



Tide Prediction in Prigi Beach using Support Vector Regression (SVR) Method

Tri Mar'ati Nur Utami^{1*}, Dian C. Rini Novitasari², Fajar Setiawan³, Nurissaidah Ulinnuha⁴,
Yuniar Farida⁵, Ghaluh Indah Permata Sari⁶

^{1,2,4,5}Department of Mathematics, Faculty of Science and Technology, UIN Sunan Ampel Surabaya, Indonesia

³ Meteorological, Climatological, and Geophysics Agency Surabaya, Indonesia

⁶ Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taiwan

Abstract.

Purpose: Prigi Beach has the largest fishing port in East Java, but the topography of this beach is quite gentle, so it is prone to disasters such as tidal flooding. The tides of seawater strongly influence the occurrence of this natural event. Therefore, information on tidal level data is essential. This study aims to provide information about tidal predictions.

Methods: In this case using the SVE method. Input data and time were examined using PACF autocorrelation plots to form input data patterns. The working principle of SVR is to find the best hyperplane in the form of a function that produces the slightest error.

Result: The best SVR model built from the linear kernel, the MAPE value is 0.5510%, the epsilon is 0.0614, and the bias is 0.6015. The results of the tidal prediction on Prigi Beach in September 2020 showed that the highest tide occurred on September 19, 2020, at 10.00 PM, and the lowest tide occurred on September 3, 2020, at 04.00 AM.

Value: After conducting experiments on three types of kernels on SVR, it is said that linear kernels can predict improvements better than polynomial and gaussian kernels.

Keywords: Support Vector Regression, Prediction, Tides, Time Series, Prigi

Received February 2021 / **Revised** February 2021 / **Accepted** November 2021

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



INTRODUCTION

Indonesia is an archipelagic country whose almost 70% of its territory is the sea [1]. Therefore, to maximize marine resources, it is necessary to refer to information on oceanographic parameters. Oceanographic parameters include waves [2], tides [3], currents [4], and others that are influenced by weather in the sea area [5]. Tidal information itself is considered crucial because, generally, tides can affect the speed of ocean currents [6]. Tidal information is needed by natural disaster management agencies such as tsunamis and El Nino storms [7]. For coastal communities, tidal information also dramatically affects the survival of people who depend on the sea, one of which is the coastal community of Prigi [8].

Prigi Beach is one of the beaches located in Trenggalek Regency, East Java Province. Prigi Beach has the largest fishing port in Indonesia in East Java, but this beach has a reasonably sloping topography, so it is vulnerable to tidal flooding [9]. In 2016 the Prigi beach area experienced tidal flooding. This incident occurred because the tide reached a height of 2-3 meters above sea level (Source: Merdeka.com). Therefore, it is necessary to predict the waves so that this event does not occur in the future and make maximum use of marine resources on the coast of Prigi.

Making a tidal prediction system cannot be separated from time series analysis. One of the prediction methods, namely Support Vector Regression (SVR) [10]. The working principle of SVR is to find the best hyperplane in the form of a regression function by making the slightest possible error. In addition, SVR is a suitable method for data with non-linear characteristics [11].

* Corresponding author

Email addresses: trimaratinurutami12@gmail.com (Utami), diancrini@uinsby.ac.id (Novitasari), geomaritim@bmkg.go.id (Setiawan), nuris.ulinnuha@uinsby.ac.id (Ulinnuha), yuniar_farida@uinsby.ac.id (Farida), ghaluhips@gmail.com (Sari).

DOI: [10.15294/sji.v8i2.28906](https://doi.org/10.15294/sji.v8i2.28906)

A tidal prediction has previously been made using the Backpropagation method. The results obtained MSE 0.0035861 with 70% training data, five nodes on the hidden network, learning speed 0.9, and target error 0.01 [12]. Another study used the Autoregressive integrated moving average (ARIMA) method combined with a deep belief network to predict tides. MAPE results are 17.21%, so the prediction results are just at a good level [13].

Another study compared the SVR method with ANN to predict groundwater levels with limited and unlimited systems. This study explains that [14], the SVR method is superior to the ANN method for estimating groundwater levels for the next 1, 2, and 3 months. Another study using the SVR method to predict building energy in South China resulted in an MSE value of less than 0.01 using seven parameters [15]. Another study compared SVR with multiple linear regression in the effect of weather on rice productivity in Indonesia. This study revealed that [16], the SVR method is superior to RMSE 25,8503 and MAE 20,912.

From previous studies using SVR, it is known that SVR is suitable for prediction. But by processing the data at the beginning is also increasing the accuracy. It can use PACF autocorrelation plot is to check the time-series input data [17]. So in this study, using PACF in data processing to improve the accuracy of SVR.

Based on this background, the purpose of this study is to provide information about tidal predictions, in this case using the SVR method with processing the initial data for the formation of time series using PACF to predict tides on the Prigi coast. This research is expected to be a reference in predicting the tidal.

METHODS

Fillmissing

In the data collection process, blank data is often encountered which can be due to technical errors so that a calculation is needed to fill in the blank data by taking into account the time series data pattern. The interpolation method is suitable for filling blank data in time series [18]. The interpolation method can be formulated using Equation (1).

$$\frac{y-y_u}{y_v-y_u} = \frac{x-x_u}{x_v-x_u} \quad (1)$$

Where y is the order of data that contains blank values, y_u is the order of data before the data is blank, y_v is the order of data after the data is blank, x is the result of interpolated data, x_u is the order data before the data is blank, x_v is the order data after the data that is blank. The purpose of Equation (1) is that the value of the order data before the data is blank, the value of the order data after the data is blank, and the value of the blank data has an influence or relationship, this is called the time series pattern.

Partial Autocorrelation function (PACF)

The Partial Autocorrelation function (PACF) on lag- k is the correlation between x_t and x_{t+k} after the linear dependencies between x_t and x_{t+k} variables between $x_{t+1}, x_{t+2}, \dots, x_{t+k-1}$ are removed. Partial autocorrelation is used to measure the level of closeness (association) between x_t and x_{t+k} , if the effect of time lag is 1, 2, 3, ..., $k - 1$ is considered separate [19].

Support Vector Regression (SVR)

SVR is the development of the Support Vector Machine (SVM) method to solve regression cases [20]. The purpose of the SVR is to obtain the function $f(x)$ as a dividing line (hyperplane) in the form of a regression function according to input data. The linear function of the SVR method can be formulated using Equation (2) [21].

$$f(x) = w^T \varphi(x) + b \quad (2)$$

Where $f(x)$ is a regression function, w is a weight vector that has dimension 1, in other words w is a normal field, functions to balance errors with a hyperplane, $\varphi(x)$ is obtained from the mapping of low-dimensional input vectors to produce a point in a high dimensional feature space, b is the bias which is the position of the plane relative to the coordinate center [22].

Determining the parameter values of w and b becomes a quadratic programming problem. To overcome optimization problems with constraints it is called lagrange. The optimal solution can be solved by the lagrange multiplier equation which is formulated using Equation (3) [23].

$$-\frac{1}{2} \sum_{i=1}^l (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \langle \varphi(x_i) \cdot \varphi(x_j) \rangle - \sum_{i=1}^l (\alpha_i - \alpha_i^*) \varepsilon \quad (3)$$

From the process of deriving the formula in Equation (3), the main variables are α_i and α_i^* . The solution to this problem is derived from the vector w then substituted into the function $f(x)$.

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \cdot (\varphi(x_i) \cdot \varphi(x_j)) + b \quad (4)$$

The algorithm can be adjusted for non-linear regression problems by adding a kernel [24]. The observations in the SVR can be mapped to a higher dimension which has a linear structure, without regard to explicit mapping. The regression model for non-linear cases can be formulated in Equation (5) [25].

$$\begin{aligned} f(x) &= \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x_j) + b \\ f(x) &= \sum_{i=1}^l \beta K(x_i, x_j) + b \end{aligned} \quad (5)$$

Where β is the difference between α_i and α_i^* , $K(x_i, x_j)$ is a kernel trick that is often used in SVM and SVR methods.

Kernel

Linear and non-linear SVM problems can be solved by adding a formulated kernel function to Equation (6), (7) and (8).

a. Kernel Linear

$$x \cdot x^T \quad (6)$$

b. Kernel Polynomial

$$(x \cdot x^T + 1)^d \quad (7)$$

c. Kernel Gaussian

$$\exp\left(-\frac{1}{2\sigma^2} \|x - x_i\|^2\right) \quad (8)$$

The Parameter $d > 0$ is constant. $\|x - x_i\|$ is the euclidean distance and σ is the value of the SVR parameter [26]. The difference between the three kernel functions is in the function of their mapping to feature space. Each kernel has advantages and disadvantages in each particular case, so it is necessary to find the best kernel function to use in some instances by trial and error [27].

MAPE (Mean Absolute Percent Error)

MAPE is obtained from the calculation of predictive data with actual data expressed in percent. The MAPE value can be formulated using Equation (9) [28].

$$MAPE = \left(\frac{1}{n} \sum_{i=1}^n \left| \frac{e_i}{y_i} \right| \right) \times 100\% \quad (9)$$

Where e_i is the result of subtracting from the actual data value to i with the predicted data value to i , y_i is the predicted value of the data to i , and n is the amount of data. The value categories of MAPE can be seen in Table 1.

| MAPE | Category |
|-----------|-----------------------|
| <10% | Very Good Predictions |
| 10% - 20% | Good Predictions |
| 20% - 50% | Worthy Predictions |
| >50% | Bad Predictions |

Technical Research

The Flowchart of Tide Prediction in Prigi Beach using SVR by Figure 1.

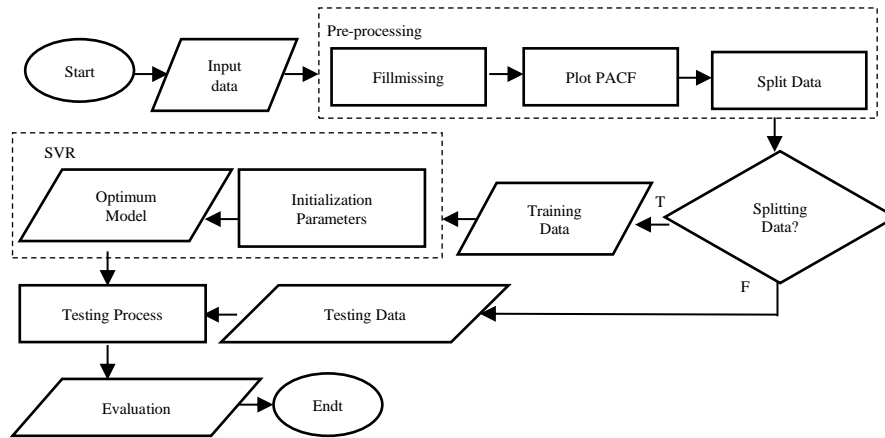


Figure 1. Flowchart of Tide Prediction System in Prigi Beach using SVR

The first step in predicting tides is to determine the input data. The data is then arranged into a time-series series using Equation (1), using the PACF autocorrelation plot, and then forming a series according to the time sequence. After that, the data is divided into training and testing processes with a ratio of 80%:20%. The parameters used in the training process are initialized. In the training process, the SVR model is obtained using Equation (5) which will later be used in the testing process. The last step is to evaluate by using a confusion matrix as in Equation (9).

RESULT AND DISCUSSION

The area used as the research location is Prigi Beach, located in Trenggalek Regency, East Java Province, as shown in Figure 2. The topography of the Prigi beach area is quite sloping, causing tidal flooding in the coastal area [9].



Figure 2. Location of Prigi Beach (Source: Google, 2020)

This study uses secondary data in tidal data from August 2nd to September 1st, in 2020, obtained from Meteorological, Climatological, and Geophysical Surabaya. This estimate uses hourly data with 720 data which means it is hourly data for 30 days. Table 2 shows an example of tidal data for the Prigi coast in rad (m).

Table 2. Tidal data samples of Prigi beach with missing value

| Date Time | August 16 th , 2020 | | | | | | |
|----------------|--------------------------------|-------|-------|------|-------|-------|-------|
| | 3:00 | 4:00 | 5:00 | 6:00 | 7:00 | 8:00 | 9:00 |
| Tidal (rad(m)) | 6.329 | 6.079 | 5.917 | | 5.794 | 5.874 | 6.039 |

The first process in studying tidal predictions at Prigi beach is data analysis. In this study, there is a preprocessing stage, namely fill-missing. Table 2 shows blank data at 6 AM, and it is necessary to fill in the empty data using Equation (1). The result is presented in Table 3.

Table 3. Data sample of fillmissing results

| Date | August 16 th , 2020 | | | | | | |
|----------------|--------------------------------|-------|-------|-------|-------|-------|-------|
| Time | 3:00 | 4:00 | 5:00 | 6:00 | 7:00 | 8:00 | 9:00 |
| Tidal (rad(m)) | 6.329 | 6.079 | 5.917 | 5.856 | 5.794 | 5.874 | 6.039 |

This study has four predictor variables (x) and one response variable (y). The time-series data in Table 4 represents the formation of the PACF plot. These data are helpful for the prediction process with (x_{t-5}) as input data and (x_t) as prediction targets. The PACF plot shows the 4th data pattern of Lag so that every 4th time-series pattern is deleted. Comparison of training data and test data is 80%:20%.

Table 4. Tidal time series data samples

| (x_{t-5}) | (x_{t-4}) | (x_{t-3}) | (x_{t-1}) | (x_t) |
|-------------|-------------|-------------|-------------|---------|
| 6.329 | 6.079 | 5.917 | 5.794 | 5.874 |
| 6.079 | 5.917 | 5.856 | 5.874 | 6.039 |

The second process is training or forecasting the measurement formulation model using Equation (5). The third process is model testing to determine the level of accuracy. The formulation of the SVR model uses several types of kernels, namely linear, polynomial and Gaussian kernels. Next, measure the error rate using Equation (9). The comparison of the error values for each kernel is presented in Table 5.

Table 5. Comparison results

| Kernel | ϵ | b | MAPE |
|------------|------------|--------|----------|
| Linear | 0.0614 | 0.6015 | 0.5510 % |
| Polynomial | 0.0614 | 4.6908 | 4.4160 % |
| Gaussian | 0.0614 | 6.2921 | 0.5565 % |

Visualization of actual data and prediction results can be done with linear, polynomial, and Gaussian kernels as seen in Figure 3, 4 and 5.

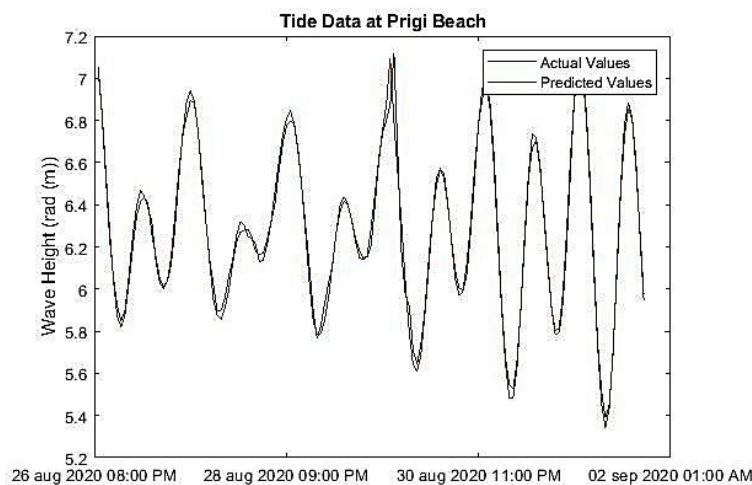


Figure 3. Plot of prediction with a linear kernel,

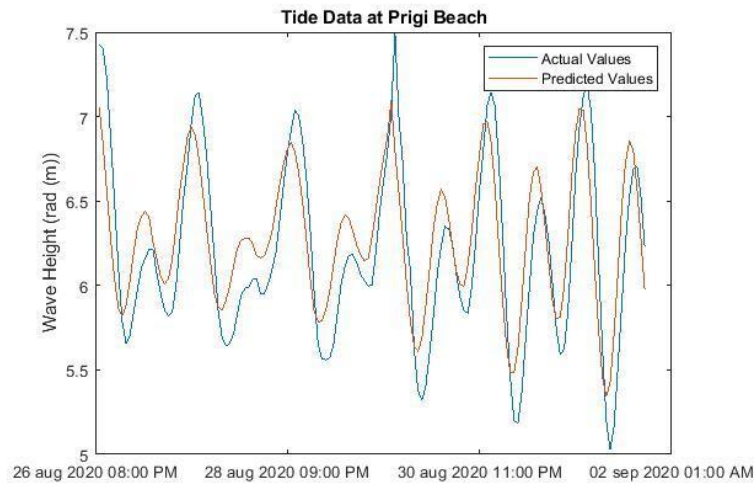


Figure 4. Plot of prediction with a polynomial kernel

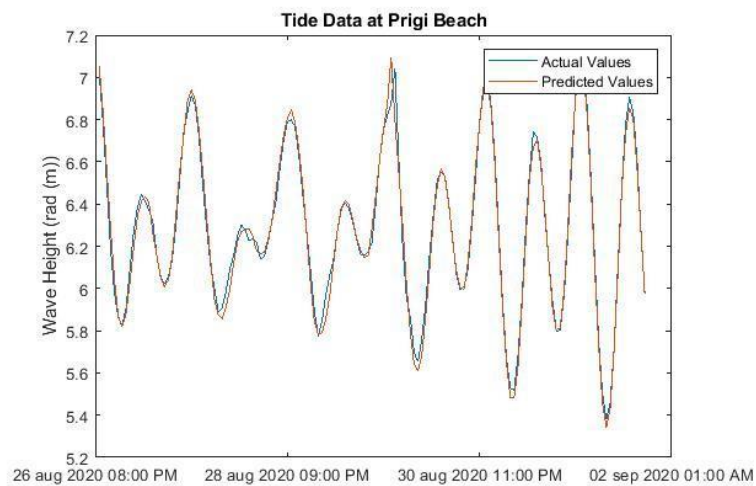


Figure 5. Plot of prediction with a gaussian kernel

In Table 5, the prediction results by comparing the three kernels, obtained of ϵ 0.0614, is the distance between the hyper tube and the hyperplane. Prediction results with three kernel trials show that the linear kernel produces the best accuracy. The resulting MAPE value is 0.5510%, the value of β [0.446;-0.335;-0.647;1.441] and the bias is 0.6015. Because MAPE is less than 10%, it is included in the very good category. Saputra, G. H., et al., in his research also revealed that the linear kernel has a fairly high prediction accuracy and produces relatively small errors in the prediction process [29].

The value of β is the result of the reduction of α_i and α_i^* . A non-linear kernel, namely the Gaussian kernel and polynomial, cannot display because of the difficulty of the calculation process. Smola stated in his research that “Even when evaluating $f(x)$ we need not compute explicitly” [32]. The error in this study is very small because the data in this study are pure tidal data without the influence of other components.

This study uses training data from August 2 at 7 PM to August 26, 2020, at 2 PM. The acquisition of test results starts from August 26th at 8 PM until September 1, 2020, at 6 PM, with one hour output to come. Based on the acquisition of the best model, the predicted tide results in Prigi Beach in September 2020 showed the highest tide occurred on September 19, 2020, at 10 PM, and the lowest tide occurred on September 3, 2020, at 4 AM. Because the outcome of this study only shows one outcome, even though the situation in society requires more tidal predictions for at least the next 3 hours. Therefore, the authors expect

further research to use a method that can display more than one output, for example, Multiple-Output Support Vector Regression (M-SVR) [30].

CONCLUSION

Based on the results of tidal research on the Prigi coast, it can be concluded that using the SVR method for tidal prediction with four previous tidal data parameters and a comparison of 3 kernels, namely Gaussian, polynomial and linear kernels, produces a very small MAPE. However, the smallest MAPE obtained by using a linear kernel is 0.5510%, so it is included in the very good category. Based on the acquisition of the best model, the predicted tide results in Prigi Pantai in September 2020 showed the highest tide occurred on September 19, 2020, at 10 PM, and the lowest tide occurred on September 3, 2020, at 4 AM.

REFERENCES

- [1] I. Chapsos and S. Hamilton, "Illegal fishing and fisheries crime as a transnational organized crime in Indonesia," *Trends Organ. Crime*, vol. 22, no. 3, pp. 255–273, 2019.
- [2] S. Darmawan, B. Irawan, C. Setianingsih, and M. A. Murty, "Design of detection device for sea water waves with fuzzy algorithm based on internet of things," *Proc. - 2020 IEEE Int. Conf. Ind. 4.0, Artif. Intell. Commun. Technol. IAICT 2020*, pp. 1–6, 2020.
- [3] D. C. R. Novitasari, F. Febrianti, and F. Setiawan, "Analisis Kecepatan Angin pada Pasang Surut Air Laut dengan Menggunakan Algoritma Forward-Backward dalam Hidden Markov Model di Wilayah Pelabuhan Tanjung Perak Surabaya," *J. Sains Mat. dan Stat.*, vol. 4, no. 1, pp. 26–35, 2018.
- [4] D. D. Kartika, D. C. R. Novitasari, and F. Setiawan, "Prediksi Kecepatan Arus Laut di Perairan Selat Bali Menggunakan Metode Exponential Smoothing Holt-Winters," *MathVisioN*, vol. 2, no. 1, pp. 12–17, 2020.
- [5] D. A. Adyanti, A. H. Asyhar, D. C. R. Novitasari, A. Lubab, and M. Hafiyusholeh, "Forecasts marine weather on java sea using hybrid methods: Ts-anfis," in *2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 2017.
- [6] T. Jansen, "Tinjauan Pengaruh Pasang Surut Terhadap Pola Arus Di Teluk Amurang, Sulawesi Utara," *TEKNO*, vol. 16, no. 70, 2018.
- [7] I. K. Missa, L. A. S. Laponi, and A. Wahid, "Rancang Bangun Alat Pasang Surut Air Laut Berbasis Arduino Uno dengan Menggunakan Sensor Ultrasonik HC-SR04," *J. Fis. Fis. Sains dan Apl.*, vol. 3, no. 2, pp. 102–105, 2018.
- [8] E. R. Heimisson and J. Avouac, "Analytical prediction of seismicity rate due to tides and other oscillating stresses," *Geophys. Res. Lett.*, vol. 47, no. 23, 2020.
- [9] I. Haqiqi, "Analisis Teknis Dan Ekonomis Kapal Tonda Di Pelabuhan Perikanan Nusantara (PPN) Prigi Trenggalek, Jawa Timur." Universitas Brawijaya, 2017.
- [10] Y. Yang, J. Che, C. Deng, and L. Li, "Sequential grid approach based support vector regression for short-term electric load forecasting," *Appl. Energy*, vol. 238, pp. 1010–1021, 2019.
- [11] G.-F. Fan, L.-L. Peng, W.-C. Hong, and F. Sun, "Electric load forecasting by the SVR model with differential empirical mode decomposition and auto regression," *Neurocomputing*, vol. 173, pp. 958–970, 2016.
- [12] N. Nikentari, N. Ritha, and T. Haryadi, "Prediksi Pasang Surut Air Laut Menggunakan Jaringan Syaraf Tiruan Backpropagation," *J. Sustain. J. Has. Penelit. dan Ind. Terap.*, vol. 7, no. 1, pp. 1–6, 2018.
- [13] M. Qin, Z. Li, and Z. Du, "Red tide time series forecasting by combining ARIMA and deep belief network," *Knowledge-Based Syst.*, vol. 125, pp. 39–52, 2017.
- [14] A. Mirarabi, H. R. Nassery, M. Nakhaei, J. Adamowski, A. H. Akbarzadeh, and F. Alijani, "Evaluation of data-driven models (SVR and ANN) for groundwater-level prediction in confined and unconfined systems," *Environ. Earth Sci.*, vol. 78, no. 15, p. 489, 2019.
- [15] Z. Ma, C. Ye, and W. Ma, "Support vector regression for predicting building energy consumption in southern China," *Energy Procedia*, vol. 158, pp. 3433–3438, 2019.
- [16] R. Yulistiani, I. Ramadhan, Q. Said, and M. M. Santoni, "Metode Regresi Linier Berganda dan SVR dalam Menentukan Tingkat Pengaruh Cuaca Terhadap Produktivitas Padi di Indonesia," *SEINASI-KESI*, vol. 1, no. 1, pp. 29–34, 2018.
- [17] A. Zakrani, A. Najm, and A. Marzak, "Support vector regression based on grid-search method for agile software effort prediction," in *2018 IEEE 5th International Congress on Information Science and Technology (CiSt)*, 2018.
- [18] Y. Pristyanto and I. Pratama, "Missing Values Estimation on Multivariate Dataset: Comparison of

- Three Type Methods Approach,” in *2019 International Conference on Information and Communications Technology (ICOIACT)*, 2019.
- [19] V. Demir, M. Zontul, and İ. Yelmen, “Drug Sales Prediction with ACF and PACF Supported ARIMA Method,” in *2020 5th International Conference on Computer Science and Engineering (UBMK)*, 2020.
- [20] X. Yuan, L. Cai-nian, X. Xiao-liang, J. Mei, and Z. Jian-guo, “A two-stage hog feature extraction processor embedded with SVM for pedestrian detection,” in *2015 IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 3452–3455.
- [21] Y. Chen *et al.*, “Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings,” *Appl. Energy*, vol. 195, pp. 659–670, 2017.
- [22] Q. N. Meng and X. Xu, “Price forecasting using an ACO-based support vector regression ensemble in cloud manufacturing,” *Comput. Ind. Eng.*, vol. 125, pp. 171–177, 2018.
- [23] C.-H. Huang, F.-H. Yang, and C.-P. Lee, “The strategy of investment in the stock market using modified support vector regression model,” *Sci. Iran.*, vol. 25, no. 3, pp. 1629–1640, 2018.
- [24] K. O. Achieng, “Modelling of soil moisture retention curve using machine learning techniques: Artificial and deep neural networks vs support vector regression models,” *Comput. Geosci.*, vol. 133, p. 104320, 2019.
- [25] J.-S. Chou and A.-D. Pham, “Nature-inspired metaheuristic optimization in least squares support vector regression for obtaining bridge scour information,” *Inf. Sci. (Ny)*, vol. 399, pp. 64–80, 2017.
- [26] B. Schölkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press, 2018.
- [27] C. López-Martín, M. Azzeh, A. Bou-Nassif, and S. Banitaan, “Upsilon-SVR Polynomial Kernel for Predicting the Defect Density in New Software Projects,” in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2018, pp. 1377–1382.
- [28] G. H. Saputra, A. H. Wigena, and B. Sartono, “Penggunaan Support Vector Regression dalam Pemodelan Indeks Saham Syariah Indonesia dengan Algoritme Grid Search,” *Indones. J. Stat. Its Appl.*, vol. 3, no. 2, pp. 148–160, 2019.
- [29] E. M. Priliani, A. T. Putra, and M. A. Muslim, “Forecasting Inflation Rate using Support Vector Regression (SVR) based Weight Attribute Particle Swarm Optimization (WAPSO),” *Sci. J. Inf.*, vol. 5, no. 2, pp. 118–127, 2018.
- [30] G. Melki, A. Cano, V. Kecman, and S. Ventura, “Multi-target support vector regression via correlation regressor chains,” *Inf. Sci. (Ny)*, vol. 415, pp. 53–69, 2017.