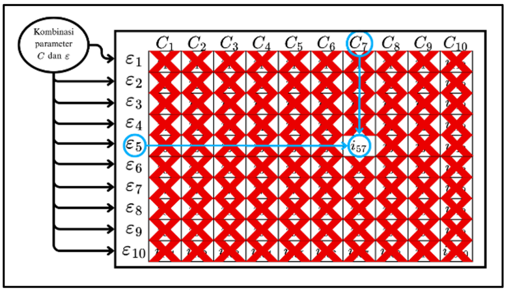

Fig. 1. Illustration of the grid search algorithm

These algorithms need to be approached with some performance metrics and generally through cross validation. The process of cross validation is done by dividing the dataset into subsets known as folds [15].

7. Construct an SVR model with optimal hyperparameters for each proportion used. Assume there is n data (x_i, y_i) with $i = 1, 2, , n$ then $x = [x_1 \quad x_2 \cdots x_n]$ is a vector of input variables in the input space, where $x \in R^d . y = [y_1 \quad y_2 \cdots y_n]$ is a vector of output variable values where $y \in R$. The continuous function for one-dimensional linear approximation can be written in Eq. (2.3) [16].

$$f(x) = w^T x + b, \qquad (2.3)$$

where $w = [w_1 \quad w_2 \cdots w_n]^T$ is the $1 \times n$ weight vector for the optimal hyperplane in the input space which can be written as some linear combination of support vectors and b is the bias. SVR will find the function $f(x)$ by minimizing the coefficient w using convex optimization in Eq. (2.4) [17].

$$\min_{w} \frac{1}{2}\|w\|^2, \qquad (2.4)$$

with constraints

$$\begin{cases} y_i - f(x) \leq \varepsilon \\ f(x) - y)i \leq \varepsilon \end{cases}.$$

The function $f(x)$ is feasible if the data points of x are within the range $|y_i - f(x) \leq \varepsilon|$, otherwise it is infeasible. This infeasible condition can be solved by including slack variables ξ_i, ξ_i^* in order to overcome the infeasible margin problem in the optimization problem, so that Eq. (2.4) can be rewritten as Eq. (2.5) [18].

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*), \quad (2.5)$$

with constraints

$$\begin{cases} y_i - f(x) \leq \varepsilon + \xi_i \\ f(x) - y)i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}.$$

by solving the optimization problem using the Lagrange function then the new weight vector will be obtained as in Eq. (2.6) [18].

$$w = \sum_{i=1}^{n}(a_i - a_i^*)x_i. \qquad (2.6)$$

Next, the data inside x is mapped through a function $\Phi(x)$ which is approximated by the $k(x, x')$ function equation. The SVR function can then be redefined as in Eq. (2.7) [16].

$$f(x) = \sum_{i=1}^{n}(a_i - a_i^*)k(x, x') + b. \tag{2.7}$$

The prediction model is constructed using Eq. (2.7).

8. Implement the prediction model that was obtained on the tested data for each proportion used.

9. Denormalize the predicted data to returned the predicted valued to the actual data scale.

10. Determine the best model by evaluating the results of the best accuracy value.

    There are several measurements that can be used to find the best accuracy value, one of which is Mean Absolute Percentage Error (MAPE). MAPE is calculated by using the absolute error in each time period divided by the real observed value for that period. Then, calculate the average absolute percentage error. MAPE is formulated in the Eq. (2.8) [19].

$$MAPE = \frac{1}{2}\sum_{i=1}^{n}|\frac{y_i - \hat{y}_i}{y_i}| \times 100, \tag{2.8}$$

    with y_i is an actual object of observation and $\hat{y}_i$ is the prediction of the observation object.

11. Plot the comparison of the actual data with the best predicted data that has been obtained.

3. Results and Discussion

3.1 Analysis of descriptive statistics

Descriptive statistics basically function to display data characteristics so that a general description of the variables used as research data can be obtained. The dependent variable in this study is the price of Indonesian coal (Y). The independent variables are historical price of natural gas futures (X_1), the historical price of WTI crude oil futures (X_2), total Indonesian exports (X_3), total Indonesian imports (X_4), the dollar exchange rate against the rupiah (X_5) and the value of Indonesian inflation (X_6).

Table 1. Descriptive statistical analysis of data.

Variable	Mean	Standard of Deviation	Minimum	Maximum
Y	112.382	71.924	49.420	330.970
X_1	3.677	1.710	1.864	10.028
X_2	69.459	20.290	29.130	130.500
X_3	16652.930	4399.142	9649.500	27928.700
X_4	15180.210	3319.113	8438.627	22150.550
X_5	14134.500	1117.430	11404.000	16421.000
X_6	0.037	0.017	0.013	0.084

During the period 2014 - 2024, coal prices in Indonesia experienced significant fluctuations, with a low of \$49.420 in September 2020 and a peak of \$330.970 in October 2022. With an average of \$112.382 and a data spread of \$71.924, coal prices show high volatility. In contrast, natural gas prices were more stable with a range of \$1,864/MMBtu in May 2020 to \$10,028/MMBtu in July 2022, as well as an average of \$3,677/MMBtu and a standard deviation of \$1,710/MMBtu, indicating more restrained movements. WTI crude oil futures prices also experienced considerable variation, from \$29.130/barrel in October 2018 to \$130.500/barrel in November 2016, with an average of \$69.459 and standard deviation of \$20.290/barrel, indicating instability in energy prices.

In addition, Indonesia's export

value ranged from \$9,649.500 million in July 2016 to \$27,928.700 million in August 2022, while import value moved from \$8,438.627 million in May 2022 to \$22,150.550 million in August 2022. The relatively restrained spread of export and import values indicates the stability of trade on a large scale. However, the rupiah exchange rate against the US dollar has weakened significantly over the past 11 years, from IDR11,404.000 in March 2014 to IDR16,421.000 in June 2024, with an average of IDR14,134.500 and standard deviation of IDR1,117.430, reflecting the downward trend of the local currency against the US dollar.

Inflation in Indonesia has remained in a relatively stable range, with a low of 0.013 in December 2014 and a high of 0.082 in June 2021. Average inflation stands at 0.037 with a standard deviation of 0.017, indicating a less extreme pattern of change.

3.2 Coal price prediction using support vector regression

The first step of this process is proportion, which divides the sample data into 70:30, 80:20 and 90:10. After dividing the proportion, the normalization process is carried out using Eq. (2.10) on each proportion. Next is to find the optimal hyperparameter value using the grid search algorithm. The parameter values used in this study were $10^{-3} \le C \le 10^{2}, 10^{-3} \le \gamma \le 10^{2}$, and $\varepsilon = 0.01, 0.02, 0.03$. This process is carried out using the help of *R* software, and the results obtained are shown in Table 2.

Table 2. Optimal hyperparameter values.

Proportion	C	γ	ε
70:30	11.500	0.050	0.035
80:20	12.500	0.150	0.035
90:10	12.500	0.100	0.025

The model was built using the optimal hyperparameter value. Model building is done using the svm function in *R* software. The output in the software will show the number of support vectors and the *b* values for each data proportion as in Table 3.

Table 3. Result of support vectors and *b* values.

Proportion	Support Vectors	b
70:30	89	0.147
80:20	86	0.177
90:10	102	0.057

Based on the output results, an SVR model with RBF kernel can be formed for each data proportion as in Eq. (3.1) until Eq. (3.3).

1. Support Vector Regression Model with 70:30 Proportion.

$$f(x) = \sum_{i=1}^{89}(a_i - a_i^*)k(x, x') + 0.147. \tag{3.1}$$

2. Support Vector Regression Model with 80:20 Proportion.

$$f(x) = \sum_{i=1}^{86}(a_i - a_i^*)k(x, x') - 0.177. \tag{3.2}$$

3. Support Vector Regression Model with 90:10 Proportion.

$$f(x) = \sum_{i=1}^{102}(a_i - a_i^*)k(x, x') + 0.057. \tag{3.3}$$

3.3 Goodness of prediction measure

Based on the prediction results that have been obtained, the best model selection is carried out using the MAPE value

Table 4. Goodness measure of coal price prediction model.

Proportion of Training and Testing Data	MAPE Training	MAPE Testing
Proportion 70:30	6.367%	19.493%
Proportion 80:20	4.661%	36.650%
Proportion 90:10	5.459%	5.983%

in Eq. (2.8). The calculation of the MAPE value is done using the help of R software which can be seen in Table 3.

Table 3 indicates that the Support Vector Regression (SVR) model with Radial Basis Function (RBF) kernel at 80:20 data proportion has the lowest Mean Absolute Percentage Error (MAPE) of training, but the highest MAPE of testing, indicating overfitting due to overexposure of the model to the training data. In contrast, the SVR-RBF model with a 90:10 data proportion has a training MAPE of 5.459% and a testing MAPE of 5.983%, both below 10%, which indicates the model does not experience overfitting and has high prediction accuracy on both training and testing data.

3.4 Plot comparison of actual data with predicted results

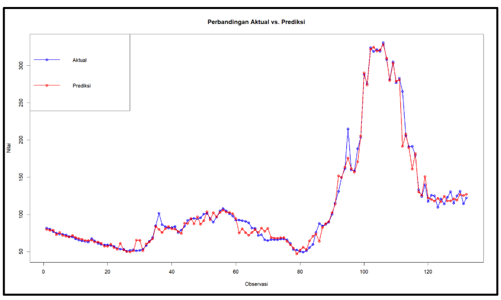

Fig. 2. Comparison plot of actual data with predicted data.

The plot in Fig. 2 shows that the prediction data pattern follows the movement of the actual data pattern, which means that the SVR model with RBF kernel at a proportion of 90:10 is very suitable for predicting Indonesian coal prices.

3.5 Discussion

The SVR model with the RBF kernel function at an 80:20 data proportion had the largest MAPE testing but the smallest training MAPE. This phenomenon suggests that the model is overfitting, a condition in which the model performs exceptionally well on training data but has a reduced capacity to generalize to testing or new data. When applied to data that has never been seen before, overfitting occurs because the model overfits itself to the patterns in the training data, lowering prediction accuracy. Alternative hyperparameter tuning techniques like Bayesian optimization or genetic algorithms could be investigated to lessen this problem. The selection of features, where variables like coal stock levels, weather, or energy policies could improve prediction capabilities, is another factor affecting prediction accuracy in addition to overfitting.

4. Conclusion

Based on the results of data analysis and discussion, it can be concluded that the SVR model with RBF kernel and 90:10 training-test data division was the best model in this studiy. This is supported by the MAPE valued of 5.459%, which indicates excellent prediction accuracy. The optimal hyperparameter configuration for this model is C=12.5; γ=0.1; and ε=0.025. Based on the best model, it was known that the prediction results obtained approximated the actual data valued.

References

[1] Luger GF. Artificial Intelligence: Structures and Strategies for Complex Problem Solving. 6th ed. United States of America: Pearson Addison Wesley; 2009.

[2] Elwirehardja GN, Suparyanto T, Pardamean B. Pengenalan Konsep Machine Learning Untuk Pemula. 1st ed. Yogyakarta: Instiper Press; 2023.

[3] Putra JWG. Pengenalan Konsep Pembelajaran Mesin dan Deep Learning. Edisi 1.4. 2020.

[4] El Naqa I, Li R, Murphy JM. Machine Learning in Radiation Oncology: Theory and Applications. 1st ed. Switzerland: Springer International Publishing; 2015.

[5] Han J, Pei J, Tong H. Data Mining: Concepts and Techniques. 4th ed. United States of America: Morgan Kaufmann Publisher; 2023.

[6] Ngoc TT, Van Dai L, Thuyen CM. Support Vector Regression based on Grid Search method of Hyperparameters for Load Forecasting. 2021;18.

[7] Purwoko CFF, Sediono S, Saifudin T, Mardianto MFF. Prediksi Harga Ekspor Non Migas di Indonesia Berdasarkan Metode Estimator Deret Fourier dan Support Vector Regression. Inferensi. 2023;6:45–55.

[8] Umiyati A, Dasari D, Agustina F. Peramalan Harga Batubara Acuan Menggunakan Metode PSOSVR Dan IPSOSVR. Jurnal EurekaMatika. 2021;9:69–94.

[9] Almas MF, Setiawan BD, Sutrisno. Implementasi Metode Backpropagation untuk Prediksi Harga Batu Bara. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. 2018;2:6502–11.

[10] Indonesian Coal Suppliers Association (ICSA). Memahami Harga Batu Bara Indonesia: Faktor yang Mempengaruhinya. Jakarta: ICSA; 2023.

[11] CNN Indonesia. 5 Faktor Harga Batu Bara Terus Melonjak Versi Eks Wamen ESDM. 2022.

[12] Du KL, Swamy MNS. Neural Networks and Statistical Learning. 2nd ed. United Kingdom: Springer; 2019.

[13] Salsabilla A, Satyahadewi N, Andani W. Implementation Grid Search of RBF and Polynomial on Support Vector Regression for Closing Stock Prices Prediction on PT Indofarma (INAF). VARIANCE: Journal of Statistics and Its Applications. 2024;6:133–42.

[14] Hutter F, Kotthoff L, Vanschoren J. Automated Machine Learning: Methods, Systems, Challenges. Switzerland: Springer Nature Switzerland AG; 2020.

[15] Mulyawan, Subagja R, Rohman D, Efendi DI. Implementasi Support Vector Regression untuk Prediksi Harga Rumah Dengan Optimasi Grid Search. Jurnal Elektronik Ilmu Komputer Udayana. 2024;12:687–96.

[16] Awad M, Khanna R. Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers. New York: Apress Media; 2015.

[17] Prakoso BH. Implementasi Support Vector Regression pada Prediksi Inflasi Indeks Harga Konsumen. MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer. 2019;19:155–62.

[18] Smola AJ, Schölkopf B. A Tutorial on Support Vector Regression. Statistics and Computing. 2004;14.

[19] Krislianti DE, Zukhronah E, Susanti Y. Peramalan Harga Minyak Menggunakan Autoregressive Integrated Moving Average dan Support Vector Regression. Prosiding Seminar Pendidikan Matematika dan Matematika. 2023;7.