Application of the Support Vector Regression Method with the Grid Search Algorithm to Predict Movement Gold Price

Reni Puspita*, Hendra Cipta, Rima Aprilia

Mathematic Study Program, Faculty of Science and Technology, North Sumatra State Islamic University, North Sumatra,

Indonesia

*E-mail: rpuspita462@gmail.com

Received: Januari 28, 2024. Accepted: March 5, 2024. Published: March 30, 2024

Abstract: Gold is an investment with the smallest risk because it can be sold anytime and anywhere. In Indonesia, gold bullion as an investment product is known for its purity level of 99.99%, namely gold bullion produced by PT. Aneka Tambang (Antam) through its Precious Metals business unit. Apart from its pure production, Antam gold bullion is easier to resell anytime and anywhere because it has an official certificate from the international gold standardisation institution, namely LBMA (London Bullion Market Association), to more easily estimate the value of gold bullion when sold. To overcome this, predictions of future gold prices are needed. In this research, one of the prediction methods is Support Vector Regression with the Grid Search Algorithm. In this method this method will be used to predict the price of gold, which aims to predict and find out the price of gold one year in the future to produce a level accuracy (MAPE) of 5.43% and the prediction of gold prices increasing from 2023-June-01 to 2024-March-23 while experiencing a decline starting in 2024-March-24. Research by examining the relationship between variables, which emphasises data consisting of numbers so that it is analysed based on statistical procedures using the Support Vector Regression method with data sourced from the daily price of gold bullion through PT. Gallery 24 Pawnshops, North Sumatra. Where this method is very well used in predicting by choosing the best kernel used is the linear kernel because, from these three kernels, the best hyperparameters were obtained for predicting gold price movements using a linear kernel with a division for training and testing data of 60: 40. The MAPE value obtained was 5.43.

Keywords: Gold Price; Prediction; Support Vector Regression.

Introduction

The capital market is a source of financing needed by business actors who need additional capital and as alternative financing for the investing community [1]. According to [2], investment is an activity that manages valuable assets. Meanwhile, [3] investment is a person's (or investor's willingness to allocate money in a certain amount in the present to obtain future income. The investment that has become the favourite because it is the safest investment is gold (precious metals), including gold, silver and platinum.

The risk of investing in gold prices is called a time series data investment type because the price of gold will increase, decrease or remain constant every day [4]. Therefore, gold price predictions are needed to avoid risks with accurate prediction techniques. Using the time series model is expected to produce excellent and optimal predictions because the characteristics of gold price data are time series data that move according to specific times [5]. The method to predict time series data on gold price movements can be used is Support Vector Regression (SVR) with the Gird Search Algorithm [6].

SVR is a method that can produce good performance because it can overcome the overfitting problem. Overfitting is the behaviour of testing or training data that makes almost perfect prediction accuracy. The concept of SVR maximises the hyperplane to obtain Support Vector data [7]. SVR uses kernel functions in its work process. The kernel function's selection and settings greatly influence the regression accuracy level [8]. The advantage of SVR is that it can overcome overfitting. One method that can be used to determine optimal model parameters is the Grid Search method [9]. This method can obtain optimal SVR parameters [10]. In the Grid Search Algorithm application, there is a performance guide. This algorithm divides the range of optimised parameters into a grid (pairs of parameters). It passes through all points to obtain optimal parameters, usually measured by cross-validation on training data [11].

Predictions require taking historical data and projecting it into the future. Prediction is a way of systematically and pragmatically estimating what will happen in the future based on relevant information from the past [12]. The determination of the projection time depends on the actual situation, circumstances, and projected goals. Commonly used periods are days, weeks, months, half a year and a year. The further the prediction is made, the less accurate the results are [13].

Support Vector Regression (SVR) is an application of the Support Vector Machine (SVM) introduced by Cortes and Vapnik (1995) for regression cases. The purpose of Support Vector Regression (SVR) is to determine a function (x) as a hyperplane (dividing line) in the form of a regression function which fits all input data with an error [14]. The SVR output results are accurate and continuous numbers. Performance in SVR itself has three parameters that influence it: cost (C) or capacity, which presents a trade-off

How to Cite:

Puspita, R., Cipta, H., & Aprilia, R. (2024). Application of the Support Vector Regression Method with the Grid Search Algorithm to Predict Movement Gold Price. *Jurnal Pijar Mipa*, 19(2), 380–385. <u>https://doi.org/10.29303/jpm.v19i2.6607</u>

between model complexity and number, up to larger than acceptable deviations. Epsilon (ϵ), which controls the width of the sensitive zone, is used to adjust the training data. And gamma (γ) is a kernel function parameter [15].

Grid Search searches for the best hyperparameters by combining all the required parameters. This method achieved the best accuracy at the expense of the time required for searching [16]. The grid search method can find the best parameter values by providing a range of parameter values. This gird search divides the range of parameters to be optimised into girds and traverses all points to get the optimal parameters. The range that will be used in this research is $\varepsilon = (0.1, 0.2, ..., 1), C =$ $(2^{-5}, 2^{-4}, ..., 2^5) dan \gamma = (2^{-3}, 2^{-2}, ..., 2^8)$ When C and Y are an exponential sequence, Gird Search will train several pairs of models and finally select the pair with the smallest average error [17].

Normalisation is an approach to building a relational database logic design that is not directly related to the data model but applies several standard rules and criteria to produce a typical table structure [18]. In this research, the data preprocessing stage is data normalisation using the Min-Max Normalization method.

This research aims to determine the accuracy level of applying the Support Vector Regression method with the Grid Search Algorithm to predict Antam's gold price movements at PT. Gallery 24 Pegadaian North Sumatra Province and find out the results of Antam's gold price predictions at PT. Gallery 24 Pawnshops in North Sumatra Province using Support Vector Regression with the Gird Search Algorithm.

Research Methods

In this research, the author used quantitative methods. Research examining the relationship between variables emphasises that data consists of numbers, so it is analysed using the Support Vector Regression method based on statistical procedures. The type of data used in this research is secondary data, and the research data used is the daily price of gold bullion through PT. Gallery 24 Pawnshops, North Sumatra.

The variables used in this research are: Variable (y) = Gold price

Variable (x) = Gold price on the previous day

Research procedure

In the research method, research procedures are carried out, namely problem identification, which is the research about the movement of Antam gold prices at PT. Gallery of 24 Pawnshops in North Sumatra Province. Next, gold price data will be collected from 1 July 2020 to 31 May 2023. Then, descriptive analysis and analysis and processing of data consisting. Next, the variable is determined, followed by data normalisation.

 $x' = \frac{(x-a)}{b-a}$

Then, divide the data into training data and testing data with the percentage division in this research, namely 60%:40%, 70%:30%, 80%:20%, and 90%:10%., next determine the optimal parameter values for the gird search algorithm with the help of Software R e.1071 using the parameter range, namely. $(0.1,0.2, ... 1), C = (2^{-5}, 2^{-4}, ..., 2^5) dan \gamma = (2^{-3}, 2^{-2}, ..., 2^8)$ Then, the SVR kernel function selection Linear Kernel K = (x, y) = x. yPolynomial Kernel $K(x, y) = (x, y + C)^d$

Next, Kernel Radial Basis Function $(x, y) = exp(-\gamma ||x_i - x||^2)$

Then, an SVR model from the regression function will be created. Where to find the parameter values w and b using the R Package e.1071 Software. $f(x) = w.K\langle x_i, x_j \rangle + b$ Next, Evaluation of the SVR model into kernel functions, denormalising SVR Model Results with Kernel Functions

x' = y(max - min) + min

Next, Comparison of Accuracy Levels by evaluating the results of the best accuracy values using MAPE

$$MAPE = \frac{\sum_{t=1}^{n} \left| \frac{X_t - F_t}{X_t} \right|}{n} \times 100\%$$

Then, Predict one year into the future with the help of the Excel software provided by the system. And last, conclude (discussion).



Figure 1. Research Flow Diagram

Results and Discussion

Descriptive Analysis

The data used in this research is time series data regarding daily gold prices. Time series data is a type of data that is collected according to a specific time sequence. Before the data is processed using the SVR – Grid Search method, we process it first to maximise it when used. The data used in this research is the price of gold bullion from 2020-07-01 to 2023-05-31.

Based on Table 1, it can be seen that in this study, the amount of data used was 1065 data with fluctuating gold prices. A gold price data plot shows movements in gold bullion price data.

Based on Graph 2, it shows that the price of gold tends to increase in the period 03-07-2020 to 05-31-2023. Then, the price of gold tends to be stable and does not fluctuate too

much. However, gold prices tend to fluctuate upwards in 2020, making the difference in gold prices fluctuate widely and high. This is due to economic instability during the Covid-19 pandemic.

Table 1. Gold	price data at PT. Gallery	24 Pawnshops
No	daily time	Daily Price

110	ually time	Daily Thee
1	1-Jul-20	Rp730,000
2	2-Jul-20	Rp730,000
3	3-Jul-20	Rp730,000
4	4-Jul-20	Rp737,000
5	5-Jul-20	Rp730,000
6	6-Jul-20	Rp729,000
7	7-Jul-20	Rp728,000
8	8-Jul-20	Rp728,000
1061	27-May-23	Rp1,048,000
1062	28-May-23	Rp1,048,000
1063	29-May-23	Rp1,047,000
1064	30-May-23	Rp1,048,000
1065	31-May-23	Rp1,056,000



Figure 2. Gold Price Chart 03-07-2020 to 31-05-2023

Table 2. Descriptive gold prices	ptive gold prices	s
---	-------------------	---

Variable	Gold Price
Jumlah	1065
Min	Rp 726.000
Median	Rp 946.000
Mean	Rp 942.175
Max	Rp 1.096.000
Std. Dev	68053.4

Determination of Variables

In this research, gold price data uses the Support vector Regression method. Where this data is a time series forecast of current data which is influenced by past data (Lag time), and the target is current data so that the input variable (x) is needed using the previous day's gold price based on the past value of the time series [19]. Therefore, this research assumes that the gold price variable is influenced by the gold price on the previous day as variable (x). Meanwhile, this research's variable (y) is the current gold price. So you get something like the following table.

No	<i>(x)</i>	(y)
1		Rp730,000
2	Rp730,000	Rp730,000
3	Rp730,000	Rp730,000
4	Rp730,000	Rp737,000
5	Rp737,000	Rp730,000
6	Rp730,000	Rp729,000
7	Rp729,000	Rp728,000
8	Rp728,000	Rp730,000
1061	Rp1,043,000	Rp1,048,000
1062	Rp1,048,000	Rp1,048,000
1063	Rp1,048,000	Rp1,047,000
1064	Rp1,047,000	Rp1,048,000
1065	Rp1,048,000	Rp1,056,000

Data Normalisation

Because the data used has a high numerical scale, the data is first normalised. Data normalisation is a data processing process that aims to standardise data so that it is within a certain distance [20]. In Support Vector Regression analysis, normalisation is essential because the Support Vector Regression model will improve when data values become small-scale. Therefore, the y and x variables will be normalised.

Table 4. Normalisation Results for variables x and y

Tipe	Normalization (x)	Normalisation (y)
1		0.01081
2	0.01081	0.01081
3	0.01081	0.01081
4	0.01081	0.02973
5	0.02973	0.01081
6	0.01081	0.00811
7	0.00811	0.00541
8	0.00541	0.00541
1061	0.87027	0.85676
1062	0.85676	0.87027
1063	0.87027	0.87027
1064	0.87027	0.86757
1065	0.86757	0.87027

Determine the parameter values of the grid search algorithm into the kernel function

The performance of the SVR method itself has three Hyperplane parameters that influence it, namely (*C*) cost, (ε) epsilon, and (γ) gamma. These parameters can potentially influence the final model of the formed SVR. One step to determine the optimal parameters of the SVR model is to use a grid search algorithm by dividing the range of parameters to be optimised. Grid search will train several pairs of models

and finally select the pair with the smallest average error. Optimising the Grid search model to search for parameters using R software, before using the grid search process in R software, the package required is e.1071. After the e.1071 package has been installed, the next step is to determine the optimal parameters with the range that will be used to find the best SVR model using the Grid search algorithm, namely:

$$\begin{split} \varepsilon &= (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1), \\ C &= (2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^{0}, 2^{1}, 2^{2}, 2^{3}, 2^{4}, 2^{5}) \ and \\ \gamma &= (2^{-3}, 2^{-2}, 2^{-1}, 2^{0}, 2^{1}, 2^{2}, 2^{3}, 2^{4}, 2^{5}, 2^{6}, 2^{7}, 2^{8}) \end{split}$$

We can see the results of the three kernels in the following table.

Table 6. RBF, Polynomial and Linear Kernel Results

NO	RBF	Polynomial	Linear Kernels
	Kernels	Kernels	
1	0.99999	0.01563	0.00000
2	1.00000	0.01565	0.00012
3	1.00000	0.01565	0.00012
4	0.99996	0.01569	0.00032
5	0.99996	0.01569	0.00032
6	1.00000	0.01564	0.00009
7	1.00000	0.01563	0.00004
8	1.00000	0.01563	0.00003
1061	0.99998	0.98689	0.74561
1062	1.00000	1.02227	0.75737
1063	1.00000	1.01513	0.75502
1064	1.00000	1.01513	0.75502
1065	0.99994	1.08064	0.77619

Creating SVR Models

To find w in the R package e.1071 program, use the equation model, which is solved using a Lagrange multiplier. Meanwhile, to see the value of parameter b using the equation, the values w=26.3 and b=0.8 are obtained. After the values of w and b are obtained, you will get a model formed based on the equation by entering the values of w and b into the equation.

$$f(x) = \sum_{i=1}^{n} (a_i - a_i^*) K(x_i \cdot y_i) + b$$
$$f(x) = w \cdot K(x_i \cdot y_i) + b$$
$$f(x) = 26.3 \cdot K(x_i \cdot y_i) - 0.8$$

Evaluation of SVR Models into Kernel Functions

First, the prediction is made using the SVR model formed with the testing data, and the prediction results for each kernel function are looked for. Then, the data goes to the denormalisation stage because we have previously converted the data into normalised data. The next step will be to get the accuracy value.

Table 7. SVR model	prediction results with kernel function

Daily Times	SVR with RBF Kernel	SVR with Polynomial Kernel	SVR with Linear Kernel
2022-04-01	27.100	11.456	13.686
2022-04-02	27.100	11.958	13.987
2022-04-03	27.100	11.788	13.886
2022-04-04	27.100	11.457	13.687
2022-04-05	27.100	11.457	13.687
2022-04-06	27.100	11.704	13.836
2022-04-07	27.099	11.537	13.736
2022-04-08	27.100	11.215	13.538
••••			
2023-05-27	27.099	26.755	20.410
2023-05-28	27.100	27.686	20.719
2023-05-29	27.100	27.498	20.657
2023-05-30	27.100	27.498	20.657
2023-05-31	27.098	29.221	21.214

Data Denormalisation

After getting the prediction results using the SVR-Grid Search model formed from each kernel, the next step is to denormalise the data, where previously, the actual data you wanted to predict had gone through data normalisation to use a small number scale. So, the next stage is that the results of the prediction output data using the kernel function must be denormalised first to return the numbers in large form to avoid empty values in the prediction results of each kernel function.

 Table 8. Denormalisation of Prediction Data with Kernel

 Functions

i unetions			
Daily	Denormali-	Denormali-	Linear
Time	zing SVR with	zing SVR with	Denormaliza
	RBF	Polynomials	tion of SVR
2022-04-01	Rp10,752,858	Rp4,964,768	Rp5,789,906
2022-04-02	Rp10,752,964	Rp5,150,459	Rp5,901,219
2022-04-03	Rp10,752,858	Rp5,087,452	Rp5,863,830
2022-04-04	Rp10,753,000	Rp4,965,235	Rp5,790,190
2022-04-05	Rp10,753,000	Rp4,965,235	Rp5,790,190
2022-04-06	Rp10,752,920	Rp5,056,664	Rp5,845,420
2022-04-07	Rp10,752,778	Rp4,994,852	Rp5,808,174
2022-04-08	Rp10,752,991	Rp4,875,427	Rp5,735,102
	•••••	•••••	
2023-05-27	Rp10,752,778	Rp10,625,403	Rp8,277,530
2023-05-28	Rp10,753,000	Rp10,969,752	Rp8,391,971
2023-05-29	Rp10,752,991	Rp10,900,235	Rp8,369,083
2023-05-30	Rp10,752,991	Rp10,900,235	Rp8,369,083
2023-05-31	Rp10,752,431	Rp11,537,671	Rp8,575,076

Accuracy Comparison Using MAPE

After the above calculations, the accuracy of each kernel function and the percentage of Testing & Training data sharing can be compared. It can be seen in the following table.

Table 9. Accuracy Comparisons

Kernel	Data Sharing	MAPE
Functions		
	60:40	9.77
DDE	70:30	9.65
КDГ	80:20	9.46
	90:10	9.19
	60:40	5.80
Dolinomial	70:30	6.37
Foinionnai	80:20	8.14
	90:10	10.72
	60:40	5.43
Linian	70:30	5.59
Limer	80:20	6.39
	90:10	7.28

Gold Price Prediction

The following is a comparison table of predicted data and actual gold price data based on testing data with one of the best models that have been determined.

 Table 10. Comparison of actual data and predicted data results

Daily Times	Actual Data	Daily Times	Prediction Data
2022-04-01	Rp987,000	2023-06-01	Rp1,050,797
2022-04-02	Rp989,000	2023-06-02	Rp1,051,209
2022-04-03	Rp985,000	2023-06-03	Rp1,051,634
2022-04-04	Rp985,000	2023-06-04	Rp1,052,043
2022-04-05	Rp985,000	2023-06-05	Rp1,052,455
2022-04-06	Rp988,000	2023-06-06	Rp1,052,871
2022-04-07	Rp983,000	2023-06-07	Rp1,053,304
2022-04-08	Rp984,000	2023-06-08	Rp1,053,717
		•••••	
2023-05-27	Rp1,048,000	2024-05-27	Rp893,989
2023-05-28	Rp1,048,000	2024-05-28	Rp899,033
2023-05-29	Rp1,047,000	2024-05-29	Rp898,976
2023-05-30	Rp1,048,000	2024-05-30	Rp902,268
2023-05-31	Rp1,056,000	2024-05-31	Rp899,882

Based on the Figure 3 plot above, the prediction results show an increase from the previous actual data, which is not much different. However, in mid-2024, the gold price prediction will decrease from the actual data, as follows is the prediction table for the next 1 (one) year.





Figure 3. Comparison plot of actual data and predicted data

The research results are almost the same as the results of research conducted by [6] analysing the use of the Support Vector Regression method with the Grid Search Algorithm in predicting stock prices. The SVR prediction model was formed with quite good results, with a MAPE value of 5.470% for C = 10, while MAPE at C = 1 had a value of 2.954%. This means that this prediction model is good enough for investors to use in making decisions regarding stock prices.

Conclusion

From the research that has been carried out, the processing results obtained using the Support Vector Regression method using the grid search algorithm are obtained from the results of tests that have been carried out with three types of kernels, namely RBF kernels, Polynomial kernels and linear kernels. It was found that the gold price prediction results had a very good level of accuracy using a linear kernel where the MAPE accuracy value was 5.43 with a division of training and testing data of 60: 40, and the prediction results for gold prices experienced increases and decreases that were not far from the actual data. In the prediction results, the price of gold will increase in 2023 - June - 01 to 2024 - March - 23 with a high price, namely above IDR 1,000,000.00, while it will experience a decline in 2024 - March - 24 with a price below IDR 1,000,000.00.

References

- [1] Mas Rahmah, S. H., & MH, L. M. (2019). *Hukum Pasar Modal*. Prenada Media.
- [2] Guntur, M., Santony, J., & Yuhandri, Y. (2018). Prediksi harga emas dengan menggunakan metode Naïve Bayes dalam investasi untuk meminimalisasi resiko. Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), 2(1), 354-360.
- [3] Komariah, S. S., Lubis, R. S., & Cipta, H. (2021). Analisis Pengendalian Persediaan Sepeda Motor Honda Dengan Penerapan Klasifikasi Always Better Control Pada Dealer Putra Suryajaya Raya II. Jurnal Sains Matematika Dan Statistika, 7(2), 13-26.
- [4] Dewi, S., Nugraha, I., Islami, M. C. P. A., Sari, R. N., & Winursito, Y. C. (2022). Pengendalian Persediaan Material Menggunakan Metode Continuous Review dengan Sistem (r, Q). JUMINTEN, 3(2), 1-12.

- [5] Rahmadayanti, C., Rabbani, H., & Rohmawati, A. A. (2018). Model GARCH dengan pendekatan conditional maximum likelihood untuk prediksi harga saham. *Indonesia Journal on Computing (Indo-JC)*, 3(2), 21-28.
- [6] Saputra, G. H., Wigena, A. H., & Sartono, B. (2019). Penggunaan support vector regression dalam pemodelan indeks saham syariah indonesia dengan algoritme grid search. *Indonesian Journal of Statistics and Its Applications*, 3(2), 148-160.
- [7] Drajana, I. C. R. (2017). Metode support vector machine dan forward selection prediksi pembayaran pembelian bahan baku kopra. *ILKOM Jurnal Ilmiah*, 9(2), 116-123.
- [8] Musmirani, M., Purnamasari, I., & Suyitno, S. (2021). Penerapan Latent Class Regression Analysis dalam Segmentasi Pasar. *EKSPONENSIAL*, 11(1), 47-56.
- [9] Purnama, D. I. (2020). Peramalan jumlah penumpang datang melalui transportasi udara di Sulawesi Tengah menggunakan Support Vector Regression (SVR). Jurnal Ilmiah Matematika dan Terapan, 17(1).
- [10] Chairani, D., Prasetya, N. H., & Cipta, H. (2021). Analisis Pengendalian Persediaan Obat Antibiotik RSU Haji Medan Dengan Menerapkan Metode Always Better Control, Economic Order Quantity, Dan Reorder Point. *TIN: Terapan Informatika Nusantara*, 1(12), 618-622.
- [11] Dewi, K. C., Ciptayani, P. I., Surjono, H. D., & Priyanto, P. (2019). Blended Learning: Konsep dan Implementasi pada Pendidikan Tinggi Vokasi.
- [12] Supuwiningsih, N. N., Kusuma, A. S., Pratiwi, E. L., & Pratami, N. W. C. A. (2022). *Statistik Forecasting Dalam Sistem Informasi Geografis*. Media Sains Indonesia.
- [13] Maulana, R., & Kumalasari, D. (2019). Analisis Dan Perbandingan Algoritma Data Mining Dalam Prediksi Harga Saham Ggrm. J. Inform. Kaputama, 3(1), 22-28.
- [14] Hendayanti, N. P. N., & Nurhidayati, M. (2020). Perbandingan metode seasonal autoregressive integrated moving average (SARIMA) dengan support vector regression (SVR) dalam memprediksi jumlah kunjungan wisatawan mancanegara ke Bali. Jurnal Varian, 3(2), 149-162.
- [15] Putra, A. L., & Kurniawati, A. K. (2021). Analisis Prediksi Harga Saham PT. Astra International Tbk Menggunakan Metode Autoregressive Integrated Moving Average (ARIMA) dan Support Vector Regression (SVR): Array. Jurnal Ilmiah Komputasi, 20(3), 417-424.
- [16] Pradana, R. S. (2021). Penerapan Analisis Jalur Dalam Mengidentifikasi Penyebab Fluktuasi Harga Cabai Merah Di Kabupaten Aceh Jaya. Jurnal Agrica, 14(1), 20-32.
- [17] Arimuko, A., Wibawa, A. S. W., & Firmansyah, A. (2019). Analisis Perbandingan Penentuan Hiposentrum Menggunakan Metode Grid Search, Geiger, dan Random Search: Studi Kasus pada Letusan Gunung Sinabung 2017. DIFFRACTION:

Journal for Physics Education and Applied Physics, 1(2), 22-28.

- [18] Kusdarwati, H., Effendi, U., & Handoyo, S. (2022). Analisis Deret Waktu Univariat Linier: Teori dan Terapannya dengan Rstudio. Universitas Brawijaya Press.
- [19] Pambudi, R. A., Setiawan, B. D., & Wijoyo, S. H. (2018). Implementasi Fuzzy Time Series untuk Memprediksi Jumlah Kemunculan Titik Api. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 2(11), 4767-4776.
- [20] Wu, Y., & He, K. (2018). Group normalisation. In Proceedings of the European conference on computer vision (ECCV) (pp. 3-19).