Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)
Vol 7, No.1, March 2024, pp. 148 – 156
p-ISSN: 2614-3372 | e-ISSN: 2614-6150

Comparative Analysis of Support Vector Regression and Linear Regression Models to Predict Apple Inc. Share Prices

[1]Resza Adistya Pangestu, [2]Anik Vega Vitianingsih*, [3]Slamet Kacung, [4]Anastasia Lidya Maukar, [5]Agustinus Noertjahyana
[1,2,3]Informatics Department, Universitas Dr. Soetomo
[4]Industrial Engineering Department, President University, Bekasi, Indonesia
[5]Informatics Department, Universitas Kristen Petra, Surabaya, Indonesia
Email: [1]reszaadistya@gmail.com, [2]vega@unitomo.ac.id*, [3]slamet@unitomo.ac.id, [4]almaukar@president.ac.id, [5]agust@petra.ac.id

Article Info

Article history:
Received Dec 5th, 2023
Revised Feb 29th, 2024
Accepted Mar 10th, 2024

Keyword:
Apple Inc.
Linear Regression
Prediction
RMSE
Support Vector Regression

ABSTRACT

Stock price prediction is a complex and important challenge for stock market participants. The difficulty of predicting stock prices is a major problem that requires an approach method in obtaining stock price predictions. This research proposes using machine learning with the Support Vector Regression (SVR) model and linear regression for stock price prediction—the dataset used in the daily Apple Inc historical data from 2018 to 2023. The hyperparameter tuning technique uses the Grid Search method with a value of k = 5, which will be tested on the SVR and Linear Regression methods to get the best prediction model based on the number of cost, epsilon, kernel, and intercept fit parameters. The test results show that the linear regression model with all hyperparameters k = 5 with the average taken performs best with a True intercept fit value. The resulting model can get an excellent error value, namely the RMSE value of 0.931231 and MSE of 0.879372. This finding confirms that the linear regression model in this configuration is a good choice for predicting stock prices.

Corresponding Author: (10 pt)
Anik Vega Vitianingsih,
Informatics Department, Universitas Dr. Soetomo
Email: vega@unitomo.ac.id

DOI: http://dx.doi.org/10.24014/ijaidm.v7i1.28594

1. INTRODUCTION

The stock market is crucial in the financial sphere and for investors [1]. Many companies utilize the stock market as a source of long-term funds [2]. Stocks dominate in the securities market. Technical analysts use stock charts to forecast future stock values [3]. However, the abundance of historical data presents a challenge to financial analysts. Stock analysts identify patterns in historical data to understand the direction of stock prices [4]. Several studies have been conducted on stock prediction using deep learning, proving its dominance over traditional models. The research examined various models, techniques, and configurations for optimal stock prediction [3]. Market trends are difficult to predict, and many seek methods to gain greater profits [5]. Forecasting stock price trends is a complicated task. Having foresight is essential for proper decision-making. Forecasting involves making predictions about events that are currently unknown [6].

The difficulty of predicting stock prices is a common problem for stock players. Planning, scheduling, purchasing, strategy formulation, policy making, supply chain operations, and stock prices require the ability to forecast precisely. Therefore, much time has been spent on forecasting [7]. Predicting stock prices is a complex and challenging task for investors to predict future profits [8]. Forecasting is about making statements about things that are not yet known. The term predictive forecasting is often used to describe the same thing. Forecasting also addresses properly presenting and using forecasting results [9].

Many studies have been conducted to predict stock prices, one of which is a study that tests using linear regression. This test found that linear regression provides prediction results with an RMSE value of 0.93 and MSE of 0.88 [10]. Furthermore, in research that predicts stock prices by testing 2 models, namely ARIMA and SVR. The results of this test found that SVR is better than ARIMA; the RMSE value in the SVR model is

0.053294, while ARIMA's is 0.067235 [11]. However, previous studies in comparing models did not perform hyperparameter tuning first to determine the optimal parameters for each model. The comparison results are not maximized.

This research applies a machine learning-focused approach using SVR and linear regression models to analyze and forecast stock price movements. The SVR method is used in this study because SVR is a method that can overcome overfitting so that it will produce good preformation, small error, and low error. Overfitting is good for overcoming the case of regression, a condition where a model does not describe the main relationship between input and output variables but rather describes random error or noise [12]. The linear regression method is used to build a prediction model. A statistical measure to determine the strength of the relationship between the dependent and independent variables. This technique is used to predict the value of given input data. Linear regression is based on previous patterns of data relationships, modeling the relationship between scalars and explanatory variables or independent variables and regression models [13]. The improvement of the model comparison used will be tested with hyperparameter tuning based on the grid search technique, namely by finding a combination of parameters based on the kernel model, cost, gamma, epsilon, and intercept fit. The parameters used include kernel using linear and rbf, Cost parameter value between 1, 10, 100, gamma parameter between 0.01, 0.1, 10, epsilon parameter between 0.01, 0.1, 0.2, and fit intercept between true and false.

This research aims to improve the performance of predicting stock price movements using a machine learning approach using SVR and linear regression models. This technique is expected to overcome problems with high volatility and market trends that are difficult to predict.

2. METHOD

The proposed research method refers to the flow diagram of Figure 1, which includes data input, data preparation, model optimization, and validation procedures.

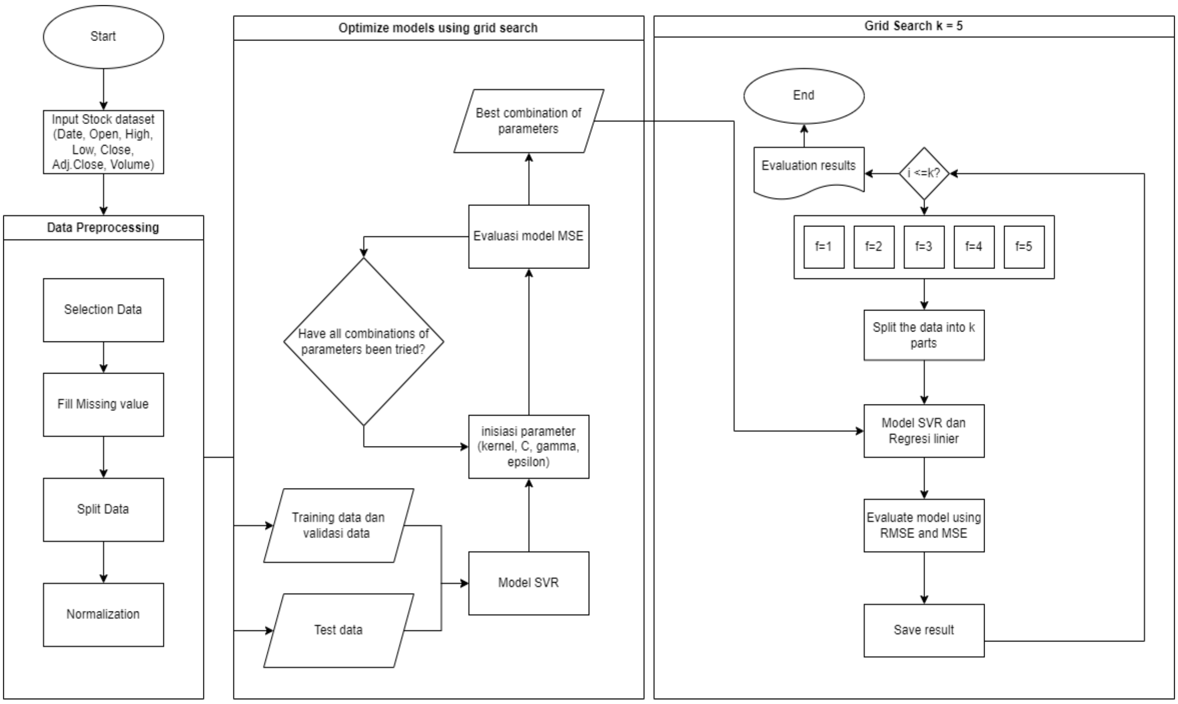

Figure 1. Flow diagram

Table 1 provides a more in-depth explanation of the flowchart and pseudocode. This section will discuss the highlights of the models used during this study. The stock data used is shown in the first section, followed by an explanation of the structure of SVR and linear regression. Below is an explanation of the flowchart.

The data that the user has input includes data selection, filling in empty values, and data normalization so that it can be implemented in the algorithm model by filling in empty values because the stock dataset stores weekday data (holidays are not recorded), it is necessary to fill in the value on the holiday date. Normalize to include all values with the required scale. Split data performs the process of dividing data into two parts, namely training data and test data. Then, build an SVR model to perform hyperparameter initiation, which aims to grid search. There are 3 parameters for each parameter with 3 values, so the process is carried out 17 times. This process is repeated until it reaches a model following the predetermined parameters. After hyperparameter tuning, the next step is to compare the model using the best hyperparameter combination obtained. Cross-validation divides the dataset into *k* parts; in this case, *k=5*.

Furthermore, at each iteration of the k-fold cross-validation, one fold becomes the test data, while the four folds become the training data. In this case, two models are used: SVR and linear regression. Each model will be built using a combination of previously determined hyperparameters. Each model will be trained with pre-distributed training data at each iteration and evaluated for performance using test data. Furthermore, the MSE and RMSE values of each model will be calculated. The program performs the calculation repeatedly as many times as the value of *k* is given.

Table 1. Pseudocode model

```
Input: stock data set, hyperparameter
Output: best configuration, evaluation models
1. Start.
2. Input the stock dataset.
3. The preprocessing of data collection involves selecting data, analyzing it, and
   addressing any missing information through data selection, analysis, and gap-
   filling.
4. Build SVR and linear regression models by initializing parameters (kernel, cost,
   gamma, and epsilon) along with training data.
   a) Initialize the combination list of SVR and linear regression parameters.
   b) For each configuration (kernel, gamma, epsilon):
      1. Build SVR and linear regression models with the following
         parameters and training data.
      2. Model evaluation with RMSE.
      3. End for
      4. Keep the RMSE and parameters if they are the best.
   c) Show the best RMSE for each model.
5. The next stage performs model comparison using cross-validation k=5
   a) Split the data preprocessed as many times as the number k has been
      determined.
   b) Initialize i =1
   c) Initialize evaluation result = []
   d) For each fold = 1 to k:
      1) Fold==i, use as test data; otherwise, use as train data.
      2) Build each model based on the best hyperparameter combination
      3) Model evaluation by finding the RMSE and MSE.
      4) Save evaluation results in evaluation results.
   e) If i==k then displays the evaluation result of all folds
   f) End for
6. End
```

2.1. Dataset

In this study, the required data consists of daily stock price data of Appel Inc. covering the period from 2018 to 2023. The data for making predictions is presented in Table 2. The date variable represents the date associated with the observed data. The open variable represents the trading price of the stock when the associated date starts. The opening price of a stock is the price at which trading begins on a given day. The variable high represents a stock's highest price on the specified date. The variable low value is the lowest price a stock trades at on the date. The close variable's value indicates the stock's last traded price on a particular date. The adj.close variable in the formula represents the adjusted closing price. Any stock splits or dividends distributed to shareholders will be considered when calculating the modified closing price. The volume variable provides information regarding the total number of shares exchanged on the specified date.

Table 2. Appel Inc. Stock Data

Date	Open	High	Low	Close	Adj.Close	Volume
10/1/2018	56.9875	57.355	56.5875	56.815	54.41364	94403200
10/2/2018	56.8125	57.5	56.6575	57.32	54.8973	99152800
…	…	…	…	…	…	…
10/9/2023	176.81	179.05	175.8	178.99	178.99	42390800
10/10/2023	178.1	179.72	177.95	178.39	178.39	43664800

2.2. Data Processing

In this study, all the features listed in Table 2 were included in the analysis. The preprocessing process started with evaluating the initial amount of data obtained from the data collection. Although not all attributes were used, only the closing date and features were focused on. The preprocessing steps involved data selection, handling of blank values, data division, and normalization. This was done to ensure the data was ready to be implemented in the algorithmic model. The focus on the closing date and price attributes is expected to increase the model's effectiveness in analyzing and processing relevant information from the data available in Table 2.

2.2.1. Data Selection

To initialize the hyperparameters for a grid search, the focus was placed on two key parameters: date and close. This step is necessary to determine the optimal values of these two parameters through different combinations in the grid search process. By incorporating these two parameters into the grid search, the best combination of hyperparameter values optimizes the model's performance in the context of the date, and the closing price is expected.

2.2.2. Filling missing value

This stock dataset is limited to weekdays without recording holidays. Therefore, values on holiday dates are filled in using the forward-filling technique. In this technique, missing values are filled with previous values. This process is important to maintain data completeness and prevent blanks on holidays, which may affect the analysis. Forward-filling ensures that each date entry has a relevant value, facilitating consistency within the dataset. Moreover, these value-filling steps are an integral part of the preprocessing stage, which is necessary before implementing the algorithmic model for accurate stock analysis [14].

2.2.3. Data normalization

In this study, *MinMax* normalization is used. *MinMax* normalization is a technique used to scale data from 0 to 1. MinMax normalization is named because it is applied to standardize values within a consistent range. The procedural steps are described in the equation (1) [15].

$$Z = \frac{x - \min(\mathrm{x})}{\max(x) - \min(x)} \tag{1}$$

Where the variable *x* represents the value that needs to be normalized, the variable *min(x)* refers to the lowest possible value for variable *x*, and *max(x)* indicates the highest possible value for variable *x*.

2.2.4. Split data

The stock data is divided into two subsets: training data (80% of the total) with 1012 initial rows and test data (20% of the total) with 253 rows. This division signifies that the training data is drawn from the initial stock dataset of 1265 rows. The 80-20 split aims to train the model at a sufficient volume while providing an independent test dataset to evaluate the model's performance. By ensuring that the training data covers 80% of the original dataset, the model is expected to learn well and produce reliable results when tested on a subset of data that has never been seen before.

2.3. Support Vector Regression (SVR)

SVR is one method that can overcome overfitting to produce good performance [16]. This algorithm aims to find a dividing line or the best hyperplane. Measuring the hyperplane margin is a way to find the best hyperplane. The distance between two hyperplanes with patterns is called the margin (2) [17]. SVR is one of the machine learning methods where SVR is designed to solve regression function problems and generate decisions based on linear combinations of inputted variables [18].

$$\mathrm{f(x, w) = wT\ \varphi(x) + b} \tag{2}$$

Where $\varphi(x)$ is a point in the feature space *F,* the result of mapping *x* in the input space. SVR also finds $||w||$ *2* .

2.4. Linear regression

Linear regression is a machine learning algorithm. It defines the relationship between two variables by fitting a regression line to the data. One of the two variables is the dependent variable, which depends on the other variable, called the independent variable [19]. Ensure a relationship between the dependent and independent variables exists before modeling [20]. The strength of the relationship between variables can be determined by using a scatterplot.

$$a11x1 + a12x2 + \ldots + a1nxn = b1 \tag{3}$$

$$a21x1 + a22x2 + \ldots + a2nxn = b2 \tag{4}$$

$$am1x1 + am2x2 + \ldots + amnxn = bm \tag{5}$$

where *a11*,*a12* etc are inputs to the system, *x1*,*x2* etc are unknown coefficients and *b1*,*b2* etc are outputs of the system. (3)(4)(5). The system has a unique solution if the number of unknown variables equals the number of linearly independent and non-conflicting equations. If the equations conflict, there is no solution. This usually happens when $m>n$. If there are fewer linear independent numbers and non-contradictory equations than unknown ones, then there are infinite solutions [21].

2.5. Hyperparameter optimization

Grid search is a hyperparameter optimization technique frequently used to perform an exhaustive search on a manually defined subset of hyperparameters. Validation is used to evaluate the suitability of parameter values in a grid search. [22]. It is known that the identification results are highly dependent on the hyperparameter selection of SVR and linear regression models. Many experiments are often needed to determine the combination of parameters, such as the Epsilon parameter, Cost parameter, kernel parameter, and intercept fit. [23]. The following parameters are used in the Grid search process. This research uses each parameter. This model was built using the parameters cost with a value of 10, epsilon with a value of 0.1, kernel with a value of Linear, and Fit intercept with True value.

2.6. Evaluation models

To evaluate the performance model in this study using. Cross-validation is very important for evaluating regression methods. It is often used to make predictions and estimate the accuracy of the prediction model [24]. A cross-validation method known as k-fold [25] is used to investigate this case. This method requires the data to be segmented into k groups of equal size each. A common statistical measure used in regression is Mean Squad Error (MSE). Using this equation reflects the average squared error between the projected and true values (6) [26]. Where *Xi* is the predicted value, *Yi* is the real value, and *M* is the number of data.

$$MSE = \frac{1}{M}\sum_{i=1}^{M}(X_i - Y_i)^2 \tag{6}$$

Another metric used to measure the difference between the value predicted by the model or estimator and the true value is the Root Mean Squad Error (RMSE). In contrast to MSE, RMSE gives the measurement error expressed in the same units as the measured variable using the equation (7) [26]. Where *Xi* is the predicted value, *Yi* denotes the true value, and *M* denotes the total number of observations.

$$RMSE = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(X_i - Y_i)^2} \tag{7}$$

3. RESULT AND DISCUSSION

Specifically, the SVR and linear regression models are represented by their respective findings in Table 4, which are based on various hyperparameter settings and obtained through the Grid Search procedure. Each configuration can be characterized by the type of activation function used for its learning speed. In addition, each configuration has been assessed based on several performance measures, and the RMSE for each configuration is also provided. The information can be seen in the table below.

Table 4. Performance model using hyperparameter tuning

Parameters	Value
Linear regression Model	
Fit intercept	True
Rank test score	1
RMSE	0.925591
SVR Model	
Epsilon	Linier

Parameters	Value
Cost	10
Kernel	Linier
Rank test score	1
RMSE	0.924529

The linear regression model using the configuration with the true intercept fit performed best and was ranked 1st due to the lowest RMSE of 0.925591. Conversely, the configuration with a false intercept fit received an RMSE value of 10.40003, which fared poorly out of all possible combinations. The SVR model using the configuration with Cost value 10, epsilon value 0.1, and linear kernel had the best results, with the lowest possible RMSE value, thus reaching rank 1. On the other hand, the configuration using other functions, such as Cost value 10, epsilon 0.01, and linear kernel, showed poor performance with an RMSE value of 0.926827.

The results of the grid search procedure for each model are reported below in Table 6, which is also arranged in order of lowest RMSE value. Table 4 contains configuration information for each of the available models. Table 6 displays the evaluations received through the comparison procedure. These results are based on the calculated hyperparameters.

Table 6. Model evaluation results based on fold

Model	K-Fold	RMSE	MSE
SVR	1	1.029455	1.059778
SVR	2	1.043636	1.089177
SVR	3	0.854454	0.730092
SVR	4	0.759360	0.576627
SVR	5	1.041207	1.084111
Linear Regression	1	0.975265	0.951141
Linear Regression	2	1.028081	1.056951
Linear Regression	3	0.844435	0.713070
Linear Regression	4	0.761733	0.580237
Linear Regression	5	1.046644	1.095463

The k-fold cross-validation method is used in the evaluation process, and the k value used in the evaluation is five, which means that the evaluation is performed five times. Table 6 shows that the model has different evaluation values, including RMSE and MSE. Furthermore, the average of these evaluation results is calculated as presented in Figure 2. Compared to other models, the linear regression model performs better. This can be seen from the RMSE value of 0.931231 and the MSE value of 0.879372. This error value is the lowest when compared to other models. Lower values indicate higher prediction quality.

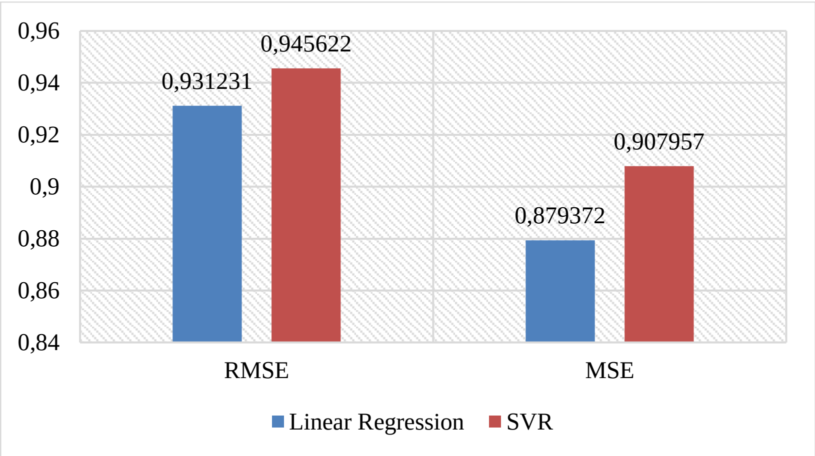

Figure 2. Average evaluations result

The comparison of each model and the actual prices can be found in Figure 3. The graph provides a visual representation of the differences between the models and the actual prices.

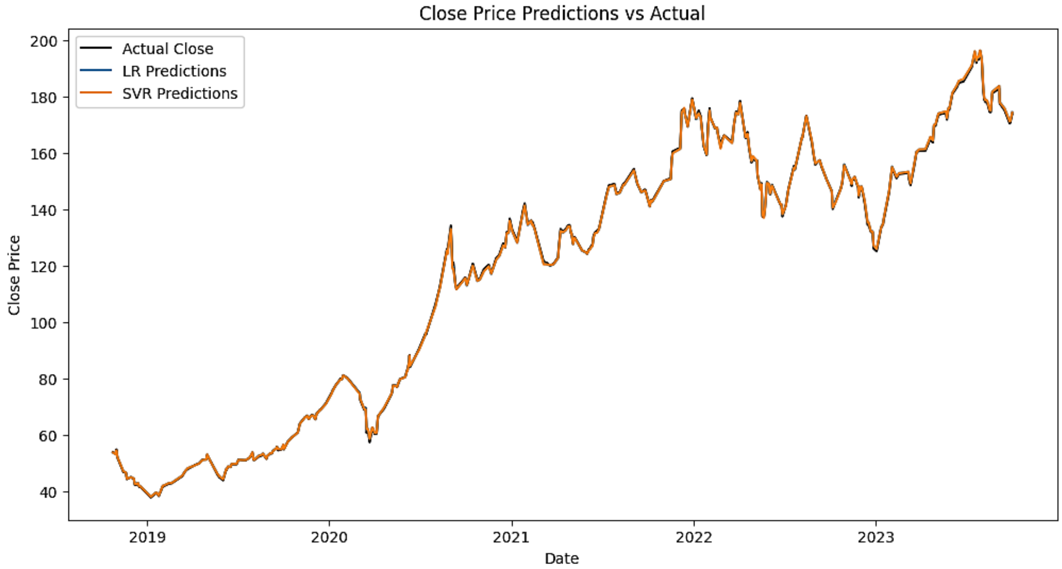

Figure 3. Model comparison chart

4. CONCLUSION

In the investigation of stock price prediction using machine learning, the Linear Regression model performed very well. The model has one intercept that matches the true value. The RMSE value is 0.931231 and the MSE value is 0.879372, both of which show excellent results. Therefore, it can be concluded that linear regression performs better than SVR, as evidenced by the smaller RMSE and MSE values of linear regression compared to SVR. In addition, integrating other deep learning algorithm models into the comparison may provide new insights or information relative to predicting stock prices. For future research on hyperparameter tuning trials, it is essential to explore a wider range of hyperparameter variants to improve the accuracy of each model. Dataset processing involves factoring variables of inflation data and other economic indicators that may affect stock prices to produce more precise predictions.

REFERENCES

[1] X. Ji, J. Wang, and Z. Yan, “A stock price prediction method based on deep learning technology,” *Int. J. Crowd Sci.*, vol. 5, no. 1, pp. 55–72, 2021, doi: 10.1108/IJCS-05-2020-0012.

[2] E. Eka Patriya, “Implementasi Support Vector Machine Pada Prediksi Harga Saham Gabungan (Ihsg),” *J. Ilm. Teknol. dan Rekayasa*, vol. 25, no. 1, pp. 24–38, 2020, doi: 10.35760/tr.2020.v25i1.2571.

[3] Y. Chen, P. Zhao, Z. Zhang, J. Bai, and Y. Guo, “A Stock Price Forecasting Model Integrating Complementary Ensemble Empirical Mode Decomposition and Independent Component Analysis,” *Int. J. Comput. Intell. Syst.*, vol. 15, no. 1, 2022, doi: 10.1007/s44196-022-00140-2.

[4] H. Hewamalage, C. Bergmeir, and K. Bandara, “Recurrent Neural Networks for Time Series Forecasting: Current status and future directions,” *Int. J. Forecast.*, vol. 37, no. 1, pp. 388–427, 2021, doi: 10.1016/j.ijforecast.2020.06.008.

[5] X. Liu, Z. Zhang, and Z. Song, “A comparative study of the data-driven day-ahead hourly provincial load forecasting methods: From classical data mining to deep learning,” *Renew. Sustain. Energy Rev.*, vol. 119, no. November 2019, p. 109632, 2020, doi: 10.1016/j.rser.2019.109632.

[6] A. Hussein, J. Agbinya, and I. Satti, “A Survey on Data mining Techniques for Water Flow Forecasting,” *Aust. J. Basic Appl. Sci.*, vol. 14, no. 3, pp. 13–27, 2020, doi: 10.22587/ajbas.2020.14.3.2.

[7] W. Budiharto, “Data science approach to stock prices forecasting in Indonesia during Covid-19 using Long Short-Term Memory (LSTM),” *J. Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00430-0.

[8] R. Julian and M. R. Pribadi, “Peramalan Harga Saham Pertambangan Pada Bursa Efek Indonesia (BEI) Menggunakan Long Short Term Memory (LSTM),” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 3, pp. 1570–1580, 2021, doi: 10.35957/jatisi.v8i3.1159.

[9] S. Tinggi, T. Kedirgantaraan, and Y. Abstrak, “PERAMALAN (FORECASTING) VOLUME PENUMPANG TERHADAP OPTIMALISASI TERMINAL PENUMPANG DI BANDAR UDARA INTERNASIONAL SUPADIO PONTIANAK 1 Ranggie Juliati,” *J. Gr. Handl. Dirgant.*, vol. 4, no. 1, pp. 2460–1594, 2022.

[10] A. Riyandi, I. Nur Ardiansyah, and R. Dany, “Analisis Data Mining Untuk Prediksi Harga Saham: Perbandingan Metode Linear Regression Dan Pola Historis Data Mining Analysis for Stock Price Prediction: A Comparison of Linear Regression Method and Historical Patterns,” *Jtsi*, vol. 4, no. 2, pp. 278–288, 2023.

[11] A. Kurniawati and A. Arima, “Analisis Prediksi Harga Saham PT. Astra International Tbk Menggunakan Metode Autoregressive Integrated Moving Average (ARIMA) dan Support Vector Regression (SVR),” *J. Ilm. Komputasi*,

vol. 20, no. 3, pp. 417–423, 2021, doi: 10.32409/jikstik.20.3.2732.

[12] A. Aulia, "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.

[13] V. R. Prasetyo, H. Lazuardi, A. A. Mulyono, and C. Lauw, "Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression," *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 1, pp. 8–17, 2021, doi: 10.25077/teknosi.v7i1.2021.8-17.

[14] M. Chiah and A. Zhong, "Tuesday Blues and the day-of-the-week effect in stock returns," *J. Bank. Financ.*, vol. 133, p. 106243, 2021, doi: 10.1016/j.jbankfin.2021.106243.

[15] A. Ambarwari, Q. J. Adrian, and Y. Herdiyeni, "Analisis Pengaruh Data Scaling Terhadap Performa Algoritme Machine Learning untuk Identifikasi Tanaman," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 1, pp. 117–122, 2020, [Online]. Available: http://jurnal.iaii.or.id

[16] H. Muthiah, U. Sa, and A. Efendi, "Support Vector Regression (SVR) Model for Seasonal Time Series Data," *Proc. Second Asia Pacific Int. Conf. Ind. Eng. Oper. Manag.*, no. September 14-16, 2021, pp. 3191–3200, 2021.

[17] L. M. Ginting, M. M. Sigiro, E. D. Manurung, and J. J. P. Sinurat, "Perbandingan Metode Algoritma Support Vector Regression dan Multiple Linear Regression Untuk Memprediksi Stok Obat," *J. Appl. Technol. Informatics Indones.*, vol. 1, no. 2, pp. 29–34, 2021, doi: 10.54074/jati.v1i2.36.

[18] A. Rahmi, "Portofolio Optimal Dengan Mempertimbangkan Prediksi Return Menggunakan Metode Support Vector Regression (SVR) Program Studi Matematika , Universitas Negeri Padang," vol. 7, no. March, pp. 23745–23753, 2023.

[19] B. Sravani and M. M. Bala, "Prediction of student performance using linear regression," *2020 Int. Conf. Emerg. Technol. INCET 2020*, pp. 1–5, 2020, doi: 10.1109/INCET49848.2020.9154067.

[20] W. M. B. Dian Pramesti, "Perbandingan Prediksi Jumlah Transaksi Ojek Online Menggunakan Linear Regression dan Random Forest," vol. 7, no. 3, pp. 21–30, 2023.

[21] N. Karlsson and N. Karlsson, "Comparison of linear regression and neural networks for stock price prediction for stock price prediction," 2021.

[22] M. A. Muslim *et al.*, "Support Vector Machine (SVM) Optimization Using Grid Search and Unigram to Improve E-Commerce Review Accuracy," *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 8–15, 2020, doi: 10.52465/joscex.v1i1.3.

[23] F. Tang, Y. Wu, and Y. Zhou, "Hybridizing Grid Search and Support Vector Regression to Predict the Compressive Strength of Fly Ash Concrete," *Adv. Civ. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/3601914.

[24] M. M. Hameed, M. K. AlOmar, W. J. Baniya, and M. A. AlSaadi, "Incorporation of artificial neural network with principal component analysis and cross-validation technique to predict high-performance concrete compressive strength," *Asian J. Civ. Eng.*, vol. 22, no. 6, pp. 1019–1031, 2021, doi: 10.1007/s42107-021-00362-3.

[25] S. Saud, B. Jamil, Y. Upadhyay, and K. Irshad, "Performance improvement of empirical models for estimation of global solar radiation in India: A k-fold cross-validation approach," *Sustain. Energy Technol. Assessments*, vol. 40, no. June, p. 100768, 2020, doi: 10.1016/j.seta.2020.100768.

[26] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.

BIBLIOGRAPHY OF AUTHORS

Anik Vega Vitianingsih, S.Kom., MT, is a lecturer at the Informatics Department at Universitas Dr. Soetomo, Surabaya, Indonesia. The Author is interested in research topics in geographic information systems within the scope of spatial analysis using machine learning.

Resza Adistya Pangestu is an active final-semester student in t the Informatics Department at Universitas Dr. Soetomo, Surabaya, Indonesia. The Author is interested in the field of application of machine learning, namely machine learning.

Slamet Kacung, S.Kom., M.Kom, is a lecturer at the Informatics Department at Universitas Dr. Soetomo, Surabaya, Indonesia. The Author is interested in research topics in computer vision.

Anastasia Lidya Maukar holds the degrees of S.T., M.Sc.,M.MT is a lecturer in Industrial Management and Information Systems Development, President University, Indonesia. Her educational background includes a Master of Science degree in Industrial Management from the Institut Teknologi Sepuluh Nopember Surabaya (ITS) and a Master of Science degree in Information Systems Development from the University of Hertfordshire. Production Systems, Statistics, Databases, and Information Systems are the subject of research.

Agustinus Noertjahyana, S.Kom., M.MT, is a lecturer at the Informatics Engineering department at Universitas Petra, Surabaya, Indonesia. The Author is interested in research topics in Computer Network, Business Intelligence, and Security