Prediction of Central Asia Bank's Stock Price using Support Vector Regression Method

Dide Guna Jayaswara[1], Isnandar Slamet[2], Yuliana Susanti[3]
[1,2,3] Statistics Department, Faculty of Mathematics and Natural Sciences, Sebelas Maret University,
Jl. Ir. Sutami No 36 Surakarta 57126, Indonesia. Tel. +62-271-646994, Fax. +62-271-636268.

Corresponding author
[1]dideguna@student.uns.ac.id , [2] isnandarslamet@staff.uns.ac.id , [3] yulianasusanti@staff.uns.ac.id

Abstract: The capital market in Indonesia is a developing market that is very vulnerable to global economic conditions and world capital markets. There is a term mostly found in the capital market, namely investment. Investment is a form of delaying consumption from the present to the future where there is a risk of uncertainty. In capital market investment, it is necessary to analyse stock price predictions to find out stock prices in the future by using historical stock prices because stock prices change quickly from time to time, making the shareholder immediately decide when the shares should be sold or retained. This study aims to obtain stock price predictions and model implementation and determine the model's accuracy in predicting a stock's movement to benefit from each stock price that changes from time to time or can be referred to as time series data. Therefore, we need a method that can predict stock prices. In this study, the predicted stocks are BCA stock data using the SVR algorithm with linear and RBF kernels. SVR has advantages in making accurate stock price predictions and can overcome overfitting problems. From several studies conducted, SVR provides optimal results for stock price prediction analysis. Based on the results of the research for BCA shares, it can be concluded that the SVR algorithm has an excellent predictive accuracy value with a linear kernel function with the help of a grid search using a k-fold Cross-validation of 3, which has an R-square value of training data of 93.79% and an R- square data testing is 92.98%, Mean Absolute Percentage Error (MAPE) training data is 0.2340, and MAPE data testing is 0.1021, Root Mean Square Error (RMSE) training data is 0.0597, and RMSE data testing is 0.0499. This algorithm shows that the SVR method is suitable for predicting stock prices.

Keywords: Prediction, Stock, Support Vector Regression, Time Series.

Introduction

The capital market in Indonesia is a developing market that is very vulnerable to global economic conditions and world capital markets (Septiningrum et al. 2015). There is a term that is often found in the capital market, namely investment. Investment is a form of delaying consumption from the present to the future where there is a risk of uncertainty (Dwi and Pengembangan, 2015).

Stock price prediction is an analytical technique to determine future stock prices using past historical stock prices. Stock price predictions are beneficial for investors who are engaged in buying and selling stocks. This technique avoids losses due to the nature of stock prices which move lucratively and tend to be dynamic at any time, so stock price predictions are needed to maximize profits for investors (Yudhawan, 2020).

Changes in stock prices that are relatively fast from time to time make shareholders immediately decide when shares should be sold or maintained. An accurate predictive model for stock price movements can assist investors in making stock transaction decisions because stock price movements tend to be non-linear, which makes it difficult for investors to make predictions. Therefore, predicting stock price movements is still a hot topic for discussion in buying and selling stocks. Thus, there is research on predictions using several models (Fadilah et al. 2020). Stock data is time series data that moves continuously over time, and this is a characteristic of time series models to produce excellent and optimal predictions. (Rahmadayanti et al. 2018).

ISSN 2829-3355 | EISSN 2828-8467

The Support Vector Machine (SVM) algorithm is a classification method that can produce a learning process or learning which is then separated by a line called a hyperplane into two groups. The SVM used for the regression approach is the Support Vector Regression (SVR) method. The concept of SVR is to maximize the hyperplane to obtain support vector data (Septiningrum et al. 2015).

Problems in the real world are rarely linear and more non-linear (Jondri and Saepudin, 2015). To solve this non-linear problem, the kernel function is used. The kernel function is a function that maps data to a higher dimensional space in the hope that the data will have a better structure so that it is easier to separate. Four kernel functions are widely used: linear, sigmoid, radial basis function (RBF), and polynomial kernel functions (Neneng and Asep, 2019). Kernels often used in the SVR method for predicting stock prices more often use linear and RBF kernels because these kernels provide better results for predicting stock prices.

Several previous studies have carried out stock prediction analysis using the SVR method. Research conducted by Yudhawan (2020) regarding the implementation of SVR for forecasting the stock prices of mining companies in Indonesia. The results showed that the prediction analysis of ADRO, PTBA, and ITMG stock prices using the SVR algorithm had good predictive accuracy. The accuracy used is the value of R-square and Mean Absolute Percentage Error (MAPE).

Henrique et al. (2018) researched predicting stock prices using SVM on daily and minute data. The results showed that the SVR obtained lower prediction results than the random walk model. Using linear kernels and RBF produces more minor errors than random walk models for almost all daily stock prices.

The author wants to do research on predicting Bank Central Asia stock prices using the SVR method with linear kernels and RBF.

Materials and Methods

Study area1. *Support Vector Regression (SVR) Model*
SVR is the application of the SVM algorithm in the regression case. The SVR method applies machine learning theory to regression cases that produce real or continuous numbers. The concept of the SVR algorithm can produce good forecasting values because SVR can solve overfitting problems.

- Forecasting using SVR considers the relationship between the values in the $\boldsymbol{y_t}$ period and the values of the previous time series data elements, using several time lags. The regression function of the SVR method is formulated in the equation:

$$\boldsymbol{f(x_i) = \langle w, x_i \rangle + b}$$

Where $\langle \cdot\ ,\ \cdot \rangle$ are the results of the dot product in the input space. The coefficients $\boldsymbol{w}$ and $\boldsymbol{b}$ are the weights and biases. The coefficient of the equation above can be estimated by minimizing the following risk function.

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

with the provision of

$$\boldsymbol{y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i}$$
$$\boldsymbol{\langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^*}$$
$$\boldsymbol{\xi_i, \xi_i^* \geq 0}$$

The constant $\boldsymbol{C}$ (cost) $> \boldsymbol{0}$ is a penalty governing the trade-off between the tolerable upper bound for errors of more than $\boldsymbol{\varepsilon}$ and the "flatness" of the function $\boldsymbol{f}$. The "flatness" in the previous equation means the minor $\boldsymbol{w}$ and can be found by regularizing the function, which is the norm of $\boldsymbol{w}$.

2. *Kernel Function*
To solve non-linear problems, kernel functions are used. To solve linear problems in high-dimensional space, all you have to do is replace the inner products (x_i and x_j) with the appropriate kernel functions. The SVR is modified by incorporating kernel methods so that the kernel function must be positive to guarantee a unique optimal solution to the square of the optimization problem. The kernel functions used in this study are shown in the table below.

Table 1. Kernel functions in SVR**.**

Kernel Function	Definition
Linear	$K(x_i x) = x_i . x$
Polynomial	$K(x_i x) = (x_i . x + c)^d$
Radial Basis Function	$K(x_i x) = \exp(-\gamma \|x_i - x\|^2)$ With $\gamma = \frac{1}{2\sigma^2}, \gamma > 0$
Sigmoid	$K(x_i x) = \tanh(\sigma(x_i x) + c)$

3. *Grid Search Algorithm*

This algorithm divides the range of parameters to be optimized into grids and traverses all points to get the optimal parameters. The grid search algorithm is usually measured by cross-validation on training data in its application. These parameters must be set before the training process is carried out, commonly called hyperparameters. The hyperparameters that play a role in the SVR method are the types of kernel functions and their parameters, namely $\boldsymbol{\gamma}$ and $\boldsymbol{d}$ as well as parameters $\boldsymbol{C}$ and $\boldsymbol{\varepsilon}$.

4. *Coefficient of Determination* $(\boldsymbol{R^2})$

The coefficient of determination $(\boldsymbol{R^2})$ measures the proportion of the variation in the dependent variable that the independent variables can explain in the model. $\boldsymbol{R^2}$ shows the goodness of the model. The bigger the $\boldsymbol{R^2}$, the better the model. $\boldsymbol{R^2}$ value between 0 to 1 or 0% to 100%.

5. *Mean Absolute Percentage Error (MAPE)*

The selection of the best forecasting method is based on the error rate. The smaller the error rate, the forecasting results are closer to the actual value. The method used to measure errors is MAPE.

$$\boldsymbol{MAPE} = \frac{\sum \left| \frac{\boldsymbol{y_t} - \hat{\boldsymbol{y}}_t}{\boldsymbol{y_t}} \right|}{\boldsymbol{n}} \times \mathbf{100\%}$$

Table 2. Criteria for Accuracy of Forecasting Results with MAPE

Value	Criteria
< 10%	Very Good
10 – 20%	Good
20 – 50%	Enough
> 50%	Bad

6. *Root Mean Square Error (RMSE)*

Root mean square error (RMSE) is a parameter used to evaluate the result value of a measurement to the actual value or the value considered correct. The smaller the RMSE value, the closer the data clustering is to accuracy. In general, the equation used to calculate the RMSE value is as follows.

$$\boldsymbol{RMSE} = \sqrt{\frac{(\boldsymbol{x'} - \boldsymbol{x})^2 + (\boldsymbol{y'} - \boldsymbol{y})^2}{\boldsymbol{n}}}$$

Procedures

There are several stages and procedures to implement the best model to predict Bank Central Asia's stock price movements using the SVR method. The following are the steps used in this study.

(1) Collect the data then the data that used in this study is secondary data obtained from yahoo.finance.com. In this dataset, there are as many as 2976 data records.
(2) Perform data pre-processing with data scaling, data is scaled using a min-max scaler with a range of 0-1 and dividing the data into training and test data with a ratio of 80% and 20%.
(3) Conduct model training using the SVR method on training data that has been pre-processed data, then proceed with test data. The model is built using the Python application.
(4) Evaluate the best model obtained using test data based on R^2 and MAPE. Evaluate the comparison of the results obtained from the SVR method using linear and RBF kernels.
(5) Forecasting the closing price of shares for the next 10 days.
(6) Implement the results of the best model accuracy.

Data analysis

This study uses stock price data from Bank Central Asia at yahoo.finance.com. The data is closing daily data from 28 September 2010 – 28 September 2022.

Results and Discussion

Identification Model

This study uses secondary data, namely daily data on the closing stock price of BCA from 2010 to 2022. There are 2,976 closing stock price data

divided into two, training data and test data, with 80% training data and 20% test data. Identification of the training data plot for the closing stock price of BCA is presented in Figure 1.

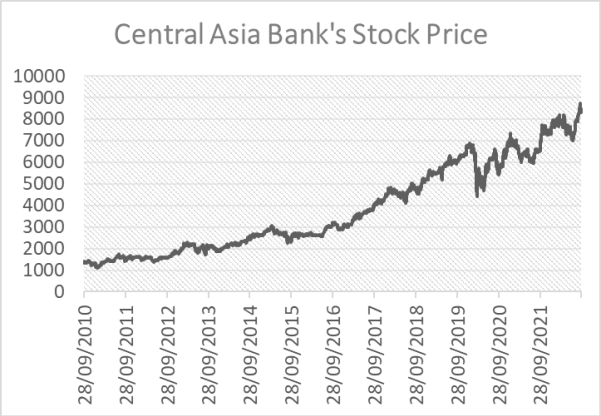

Figure 1. The plot of output indicators

The plot above shows that the plot of the closing stock price data for BCA indicates an upward trend pattern, even though there is a decline at certain times.

Analysis with the SVR method

In the initial step, after dividing the data into training and test data, the data is scaled using the min-max scale. Forecasting using SVR considers the relationship between the value in the period yt and the value of the previous time series data elements, using several time lags. SVR has the concept of determining parameters to form a model so that the support vector enters the hyperplane area to form an optimal regression model.

The parameters used to form the model in this study are linear and radial basis function kernel parameters. The linear kernel has C parameters, while the RBF kernel uses C and gamma parameters. Next, the determination of the range of parameters C and gamma is carried out, as shown in Table 3.

Table 3. Parameter range

Parameter	Value Range
C	$10^{-4}, 10^{-3}, 10^{-2}, \ldots, 10^{3}$
γ	$10^{-5}, 10^{-4}, 10^{-3}, \ldots, 10^{1}$

The best parameters and kernels are determined by tuning the SVR using a grid search and k-fold cross-validation with k = 3 in training on training data. The performance of the model formed is measured using the accuracy values of R-square, MAPE, and RMSE. The closer the R-square value is to 1, the better the model. The best parameter is seen from the minor error value. The results are presented in Table 4.

Table 4. Best parameter

C	Kernel
10^{1}	Linear

The analysis results in predicting the closing stock price of BCA using the SVR method obtained the best parameters, namely C of 10 and a linear kernel with MAPE, R2, and RMSE values on the test data of 0.1021, 97.8%, and 0.0499. Then these results are used to predict the closing share price of BCA. The comparison plot between the training data and the forecast results can be seen in Figure 2. The orange plot shows the value of the training data, while the blue one is the forecast result. It can be seen that the forecast data pattern follows the original data pattern.

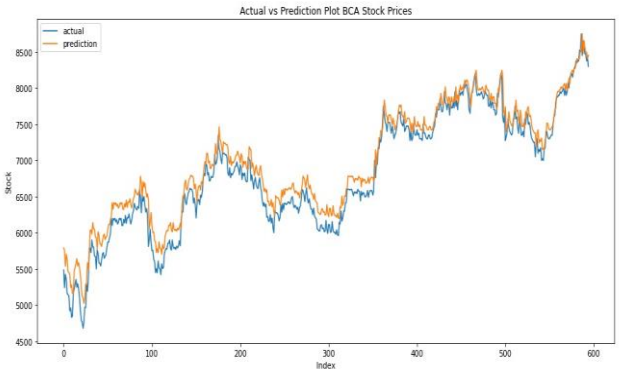

Figure 2. Actual vs prediction BCA stock prices plot

After the model and forecasting results are obtained, an evaluation is needed to measure the accuracy of these results. Evaluation of forecasting results using MAPE, R-square and RMSE values. Table 5. Displays the evaluation value of forecasting results using the SVR method.

Table 5. Evaluation of SVR Method Forecasting Results

Data	MAPE	R-SQUARE	RMSE
Training	0.2340	99.8%	0.0597
Testing	0.1021	97.8%	0.0499

It can be seen in Table 5 above. The table shows that the accuracy of forecasting results on training and test data has excellent results.

Forecasting the next 10 days periods

From the experimental results of the parameters to form the support vector regression model that has been done, the next step is to forecast stock prices using the best model that has been formed before. The model generated from stock data shows good performance. The line graph of actual stock data and stock predictions shows that the forecasting data plot follows the actual data plot, which means that the forecasting data is not much different from the actual data.

The following table will display prediction and actual data on BCA shares using the best model that has been determined.

Table 6. Comparison of Actual and Predictive Data on BCA Shares

Date	Actual	Predictive
4/14/2020	5485	5788
4/15/2020	5240	5761
4/16/2020	5425	5536
4/17/2020	5355	5706
4/20/2020	5160	5642
⋮	⋮	⋮
9/21/2022	8475	8567
9/22/2022	8475	8498
9/23/2022	8375	8498
9/26/2022	8425	8407
9/27/2022	8300	8452

The results of predictions made for the following 10 periods on BCA daily shares with a predetermined model are as follows:

Table 7. Forecasting closing price of the stock for the following 10 periods

Date	Forecast
28-Sep-22	7974
29-Sep-22	8039
30-Sep-22	8099
03-Oct-22	8154
04-Oct-22	8205
05-Oct-22	8251
06-Oct-22	8293
07-Oct-22	8331
10-Oct-22	8367
11-Oct-22	8399

Table 7 shows the stock price forecasting for BCA, which shows that the stock price forecasting results use the SVR model with a linear kernel with parameter C = 10 for the following 10 periods. From the forecasting results, the share price of BCA has increased.

Discussion

This research used the support vector regression method to predict the closing price of shares, which can assist in predicting the movement of Bank Central Asia's stock prices using the SVR method with a linear kernel and RBF. In addition, the model can be applied for forecasting many stock predictions to detect future predictions so that the shareholders of the stocks can minimize the impact of loss.

Conclusions

Based on the results of the research for BCA shares, it can be concluded that the SVR algorithm has an excellent predictive accuracy value with a linear kernel function with the help of a grid search using a k-fold Cross-validation of 3, which has an R-square value of training data of 93.79% and an R-square data testing is 92.98%, Mean Absolute Percentage Error (MAPE) training data is 0.2340, and MAPE data testing is 0.1021, Root Mean Square Error (RMSE) training data is 0.0597, and RMSE data testing is 0.0499. This algorithm shows that the SVR method with linear kernel is suitable for predicting stock prices.

References

Ahmad, A. 2017. Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning. *Yayasan Cahaya Islam Jurnal Teknologi Indonesia*.

Anoraga, P., & Pakarti, P. 2008. *Pengantar Pasar Modal.* Jakarta: Rineka Cipta.

Astudillo, G., Carrasco, R., Fernández-Campusano, C., & Chacón, M. 2020. Copper price prediction using support vector regression technique. *Applied Sciences (Switzerland)*, *10*(19). https://doi.org/10.3390/APP10196648

Cortez, P. dan Donate, J.P. 2012. Evolutionary Support Vector Machines for Time Series Forecasting. *ICANN Part II, LNCS* 7553: 523-530

Darmaji, T., & Fakhrudin, H. 2012. *Pasar Modal di Indonesia*. Jakarta: Salemba Empat.

Dwi, N., & Pengembangan, S. B. (n.d.). 2015. *Penerapan Algoritma Support Vector Machine untuk Prediksi Harga Emas*.

Enri, U. 2018. Optimasi Parameter Support Vector Machine untuk Prediksi Nilai Tukar Rupiah Terhadap Dollar Amerika Serikat. *Jurnal Gerbang, Vol. 8(1): 65-72*.

Fadilah, W. R. U., Agfiannisa, D., & Azhar, Y. 2020. Analisis Prediksi Harga Saham PT. Telekomunikasi Indonesia Menggunakan Metode Support Vector Machine. *Fountain of Informatics Journal*, *5*(2), 45. https://doi.org/10.21111/fij.v5i2.4449

Feta, N. R., & Asep, R. G. 2019. Komparasi Fungsi Kernel Metode Support Vector Machine Untuk Pemodelan Klasifikasi Terhadap Penyakit Tanaman Kedelai. *BRITech (Jurnal Ilmiah Ilmu Komputer, Sains dan Teknologi Terapan), Vol. 1(1): 33-39.*

Furi, R.P., Jondri, dan Saepudin, D. 2015. Prediksi Financial Time Series Menggunakan Independent Component Analysis dan Support Vector Regression Studi Kasus: IHSG dan JII. *E-proceeding of Engineering, Vol. 2(2): 3608-3618.*

Gujarati, D. 1999. *Ekonometrika Dasar*. Jakarta: Erlangga.

Hedianti, E.S. 2019. *Peramalan Harga Saham Dengan Menggunakan Metode Support Vector Regression (SVR)*. Tugas Akhir. Departemen Sistem Informasi ITS, Surabaya.

Heizer, J., & Barry, R. 2009. *Manajemen Operasi Buku 1 Edisi* 9. Jakarta: Salemba Empat.

Henrique, B., Sobreiro, V. and Kimura, H., 2018. Stock price prediction using support vector regression on daily and up to the minute prices. *The Journal of Finance and Data Science*, 4(3), pp.183-201.

Lahmiri, S., 2018. Minute-ahead stock price forecasting based on singular spectrum analysis and support vector regression. *Applied Mathematics and Computation*, 320, pp.444-451.

Lin, K., Lin, Q., Zhou, C., dan Yao, J. 2007. Time Series Prediction Based on Linear Regression and SVR. *Third International Conference on Natural Computation (ICNC)* 0-7695-2875-9/07.

Maulana, N. D., Setiawan, B. D., & Dewi, C. 2019. Implementasi Metode Support Vector Regression (SVR) Dalam Peramalan Penjualan Roti. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*.

Maulida, K. H., Gede Susrama Mas Diyasa, I., Aji, P. R., & Maulana, T. F. (n.d.). *Seminar Nasional Informatika Bela Negara (SANTIKA) Studi Literatur Mengenai Prediksi Harga Saham Menggunakan Machine Learning*.

Mutmainnah, A. dan Widodo, E. 2018. Application of Support Vector Machine (SVM) Methods on Stock Price Forecasting of PT Telekomunikasi Indonesia Tbk. *Seminar Nasional Edusainstek FMIPA UNIMUS: 50*

Naufal, M.F. 2017. *Peramalan Jumlah Wisatawan Mancanegara yang Datang ke Indonesia Berdasarkan Pintu Masuk Menggunakan Metode Support Vector Machine (SVM).* Tugas Akhir. Departemen Sistem Informasi ITS, Surabaya.

Purnama, D.I. 2020. Peramalan Jumlah Penumpang Datang Melalui Transportasi Udara di Sulawesi Tengah Menggunakan Support Vector Regression (SVR). *Jurnal Ilmiah Matematika dan Terapan (JIMT), Vol. 17(1): 109-117.*

Rahmadayanti, C., Rabbani, H., & Rohmawati, A. A. 2018. Model GARCH dengan Pendekatan Conditional Maximum Likelihood untuk Prediksi Harga Saham. *Indonesian Journal on Computing (Indo-JC)*, *3*(2), 21. https://doi.org/10.21108/indojc.2018.3.2.223

Ramdhani, Y., & Mubarok, A. 2019. Analisis Time Series Prediksi Penutupan Harga Saham Antm.Jk Dengan Algoritma SVM Model Regresi. *JURNAL RESPONSIF*, *1*(1). http://ejurnal.univbsi.id/index.php/jti

Rijin, J.N.V. dan Hutter, F. 2018. Hyperparameter Importance A cross Datasets. *Proceedings of ACM Conference on Knowledge Discovery and Data Mining: 2367-2376*

Septiningrum, L., Yasin, H., Jurusan Statistika FSM UNDIP, M., & Pengajar Jurusan Statistika, S. 2015. *Prediksi Indeks Harga Saham Gabungan Menggunakan Support Vector Regression (Svr) Dengan Algoritma Grid Search*. *4*(2), 315–321. http://ejournal-s1.undip.ac.id/index.php/gaussian

Smola, A.J. dan Scholkopf, B. 2004. A tutorial on Support Vector Regression. *Statistics and Computing , Vol. 14: 199-222*

Thissen, U., van Brakel, R., de Weijer, A.P., Melssen, W.J., dan Buydens L.M.C. 2003. Using support vector machines for time series prediction. *Chemometrics and Intelligent Laboratory Systems, Vol. 69: 35-4*

Yasin, H., Prahutama, A., & Utami, T. W. 2014. Prediksi Harga Saham Menggunakan Support Vector Regression Dengan Algoritma Grid Search. *Media Statistik*.

Yudhawan, D. H. 2020. *Implementasi Support Vector Regression Untuk Peramalan Harga Saham Perusahaan Pertambangan di Indonesia.* SKRIPSI. Yogyakarta. Universitas Islam Indonesia.