



Implementation of Support Vector Regression (SVR) Analysis in Predicting Gold Prices in Indonesia

Suwardi Annas¹, Zulkifli Rais¹, Aswi Aswi¹, Indrayasaro¹, Nurfajriani¹

¹ Statistics Department, Universitas Negeri Makassar, Indonesia
suwardi_annas@unm.ac.id

Abstract. Gold is jewelry made from precious metals which are soft and easy to shape. The value of gold tends to fluctuate every year. Gold investments, like other investments, are not risk-free. So, investors can experience profits or even losses. Gold price predictions are needed to determine investors' opportunities in the future. Support Vector Regression (SVR) is an application of Support Vector Machine (SVM) for regression cases whose output is real or continuous numbers. The use of the SVR method has been carried out in several studies. However, the use of the SVR method to predict the rate of gold prices in Indonesia has not been carried out. The aim of this research is to predict the rate of gold prices in Indonesia using the SVR method. The data that will be predicted is gold price data in Indonesia from October 2017 to October 2022. The SVR method is used due to the non-linear and fluctuating nature of gold price data. The study employs the Radial Basis Function (RBF) kernel considering the three parameters in the RBF kernel as gamma (γ), cost (C) and epsilon (ϵ), which will be optimized using the grid search method. The results shows that the best parameters obtained are when $\gamma = 1$, $C = 1$, and $\epsilon = 0.1$, with $k = 10$ because it yields the smallest error value. The level of accuracy in the prediction results is obtained using the Mean Absolute Percentage Error (MAPE) value. The MAPE value is approximately 1.28% suggesting that the prediction accuracy was very good because the MAPE value obtained was $<10\%$.

Keywords: Gold, Kernel, Support Vector Regression (SVR)

1 Introduction

Forecasting is a crucial tool for decision makers and policy makers, both individual and organizational. It is a method used to predict future events within a specified time frame by analyzing relevant historical data. One forecasting method whose results can be used as a reference for forecasting future values is time series forecasting [1].

Common time series methods used for linear data include Autoregressive (AR), Moving Average (MA), Autoregressive Integrated Moving Average (ARIMA), and Seasonal Autoregressive Integrated Moving Average (SARIMA). On the other hand, for time series forecasting with non-linear data, commonly employed methods are Support Vector Machine (SVM), Backpropagation (BPG), K-Nearest Neighbors (KNN), and Fuzzy Time Series (FTS). SVM is a supervised learning method typically used for both classification and regression tasks [2]. Initially designed for classification,

SVM can be adapted to handle regression tasks by incorporating the concept of the ε -insensitive loss function [3].

SVR is used to find a function, denoted as $f(x)$, which takes the form of a hyperplane (a separator line) in the shape of a regression function that effectively fits all input data while minimizing errors. SVR is not only capable of addressing issues with linear data but can also handle problems associated with non-linear data. One of the advantages of SVR is its ability to mitigate overfitting problems. As a result, SVR can deliver improved performance using the available data [4].

In this study, we will utilize gold price data from Indonesia for the past five years (from October 2017 to October 2022). Gold, a precious metal with soft and malleable properties, is highly regarded as an investment instrument in Indonesia. One primary reason for its popularity is its perceived profitability. Additionally, gold is favored due to its tendency to exhibit stability or fluctuations with an overall annual increase in value. This, in turn, makes gold a low-risk investment choice and an excellent option for novice investors. Furthermore, gold serves as a hedge against inflation, which is why it is frequently employed as an investment tool [5].

Gold investment can be categorized into two main types: physical investment and digital investment. Physical investment involves owning gold directly, while digital investment can be further divided into two categories: savings and gold trading. Gold trading is conducted online, without the need for physical ownership. As a result, such investments may lead to both profits and losses [5]. To minimize investor losses, this study employs a method to predict gold prices. Specifically, the SVR method is used due to the non-linear and fluctuating nature of gold price data. The study employs the Radial Basis Function (RBF) kernel with three parameters, which will be optimized using the grid search method [4].

The use of the SVR method has been carried out in several studies. For example, forecasting the number of passengers arriving via air transportation in Central Sulawesi using SVR [4], predicting the price of palm oil in Indonesia and the EUR/USD currency exchange rate using the SVR method [6], and SVR analysis with the RBF kernel to predict the inflation rate in Indonesia [7]. However, the use of the SVR method to predict the rate of gold prices in Indonesia has not been carried out. The aim of this research is to predict the rate of gold prices in Indonesia using the SVR method.

2 Literature Review

2.1 Support Vector Regression (SVR)

Support Vector Regression (SVR) is a form of application of the SVM method for regression cases so that it can be used for forecasting like the regression method in a statistical approach. Smola and Scholkopf explain SVR by splitting an n training data set, (x_1, y_1) with $i = 1, 2, \dots, n$. Whereas $x_i = \{x_1, x_2, \dots, x_p\}^T \in R^n$ is a vector of the input space and $y_i = \{y_1, y_2, \dots, y_n\} \in R$ is an output value based on the corresponding x_1 value [8]. Let $f(x)$ be a non-linear function (hyperplane) in the following form:

$$f(x) = w^t \varphi(x) + b \quad (1)$$

where:

\mathbf{w} = weighting vector;

$\varphi(\mathbf{x})$ = function that maps input \mathbf{x} into feature space;

b = coefficient of bias;

\mathbf{x} = input vector;

$f(x)$ = regression function.

In order for the function $f(x)$ to get a good generalization, it can be done by minimizing the norm of \mathbf{w} . Therefore, it is necessary to solve the optimization problem:

$$\min_{\frac{1}{2}} \|\mathbf{w}\|^2 \quad (2)$$

that fulfills $\begin{cases} \mathbf{y}_i - \mathbf{w}^t \varphi(\mathbf{x}_i) - b \leq \varepsilon, i = 1, 2, \dots, n \\ \mathbf{w}^t \varphi(\mathbf{x}_i) - \mathbf{y}_i + b \leq \varepsilon, i = 1, 2, \dots, n \end{cases}$

with:

\mathbf{y}_i : actual value of the i-period;

(\mathbf{x}_i) : estimated value of period i;

ε : epsilon value.

It is assumed that there is a function $f(x)$ which can approximate all points $(\mathbf{x}_1, \mathbf{y}_1)$ with precision ε . In this case, it is assumed that all points are said to be feasible if they are within the range $f(x) \pm \varepsilon$, and it is said to be infeasible if there are several points that may be out of the range $f(x) \pm \varepsilon$, so as to overcome the problem Inadequate delimiters can be done by adding slack variables ξ and ξ^* [9].

$$\min_{\frac{1}{2}} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (3)$$

That fulfills $\begin{cases} \mathbf{y}_i - \mathbf{w}^t \varphi(\mathbf{x}_i) - b \leq \varepsilon + \xi_i \\ \mathbf{w}^t \varphi(\mathbf{x}_i) - \mathbf{y}_i + b \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$

The solution to the optimization problem of equation (3) can be more easily solved with the following lagrange function [10]:

$$L_p = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i - \xi_i^*) - \sum_{i=1}^n \alpha_i (\varepsilon + \xi_i - \mathbf{y}_i + \mathbf{w}^t \varphi(\mathbf{x}_i) + b) - \sum_{i=1}^n \alpha_i^* (\varepsilon + \xi_i^* + \mathbf{y}_i - \mathbf{w}^t \varphi(\mathbf{x}_i) - b) - \sum_{i=1}^n (\eta_i \xi_i + \eta_i^* \xi_i^*) \quad (4)$$

Thus, the partial derivatives of the Lagrange function with respect to \mathbf{w} , b , ξ_i , ξ_i^* are as follows:

$$\frac{\partial L_p}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \varphi(\mathbf{x}_i) \quad (5)$$

$$\frac{\partial L_p}{\partial b} = 0 \Rightarrow \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \quad (6)$$

$$\frac{\partial L_p}{\partial \xi_i} = 0 \Rightarrow \alpha_i + \eta_i = C \quad (7)$$

$$\frac{\partial L_p}{\partial \xi_i^*} = 0 \Rightarrow \alpha_i^* + \eta_i^* = C \quad (8)$$

Based on this equation, the optimal solution is obtained from the parameter estimation

constraint w in the form of lagrange coefficients α_i and α_i^* which can be written as follows:

$$w = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \varphi(\mathbf{x}_i)$$

By using the Lagrange multiplier and optimality conditions, Equation (1) can be rewritten as follows:

$$f(\mathbf{x}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \varphi(\mathbf{x}_i)^t \varphi(\mathbf{x}_j) + b \quad (9)$$

explanation:

α_i, α_i^* = lagrange multiplier;

b = coefficient of bias.

However, there are difficulties in mapping because the transformation φ is generally unknown and difficult to understand. This problem can be overcome with a kernel trick, namely the scalar multiplication (dot product) $\varphi(\mathbf{x}_i)^t \cdot \varphi(\mathbf{x}_j)$ in the feature space can be replaced by the kernel function [11]:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i)^t \varphi(\mathbf{x}_j)$$

The kernel function is able to implicitly define the transformation φ , so the SVR regression function is written as:

$$f(\mathbf{x}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}_j) + b \quad (10)$$

2.2 Kernel Function

The purpose of using the kernel is to implement a model in a higher-dimensional space (feature space) without the need to explicitly define a mapping from the input space to the feature space. This allows handling non-linear cases in the input space, with the expectation that they become linear in the feature space. The kernel equation used in the SVR method can be seen as follows [12]:

- 1) Linier : $K(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i)^t \cdot \varphi(\mathbf{x}_j)$
- 2) Polynomial : $K(\mathbf{x}_i, \mathbf{x}_j) = ((\gamma \varphi(\mathbf{x}_i)^t \cdot \varphi(\mathbf{x}_j)) + 1)^d$

where d is the degree of the polynomial

- 3) Radial Basis Function (RBF) : $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma |\varphi(\mathbf{x}_i)^t - \varphi(\mathbf{x}_j)|^2)$

where γ is the kernel RBF parameter

- 4) Sigmoid : $K(\mathbf{x}_i, \mathbf{x}_j) = \tan(\gamma(\varphi(\mathbf{x}_i)^t, \varphi(\mathbf{x}_j))) + c$

2.3 Data Normalization

Data normalization is employed to reduce data complexity and ambiguity, as well as to facilitate the simplification of data modification. One of the normalization methods utilized is min-max normalization, which transforms the original linear data, producing a comparative value that correlates the values before and after the process [13]. The normalized data typically falls within a scale of 0 to 1 or -1 to 1. The formula for min-

max normalization can be seen in the equation (11) [14].

$$N = \frac{(x-x_{min})}{(x_{max}-x_{min})} \quad (11)$$

with:

x : attribute data;

x_{min} : minimum data value;

x_{max} : maximum data value.

2.4 Grid Search Optimization

Grid search is a method employed to identify the optimal parameters in a model, ensuring accurate data prediction. One commonly used grid search technique is k-fold Cross Validation, which helps maintain the accuracy of the estimation. The process involves dividing the data into two parts, k-1 parts are utilized as training data, and one part serves as testing data. This process is then repeated k times for each combination of testing and training data [15].

2.5 Root Mean Square Error (RMSE)

RMSE is used to calculate the evaluation results of a model, the smaller or closer to zero the RMSE value, the better [16]. Error calculation using RMSE is formulated as follows [17]:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (12)$$

with:

y_i = actual value at time-i;

\hat{y}_i = estimated value at time-i;

n = amount of data.

3 Research Method

In analyzing the research data, the authors compiled the following steps:

1. Collection of secondary data: monthly reference gold price data from October 2017 to October 2022 were obtained from the website <https://harga-emas.org/> which provides historical gold data in Indonesia.
2. Normalize the data.
3. Determine training and testing data. In this study, different ratios of training to testing data (60:40, 70:30, 80:20, and 90:10) were tested.
4. Determine parameter values in the training data. This involves setting values for parameters C (cost), ϵ (Epsilon), and γ (Gamma) using Grid Search Optimization.
5. Identify the best model by evaluating the optimal parameters or the smallest error among the three parameters C (cost), ϵ (Epsilon), and γ (Gamma).
6. Implement the best model on data testing.

7. Make predictions for the next 6 months.
8. Assess the accuracy of the prediction results using RMSE calculations.
9. Draw conclusions

4 Result and Discussion

4.1 Descriptive Analysis

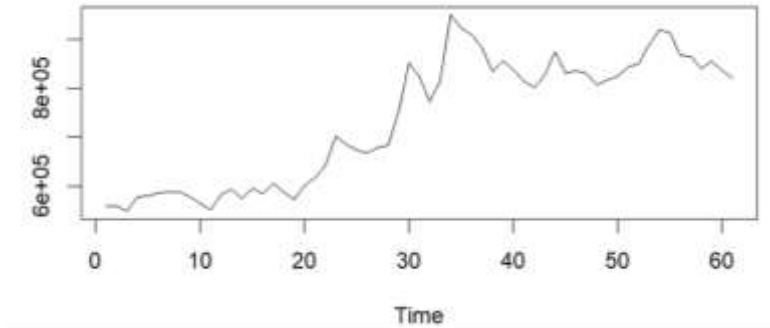


Figure 1 Time Series Plot of Gold Prices in Indonesia.

From Figure 1, the x-axis represents time, and the y-axis represents the price of gold. It can be observed that over the past five years (2017 – 2022), gold prices in Indonesia have fluctuated, showing non-linear patterns rather than forming a straight line. This indicates that the studied data is non-linear. Descriptive statistics of gold price data in Indonesia is given in Table 1.

Table 1 Descriptive Statistics of Gold Price Data in Indonesia

Gold Price	Value (IDR.)
Minimum	549.730,0
Maximum	949.729,0
<i>Mean</i>	736.121,0
Standard Deviation	129.117,7

In Table 1, the minimum value is 549.730,0, and the maximum value is 949.729,0. The mean is 736.121,0, and the standard deviation is 129.117,7.

4.2 Application of Support Vector Regression Method

a. Selection of the Best Model with a Grid Search

Grid search is used to identify the best parameters in the training data using cross-validation procedures. In this study, grid search will be applied with the RBF kernel, considering the parameters in the RBF kernel as *gamma* (γ), *cost* (C) and *epsilon* (ϵ).

Table 2 Grid Search Result Parameters.

k	γ	C	ϵ	<i>Error (IDR)</i>
3	1	1	0.1	1,500,860,936
5	1	1	0.1	1,465,836,090
10	1	1	0.1	1,371,078,914

Based on Table 2, the error values obtained are 1,500,860,936 when $k = 3$, 1,465,836,090 when $k = 5$, and 1,371,078,914 when $k = 10$. The best parameters obtained are when $\gamma = 1$, $C = 1$, and $\epsilon = 0.1$. Considering the explanations above, the parameters used are $\gamma = 1$, $C = 1$, and $\epsilon = 0.1$, with $k = 10$ because it yields the smallest error value.

b. Implementation on Testing Data

The best parameters obtained previously will be used to make predictions on the test data. The predictive plots can be seen in Figure 2, while details of the actual and predicted data on the testing data are provided in Table 3.

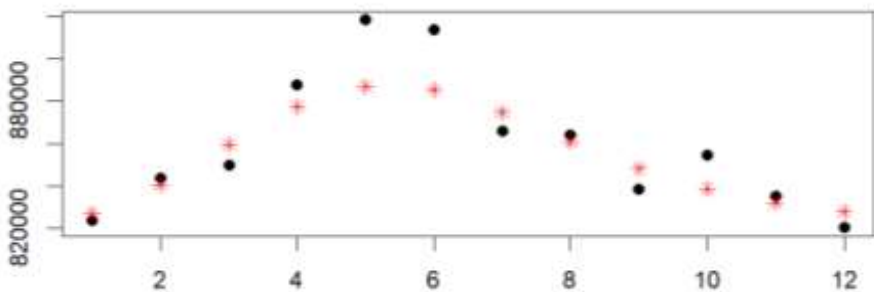


Figure 2 Predictive Plots on Testing Data.

Figure 2 shows time (month) on the x-axis and the price of gold on the y-axis. The predicted results on the testing data are marked with red dots, while the actual data is represented by black dots. It is evident that the predicted results on the testing data closely align with the actual data. Furthermore, based on Table 3, it can be seen that the predicted data is close to the actual data. The following is a line chart comparing actual data with predictive data. Comparison of actual data with predicted data using line charts is given in Figure 3.

Table 3 Actual Data and Predictive Data.

Month	Actual Data (IDR)	Predictive Data (IDR)
50	823903,75	827138,40
51	843837,08	840617,40
52	850051,63	859623,00
53	887938,77	877436,50
54	918595,93	887042,90
55	913666,77	885376,40
56	865931,22	874863,80
57	864259,47	861009,90
58	838738,99	848419,20
59	854651,11	838810,90
60	835253,80	832031,20
61	820459,06	827944,70

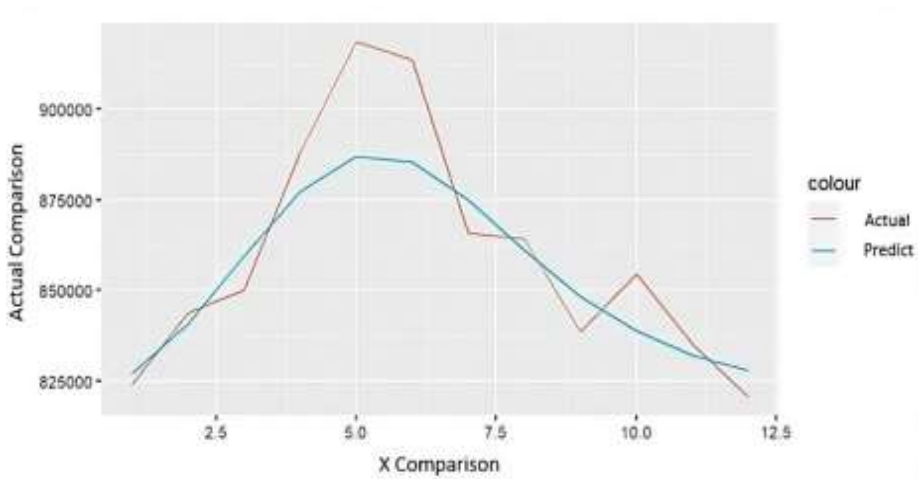


Figure 3 Comparison of Actual Data with Predicted Data using Line Charts

To find out the level of accuracy in the prediction results in Table 3, the MAPE were used with the help of R software. The results were obtained, and the MAPE value obtained was 1.27729%, approximately 1.28%. This suggests that the prediction accuracy was very good because the MAPE value obtained was <10%.

c. Prediction Results of Gold Prices in Indonesia

The prediction results of gold prices in Indonesia in the period November 2022 – October 2023 using the best parameters are shown in Table 4.

Table 4 Prediction Results of Gold Prices in Indonesia

Month	Gold Price (IDR.)
November 2022	827.138,40
December 2022	840.617,40
January 2023	859.623,00
February 2023	877.436,50
March 2023	887.042,90
April 2023	885.376,40
May 2023	874.863,80
June 2023	861.009,90
July 2023	848.419,20
August 2023	838.810,90
September 2023	832.031,20
October 2023	827.944,70

5 Conclusion

Based on the results of the research and discussion that has been carried out to predict the price of gold in Indonesia, the conclusions are obtained. The results of gold price predictions in Indonesia in November 2022 – October 2023 tend to be the same. The accuracy of the prediction results made using the MAPE method is very good because the values obtained is 1,28% which is less than 10%.

Acknowledgement

The authors wish to acknowledge the Universitas Negeri Makassar. This work was funded by DIPA Universitas Negeri Makassar Number: SP DIPA 023.17.2.677523/2023, February 15, 2023. according to the Decree of the Rector of Universitas Negeri Makassar Number: 497/UN36/HK/2023 on May 31, 2023.

References

- [1] A. Purba, "Perancangan Aplikasi Peramalan Jumlah Calon Mahasiswa Baru yang Mendaftar Menggunakan Metode Single Exponential Smothing (Studi Kasus: Fakultas Agama Islam UISU)," *J. Ris. Komput.*, vol. 2, no. 6, pp. 8–12, 2015.
- [2] A. S. Ritonga and E. S. Purwaningsih, "Penerapan Metode Support Vector Machine (SVM) dalam Klasifikasi Kualitas Pengelasan SMAW (Shield Metal Arc Welding)," *J. Ilm. Educ.*, vol. 5, no. 1, pp. 17–25, 2018.
- [3] A. Zeputra and F. Utamingrum, "Perbandingan Akurasi untuk Deteksi Pintu Berbasis HOG dengan Klasifikasi SVM Menggunakan Kernel Linear, Radial Basis Function dan Polinomial pada Raspberry Pi," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 11, pp. 4746–4757, 2021, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/10090%0Ahttp://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/10090/4483>.
- [4] D. I. Purnama and O. P. Hendarsin, "Peramalan Jumlah Penumpang Berangkat Melalui

- Transportasi Udara di Sulawesi Tengah Menggunakan Support Vector Regression (SVR),” *Jambura J. Math.*, vol. 2, no. 2, pp. 49–59, 2020, doi: 10.34312/jjom.v2i2.4458.
- [5] A. Aulia, B. Aprianti, Y. Supriyanto, and C. Rozikin, “Prediksi Harga Emas dengan Menggunakan Algoritma Support Vector Regression (SVR) dan Linear Regression (LR),” *J. Ilm. Wahana Pendidik.*, vol. 8, no. 5, pp. 84–88, 2022, doi: 10.5281/zenodo.6408864.
- [6] S. Saadah, F. Zahra, and H. Haifa, “Support Vector Regression (SVR) Dalam Memprediksi Harga Minyak Kelapa Sawit di Indonesia dan Nilai Tukar Mata Uang EUR/USD,” *J-Cosine*, vol. 5, no. 1, pp. 85–92, 2021, [Online]. Available: <http://jcosine.if.unram.ac.id/>.
- [7] Isnaeni, Sudarmin, and Z. Rais, “Analisis Support Vector Regression (SVR) dengan Kernel Radial Basis Function (RBF) untuk Memprediksi Laju Inflasi di Indonesia,” *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 4, no. 1, pp. 30–38, 2022, doi: 10.35580/variasiunm13.
- [8] A. J. Smola and B. Scholkopf, “A Tutorial on Support Vector Regression,” *Stat. Comput.*, vol. 14, no. 3, pp. 199–222, 2004, doi: 10.1210/me.10.7.813.
- [9] R. E. Cahyono, J. P. Sugiono, and S. Tjandra, “Analisis Kinerja Metode Support Vector Regression (SVR) dalam Memprediksi Indeks Harga Konsumen,” *J. Teknol. Inf. dan Multimed.*, vol. 1, no. 2, pp. 106–116, 2019, [Online]. Available: www.siskaperbapo.com.
- [10] R. E. Caraka, H. Yasin, and A. W. Basyiruddin, “Peramalan Crude Palm Oil (CPO) Menggunakan Support Vector Regression Kernel Radial Basis,” *J. Mat.*, vol. 7, no. 1, pp. 43–57, 2017, doi: 10.24843/jmat.2017.v07.i01.p81.
- [11] R. Amanda, H. Yasin, and A. Prahutama, “Analisis Support Vector Regression (SVR) dalam Memprediksi Kurs Rupiah Terhadap Dolar Amerika Serikat,” *J. Gaussian*, vol. 3, no. 4, pp. 849–857, 2014.
- [12] D. G. Jayaswara, I. Slamet, and Y. Susanti, “Prediction of Central Asia Bank’s Stock Price using Support Vector Regression Method,” in *Proceeding International Conference on Religion, Science and Education*, 2023, pp. 7–12.
- [13] Wartariyus, H. D. Sengaji, S. Wulandari, and R. Alnida, “Sistem Informasi Bimbingan Konseling Pada SMA N 4 Kotabumi dengan Menggunakan Metode Waterfall,” *J. Ilm. Inform. dan Komput. Surya Intan*, vol. 8, no. 2, pp. 105–119, 2021.
- [14] A. W. Ishlah, Sudarno, and P. Kartikasari, “Implementasi Gridsearchcv Pada Support Vector Regression (Svr) Untuk Peramalan Harga Saham,” *J. Gaussian*, vol. 12, no. 2, pp. 276–286, 2023, doi: 10.14710/j.gauss.12.2.276-286.
- [15] N. R. Robynson and Y. Sibaroni, “Analisis Tren Sentimen Masyarakat Terhadap Pembatasan Sosial Berskala Besar Kota Jakarta Menggunakan Algoritma Support Vector Machine,” *e-Proceeding Eng.*, vol. 8, no. 5, pp. 10166–10178, 2021.
- [16] A. D. Arisandi, Ferdiansyah, L. Atika, E. S. Negara, and K. R. N. Wardani, “Prediksi Mata Uang Bitcoin Menggunakan LSTM dan Analisis Sentimen pada Sosial Media,” *J. Ilm. Komputasi*, vol. 19, no. 4, pp. 559–566, 2020.
- [17] L. Wiranda and M. Sadikin, “Penerapan Long Short Term Memory pada Data Time Series untuk Memprediksi Penjualan Produk Pt. Metiska Farma,” *J. Nas. Pendidik. Tek. Inform.*, vol. 8, no. 3, pp. 184–196, 2019.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

