**Prediction of Sea Water Installation Using the Support Vector Regression (SVR) Method**

**(Case Study: Prigi Beach)**

**Tri Mar’ati Nur Utami1, Dian C. Rini Novitasari2, Fajar Setiawan3, Nurissaidah Ulinnuha4, Yuniar Farida5**

1,2,4,5 Department of Mathematics, Faculty of Science and Technology,

UIN Sunan Ampel Surabaya, Indonesia

3 Meteorogical, Climatological, and Geophysics Agency Surabaya, Indonesia

Email: 2diancrini@uinsby.ac.id

**Abstract**

Indonesia lies on the equator so that the wind, currents, tides and waves of the sea are very large. The phenomenon of tidal sea water is one of the important references in oceanographic phenomena. The tidal sea water has a high effect on the residents of the coast of Prigi, whose livelihoods depend on the sea. The data used was the secondary data in the form of hourly tidal data of Prigi beach from 2 August to 1 September 2020 obtained from the Meteorogical, Climatological, and Geophysics Agency Surabaya. This study used the Support Vector Regression (SVR) method. The SVR method can learn various data patterns, dimension requirements, and is easy to implement. This study divided the training data and test data with a ratio of 80% and 20%, and used 5 predictor variables and 1 response variable This study also compared three kernels, namely gaussian, polynomial and linear kernels. The best prediction results were obtained at a linear kernel with the epsilon value of 0.0614, the bias of 0.6021 and produced a MAPE of 0.5301%, so it can be said that the model used was very good because it was less than 10%.

**Keywords**: Support Vector Regression, SVR, Prediction, Tides, Time Series

1. **INTRODUCTION**

Indonesia is known as a country of archipelago, where two-thirds of its areas consist of sea [1]. Viewed from a geographical point of view, Indonesia is located on the equator so that the wind, currents, tides and waves of the sea are very large [2]. Therefore, coastal areas in Indonesia often have disasters caused by the sea, one of which is tidal flooding. In 2016, a tidal flooding took place in a coastal area of Prigi in Trenggalek district, East Java province. This incident occurred because the tidal waves reached a height of 2-3 meters above the sea level [3-4].

Prigi Beach has a fairly sloping topography, so that it is one of the factors which causes tidal flooding to occur in coastal areas [5]. This incident certainly caused considerable losses for coastal residents. Many ships were damaged, food stalls and even houses near the coast were flooded. If this happens again, it will be unsettling for coastal residents, because the majority of the residents in Prigi coastal area make their living on the sea. Therefore, it is necessary to have references regarding oceanographic phenomena. The phenomenon of tidal sea water is one of the important references in oceanographic phenomena.

Making tidal predictions cannot be apart from time series analysis. Many studies use time series analysis as a method to predict. One of them is the Support Vector Regression (SVR) method. In 1999 Vijayakumar Si Wu implemented SVR to solve forecasting problems in the form of regression [6]. The SVR method can be used to predict with a fairly low error rate. In addition, SVR is suitable for non-linear data [7].

There are several references in this study, including a research that discusses tidal predictions using the Backpropagation method. In this study, 1000 data were used with a comparison of 70: 30 of training data and test data and a modeling network with 5 nodes in a hidden network. The learning rate value of 0.9 and the target error of 0.01, resulting in an MSE value of 0.0079440 [8]. In addition, there is also a research using the Artificial Neural Networks (ANN) method to predict tides. In this study, the RMSE was 0.15 with η = 0.1 and α = 0.8 with 2000 iterations [9]. The weakness of the Neural Networks method is of course it uses a lot of iterations so it takes a long time.

Another study compared the SVR method with ANN in the case of prediction of groundwater levels with a limited and unlimited system. This study explained that the SVR method was superior to the ANN method for estimating groundwater levels for the next 1, 2, and 3 months [10]. The SVR method has been compared with the Auto Regressive (AR) and Moving Average (MA) methods in link hold prediction problems on the network. In that study, it was stated that SVR had the smallest error compared to AR and MA [11]. Research using the SVR method to predict building energy in South China resulted in an MSE value of less than 0.01 using seven parameters [12]. Another study used SVR to focus on queries to make multi-document summarization based on the significance of sentence positions. SVR estimated sentence weight in a set of documents to be used as a summary through predefined features. The results of the test obtained an average precision value of 0.0580 and a recall of 0.0590. Measurements with ROUGE-2 and ROUGE obtained values of 0.0997 and 0.1019 [13]. Some of the explanations from previous research, the SVR method is a good method to use in prediction. This is because the SVR method can learn various data patterns, dimensional requirements, and is easy to implement [14].

Based on the problems and references that have been described regarding the importance of predicting tidal sea water on the coast of Prigi, and the superiority of the SVR method in predicting in various fields such as engineering, energy, and natural resources, this study made the prediction on the tidal sea water at Prigi Beach using the SVR method. This research is expected to be used as a reference in predicting tidal sea water.

1. **METHODS**

The steps taken in this study starts from the pre-processing, data processing, making a model, prediction, and evaluation. The steps taken in this study are described by the theoretical framework in accordance with Figure 1.



Figure 1. Theoretical framework of SVR

* 1. **Fillmissing**

Not all data has a perfect value, sometimes data has blank values that must be filled. The interpolation method is suitable for filling blank data in time series [15]. The interpolation method can be formulated using Equation (1).

 (1)

Where is the order of data that contains blank values, is the order of data before the data is blank, is the order of data after the data is blank, is the result of interpolated data, is the order data before the data is blank, 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.

* 1. **Support Vector Regression (SVR)**

SVR is the development of the Support Vector Machine (SVM) method to solve regression cases. The purpose of the SVR is to obtain the function 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) [16-18].

 (2)

Where is a regression function, is a weight vector that has dimension l , in other words w is a normal field, functions to balance errors with a hyperplane, is obtained from the mapping of low-dimensional input vectors to produce a point in a high dimensional feature space, is the bias which is the position of the plane relative to the coordinate center [19-21].

Determining the parameter values of and 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) [22].

 (3)

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

 (4)

The algorithm can be adjusted for non-linear regression problems by adding a kernel. 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) [23-24].

 (5)

Where β is the difference between and , is a kernel trick that is often used in SVM and SVR methods.

* 1. **How to Reconciliation and Citations**

Linear and non-linear SVM problems can be solved by adding a formulated kernel function to Equation (6), (7) and (8) [25-26].

1. Kernel Linear

 (6)

1. Kernel Polynomial

 (7)

1. Kernel Gaussian

 (8)

* 1. **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) [27-28].

 (9)

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

Table 1. Value category of MAPE

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

1. **RESULT AND DISCUSSION**

The area used as a research site was Prigi beach, which is located in Trenggalek district, East Java province as shown in Figure 2. The topography of the Prigi coastal area is quite gentle, causing tidal flooding in the coastal areas [5].



**Prigi Beach, Indonesia**

**Latitude**: -8.28674 **| Longitude:** 111.72586

**BALI**

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

This study used secondary data in the form of tidal data from August 2 to September 1 2020 obtained from the Meteorogical, Climatological, and Geophysics Agency Surabaya. This forecasting used data per hour with a total of 720 data, which means that it is an hourly data of 30 days. Table 2 shows a sample of tidal data of the Prigi coast in rad (m).

Table 2. Tidal data samples of Prigi beach

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

The first process in studying the tidal prediction at Prigi beach was data analysis. The input data was in the form of tidal data. In this study, there is a preprocessing step, where there is blank data in Table 2 which is a sample data, so it is necessary to do fillmissing to fill in the blank data using Equation (1). The data sample is presented in Table 3.

Table 3. Data sample of fillmissing results

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

In this study, 5 predictor variables and 1 response variable were used. The time series data in Table 4 is obtained from the previous 5 tides and the current tides . The data is used for the prediction process with as input data and as the prediction target. The training data and test data are distributed in the ratio of 80% and 20%.

Table 4. Tidal time series data samples

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| 6.329 | 6.079 | 5.917 | 5.856 | 5.794 | 5.874 |
| 6.079 | 5.917 | 5.856 | 5.794 | 5.874 | 6.039 |

The second process was training or forecasting formulation model measurement was formulated by Equation (5). The third process was testing the model to determine the level of accuracy. The SVR model formulation used several types of kernels, such as linear, polynomial, and Gaussian. The error rate measurement was formulated by Equation (9). The comparison of error values for each kernel is presented in Table 5. According to the error values obtained, the linear kernel achieved the best result, with a MAPE value of 0.5301.

Table 5. Comparison results

|  |  |  |  |
| --- | --- | --- | --- |
| Kernel  |  |  | MAPE |
| Linear  | 0.0614 | 0.6021 | 0.5301% |
| Polynomial  | 0.0614 | 5.5507 | 1.8998% |
| Gaussian  | 0.0614 | 6.2982 | 0.5900% |

As seen from Table 5 on the prediction results by comparing the three kernels,

ε of 0.0614 is obtained, where ε represents the distance between the hyper-tube and the hyperplane. The best tidal prediction using SVR is obtained using a linear kernel where the MAPE value is 0.5301, with a value of 0.9048 and bias () of 0.6021. Because the MAPE obtained is less than 10%, it can be said that the model used is very good.

The use of non-linear kernels, namely Gaussian kernels and polynomial values of cannot be displayed in Matlab. This is due to the difficulty of the calculation process. Smola stated in his research that “Even when evaluating we need not compute β explicitly” [16]. The error in this study is very small because the data obtained in this study is pure tidal data without any influence from other components.



(a)

 

(b)



(c)

Figure 2. (a) Plot of prediction with a linear kernel, (b) Plot of prediction with a polynomial kernel, (c) Plot of prediction with a gaussian kernel

The graphs in Figures 2 (a), (b) and (c) are the results of tidal predictions with linear, polynomial, and Gaussian kernels. The training data starts from August 2 at 07:00 PM to August 26 2020 at 02:00 PM. So that the test results are obtained from August 26 at 08:00 PM to September 1 2020 at 06:00 PM. This research uses a time series pattern, but the data pattern is not yet known for being optimal, so it is hoped that further research can use ACF and PACF auto-correlation plots to form time series patterns, so that the optimal time series patterns are obtained [29]. In addition, this research predicts the sea tides one hour in advance. In fact, the community requires more tidal predictions for at least 3 hours in advance. Therefore, it is hoped that further research will use a method that can display more than one output, for example multiple-output support vector regression (M-SVR) [30].

1. **CONCLUSION**

Based on the results of the tidal research on Prigi beach, it can be concluded that the use of the SVR method for tidal prediction with the previous 5 tidal data parameters and comparison of 3 kernels namely gaussian, polynomial and linear kernels produced a very small MAPE with a small difference between the kernels. However, the smallest MAPE was obtained using a linear kernel, equal to 0.5301%, so it is said that the prediction was very good because it was less than 10% with the highest and lowest sea level predictions of 7.1297 and 5.3872.

1. **REFERENCES**

[1] Saragih, H. M., & Siregar, R. (2018, July). Indonesian national power to achieve the global maritime fulcrum. In *2018 3rd International Conference on Education, Sports, Arts and Management Engineering (ICESAME 2018)* (pp. 47-49). Atlantis Press.

[2] Talarosha, B. (2005). Menciptakan kenyamanan thermal dalam bangunan. *Jurnal Sistem Teknik Industri*, *6*(3).

[3] Suci, C. W., Wijaya, A. R., Santoso, A., & Kusumaningrum, I. K. (2020, April). Fe leaching in the sludge sediment of the Prigi beach with Tessier-Microwave method. In *AIP Conference Proceedings* (Vol. 2231, No. 1, p. 040025). AIP Publishing LLC.

[4] Sari, D. E. (2017). *Pendampingan masyarakat Desa Tasikmadu dalam upaya pengurangan risiko bencana hidrometeorologi melalui pemetaan partisipatif tata ruang desa dengan SIG (Sistem Informasi Geografis) dan SID (Sistem Informasi Desa) sebagai media penyadaran masyarakat* (Doctoral dissertation, UIN Sunan Ampel Surabaya).

[5] Ermawan, R. W. (2008). Kajian Sumberdaya Pantai untuk Kesesuaian Ekowisata di Pantai Prigi, Kabupaten Trenggalek, Provinsi Jawa Timur. *Skripsi. Fakultas Peikanan dan Ilmu Kelautan, Institut Petanian Bogor*.

[6] Vijayakumar, S., & Wu, S. (1999, June). Sequential Support Vector Classifiers and Regression. In *IIA/SOCO*.

[7] Kavousi-Fard, A., & Kavousi-Fard, F. (2013). A new hybrid correction method for short-term load forecasting based on ARIMA, SVR and CSA. *Journal of Experimental & Theoretical Artificial Intelligence*, *25*(4), 559-574.

[8] Nikentari, N., Ritha, N., & Haryadi, T. (2018). Prediksi Pasang Surut Air Laut Menggunakan Jaringan Syaraf Tiruan Backpropagation. *Jurnal Sustainable: Jurnal Hasil Penelitian dan Industri Terapan*, *7*(1), 1-6.

[9] Lee, T. L., & Jeng, D. S. (2002). Application of artificial neural networks in tide-forecasting. *Ocean Engineering*, *29*(9), 1003-1022.

[10] Mirarabi, A., Nassery, H. R., Nakhaei, M., Adamowski, J., Akbarzadeh, A. H., & Alijani, F. (2019). Evaluation of data-driven models (SVR and ANN) for groundwater-level prediction in confined and unconfined systems. *Environmental Earth Sciences*, *78*(15), 1-15.

[11] Priyadarshini, D., Acharya, M., & Mishra, A. P. (2011). Link load prediction using support vector regression and optimization. *International Journal of Computer Applications*, *24*(7).

[12] Ma, Z., Ye, C., & Ma, W. (2019). Support vector regression for predicting building energy consumption in southern China. *Energy Procedia*, *158*, 3433-3438.

[13] Fanani, A., Farida, Y., Arhandi, P. P., Hidayat, M. M., Muhid, A., & Montolalu, B. (2019). Regression model focused on query for multi documents summarization based on significance of the sentence position. *Telkomnika*, *17*(6), 3050-3056.

[14] Suwanto, S., & Novitasari, D. C. R. (2020). Forecasting Solar Activities based on Sunspot Number Using Support Vector Regression (SVR). *JPSE (Journal of Physical Science and Engineering)*, *5*(1), 6-10.

[15] Moon, T., Hong, S., Choi, H. Y., Jung, D. H., Chang, S. H., & Son, J. E. (2019). Interpolation of greenhouse environment data using multilayer perceptron. *Computers and Electronics in Agriculture*, *166*, 105023.

[16] Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, *14*(3), 199-222.

[17] Vapnik, V., Golowich, S. E., & Smola, A. (1997). Support vector method for function approximation, regression estimation, and signal processing. *Advances in neural information processing systems*, 281-287.

[18] Chen, Y., Xu, P., Chu, Y., Li, W., Wu, Y., Ni, L., ... & Wang, K. (2017). Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings. *Applied Energy*, *195*, 659-670.

[19] Lu, Y., & Roychowdhury, V. (2008). Parallel randomized sampling for support vector machine (SVM) and support vector regression (SVR). *Knowledge and Information Systems*, *14*(2), 233-247.

[20] Parveen, N., Zaidi, S., & Danish, M. (2016). Support vector regression model for predicting the sorption capacity of lead (II). *Perspectives in Science*, *8*, 629-631.

[21] Meng, Q. N., & Xu, X. (2018). Price forecasting using an ACO-based support vector regression ensemble in cloud manufacturing. *Computers & Industrial Engineering*, *125*, 171-177.

[22] Shokry, A., Audino, F., Vicente, P., Escudero, G., Moya, M. P., Graells, M., & Espuña, A. (2015). Modeling and simulation of complex nonlinear dynamic processes using data based models: application to photo-Fenton process. In *Computer Aided Chemical Engineering* (Vol. 37, pp. 191-196). Elsevier.

[23] Wang, Y., Wang, B., & Zhang, X. (2012). A new application of the support vector regression on the construction of financial conditions index to CPI prediction. *Procedia Computer Science*, *9*, 1263-1272.

[24] Kavaklioglu, K. (2011). Modeling and prediction of Turkey’s electricity consumption using Support Vector Regression. *Applied Energy*, *88*(1), 368-375.

[25] Kivinen, J., Smola, A. J., & Williamson, R. C. (2004). Online learning with kernels. *IEEE transactions on signal processing*, *52*(8), 2165-2176.

[26] Novitasari, D. C. R., Rohayani, H., Junaidi, R., Setyowati, R. D., Pramulya, R., & Setiawan, F. (2020, March). Weather Parameters Forecasting as Variables for Rainfall Prediction using Adaptive Neuro Fuzzy Inference System (ANFIS) and Support Vector Regression (SVR). In *Journal of Physics: Conference Series* (Vol. 1501, No. 1, p. 012012). IOP Publishing.

[27] De Myttenaere, A., Golden, B., Le Grand, B., & Rossi, F. (2016). Mean absolute percentage error for regression models. *Neurocomputing*, *192*, 38-48.

[28] Priliani, E. M., Putra, A. T., & Muslim, M. A. (2018). Forecasting Inflation Rate using Support Vector Regression (SVR) based Weight Attribute Particle Swarm Optimization (WAPSO). *Scient. J. Inform*, *5*(2), 118-127.

[29] Berryman, A., & Turchin, P. (2001). Identifying the density‐dependent structure underlying ecological time series. *Oikos*, *92*(2), 265-270.

[30] Bao, Y., Xiong, T., & Hu, Z. (2014). Multi-step-ahead time series prediction using multiple-output support vector regression. *Neurocomputing*, *129*, 482-493.