Jurnal Mandiri IT ISSN 2301-8984 (Print)
Vol. 11 No. 1, July (2022), pp. 34-40

Analysis of Using Support Vector Regression (SVR) Algorithm to Predict The Occurrence of Sea Tides in Tanjung Medang, Riau by Saas Storage in Cloud Computing

Dinda Faatihah Ramadhani Putri[1], Ita Arianti[2], Julydio Windu Nugroho[3], Silviya[4]
Universitas Pendidikan Indonesia

A R T I C L E I N F O

Article history:

Received May 11, 2022
Revised Jun 13, 2022
Accepted Jul 30, 2022

Kata Kunci

Cloud Computing
Data Mining
SaaS
Support Vector Regression
Tides

ABSTRACT

Tides in the waters of Tanjung Medang, Riau, have an important role of lives in communities around the coast, one of them is to help a fishermen when they want to go to sea. Cloud computing is an online based data storage that can make various human activities easier. The method that we used for this research is Support Vector Regression (SVR) with Cost (C) and Gamma parameters. The purpose of this research is to obtain the data accuracy and determine the RMSE value of the model in data processing process. The data will be stored through one of the cloud computing technologies, namely Software As A Service (SaaS) storage. Based on the research result, the Support Vector Regression (SVR) model was good enough to be used in processing tidal prediction data used Gamma parameter with a gain value of 100 and RMSE value of 0.451896.

Corresponding Author:

Dinda Faatihah Ramadhani Putri,
Program Studi Sistem Informasi Kelautan,
Universitas Pendidikan Indonesia,
Jl. Dr. Setiabudi No.229, Isola, Kec, Sukasari, Kota Bandung, Jawa Barat 40154 .
Email: julydiown@upi.edu

1. INTRODUCTION

Weather and climate is a natural phenomena in the atmosphere with their impact or influence on the environment. Weather or climate can increase drastically if natural conditions are not balanced, one of them is the tidal wave of sea water(Supriyadi & Siswanto, 2019). Tides are the phenomenon of rising and falling sea water according to the schedule for each period, which is about 24 hours and 50 minutes. There are two types of tides, first, spring tides, the tides with gravitational attraction of the sun and moon on sea water in the same direction. Second, neap tides, the tides that have a small amplitude due to position of the sun and the moon which stand perpendicular, because of a small attraction (Asyhar et al., 2018; Wakhid & Syahbuddin, 2018).

The waters in Tanjung Medang area, located in North Rupat District, Bengkalis Regency, Riau Province, located at coordinates 2°1'N 101°34' East Longitude with the island being surrounded by the Malacca Strait. The Tanjung Medang area, Riau, has a tropical climate with relatively good water conditions and high rainfall. The tides in waters of Tanjung Medang, Riau, have an important role in of lives in communities around the coast, one of them is to help fishermen when they want to go to sea. With a prediction or forecast of when the tide will occur, the data can be used as a reference for local communities in carrying out an activity, especially activities related to the sea. Forecasting is

an activity of estimating an event in the future based on data obtained at this time and data from the previous years (Arifin et al., 2021; Nikentari et al., 2018).

The selection of methods for making a predictions (forecasting), must be in accordance with several supporting aspects, namely data patterns, system models, time and the level of accuracy of the data obtained (Utami et al., 2021), (Pamungkas, 2018). One such method is Support Vector Regression (SVR). SVR is a method that able to create forecasting or prediction results that are quite accurate because SVR is able to solve the problem of overfitting (Ardhiyah, 2019),(Putri et al., 2020). Prediction data for the occurrence of sea tides in the Tanjung Medang area, Riau, can be stored in one of the internet-based software because the use of technology in the current digital era can help make various kinds of human activities easier. The data storage used is cloud computing technology or called cloud computing(Khaliq, 2021; Muttaqin et al., 2022).

Cloud computing is a computing model, using applied resources, namely computing power, storage, network and software as one of the services on the internet network using remote (Pratama, 2021). Cloud computing can be freely accessed, and users are able to access all service and information contained in the storage freely (Anharudin et al., 2022). The software name is, Software As A Service (SaaS) storage, which is a cloud computing storage container and provides cloud applications and can be use or consume by consumers when connected to the network connection (Suprayitno, 2018).

To determine the accuracy of the data and predict the occurrence tides in Tanjung Medang area, Riau, this research used Support Vector Regression (SVR) algorithm. The principle of SVR algorithm is to find the best hyperplane with very small errors. The purpose of this research is to obtain data accuracy and to determine the RMSE value of the model used in the data processing, it can be analyzed that the model used is appropriate for displaying predictive data or forecasting the tides in the Tanjung Medang area, Riau, and the data will be stored through one of the cloud computing technologies, namely Software As A Service (SaaS) storage.

2. METHOD

The research and the data collection of tidal waves was carried out in the Tanjung Medang area, Riau, using time series data and using the Support Vector Regression (SVR) algorithm. The data collection of tidal waves will be used to determine a prediction or forecast the occurrence of tides in the future. Here are a several stages of data processing used in research and visualized using a flowchart as shown below.

ISSN 2301-8984 (Print)

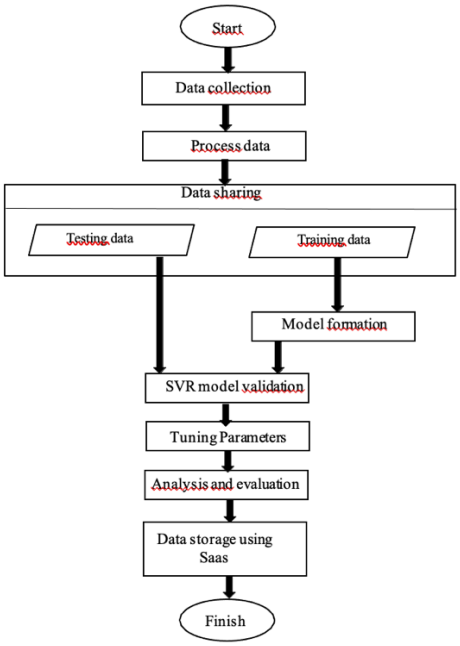

Figure 1. Research Flowchart

a. Data Collection

The dataset used is tidal waves data in Tanjung Medang, Riau, obtained from the website of the Naval Hydrographic and Oceanographic Center (PUSHIDROSAL). This dataset is based on time series data, which is annual data for three consecutive years, starting from 2018 to 2021, with 11,487 tidal data recorded. The arameters in this research used Cost (C) and Gamma. Tidal data in Tanjung Medang area, Riau in 2018-2021 can be seen in Figure 2.

Time	WaterLevel
2018-02-23 17:17:09	7607.000000
2018-02-26 15:05:12	-0.092000
2018-02-26 16:00:11	-0.087000
2018-02-26 17:00:18	-0.082000
2018-02-27 04:57:59	-0.088000
...	...
2021-01-01 13:00:00	1073.000000
2021-01-01 14:00:00	0.689583
2021-01-01 15:00:00	1071.000000
2021-01-01 16:00:00	1360.000000
2021-01-01 17:00:00	1904.000000

11487 rows × 1 columns

Figure 2. Raw Data for Sea Tides in Tanjung Medang, 2018-2021

b. Process Data

Data scaling was carried out after all of data tides are collected, namely to provide the same range of values for each data. The next step is, split data into two, namely training data (training data/train data) and test data (testing data/test data). Train data is, the data used to train the algorithm for find a suitable model with 70% data. Test data is the data used to test the results of determining the model from the train data by 30%.

c. SVR Model Validation

Support Vector Regression (SVR) is an algorithm from the development of the Support Vector Machine (SVM) algorithm. SVR is usually used to solve regression cases. The purpose of SVR algorithm is, get the value of the function (x) as a dividing line or hyperplane in the form of regression. The linear function of the SVR method is formulated using the following equation.

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

Description:
w : weight vector
T : a function mapping of dimension
b : bias

It is known that (x) is a regression function, w is a plane normal, which serves to balance the error with the hyperplane. The value (x) is obtained based on the mapping of the low dimensions of the input vector in producing a bias which is the position of the plane relative to a coordinate. The SVR algorithm has 3 kernels, namely Linear, Polynomial and Radial Basis Function (RBF). This research uses a linear kernel because this kernel is able to process data that has been classified by hyperline.

d. Tuning Parameters

The formation of the prediction model requires a parameter tuning method and be able to produce parameter values with maximum accuracy in prediction. Grid search is used to obtain the optimum parameter values in the formation of predictions. This method applies when the desired maximum value can be known within the limits that have actually been determined by the upper and lower limits on the independent variables (Gunawan et al. 2020).

e. Analysis and Evaluation

Model evaluation is carried out the final stage, when all data has been tested, namely by evaluating the model used. Evaluation of the prediction model used is RMSE (Root Mean Square). RMSE is used to measure the quality of the model on the processed data based on the error estimation value (Putri et al., 2020). In addition, RMSE can find out the magnitude of the deviation that occurs between the estimated value and the mean square error value which is used to calculate the error or error that comes from the two tidal models of the experimental results (Wilks, 1995). The value of the coefficient of determination is used to determine the magnitude of the influence of the variable x (input) with the variable y (output). If the RMSE value is getting smaller and the coefficient of determination is getting bigger, then the model used is good. The RMSE formula can be seen in the equation below.

$$\text{RMSE}=\sqrt{\sum} \; ei2 \; n \; i=1 \; n$$

Description:
n : prediction amount of data
i : data order
e : error between actual and prediction values

f. Data Storage Using Saas Method

The data from the tidal waves prediction test in Tanjung Medang area, became an embedded into a software-based cloud computing container with high storage speed, namely SaaS (Software as a Service) and can be accessed by the public if connected to the internet network. For example of cloud computing-based storage service provider is Amazon S3, a website service interface that offers access to a scalable, reliable, fast and inexpensive data storage infrastructure. Amazon S3 Bucket is a container for storing an object, such as a folder or directory (Andreozzi et al. 2008). The dataset of SVR model test results in tidal prediction will be stored on the Amazon S3 Bucket.

3. RESULT AND DISCUSSION

The dataset tidal waves is time series data from 2018 to 2021. After the dataset has been split into two data, namely train data and test data, the model used to test tidal predictions uses the SVR

ISSN 2301-8984 (Print)

algorithm with Linear kernel. After predicting modeling, parameter tuning is performed using Gamma parameters. To calculate the level of error or error in this study, it is calculated using the Root Mean Square (RMSE) formula. If the RMSE value is getting smaller, then the selection of these parameters can be said to be good. The Gamma parameters tested were carried out in the range of values from 0.01 to 100. The following is a comparison of the best Gamma values seen in Table 1.

Table 1. Best Parameters Using RMSE Value

No	Parameter *Gamma*	RMSE (%)
1	0.01	1.023697
2	0.1	1.070205
3	1	0.980224
4	10	0.621802
5	100	0.451896

Based on Table 1, it is known the best Gamma parameter at value of 100. The value of this parameter can be used to evaluate SVR algorithm model testing on tidal waves datasets in Tanjung Medang, Riau. From the several trials, obtained an accuracy of 0.89 or equivalent to 89%. The following is a graph of the test results of several Gamma parameters so as to produce different RMSE and R2 values, which can be seen in Figure 3.

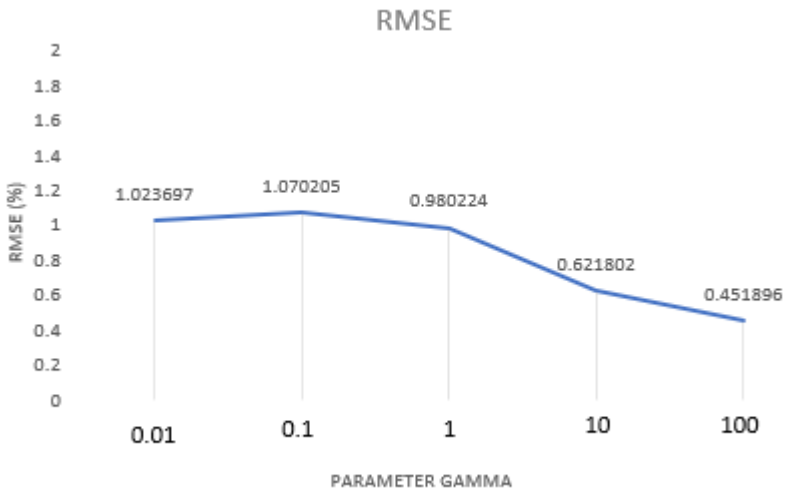

Figure 3. RMSE Value Graph with Parameter Testing

The best RMSE prediction model is the one with the smallest value. The value of the coefficient of determination/R2 has a relationship with the RMSE to determine the suitability of a model. In Figure 3, it can be seen that the tide prediction model has the best value of 0.451896. It can be concluded that the Support Vector Regression (SVR) model is good enough to be used to process data in predicting tides.

The results of the validation of using the Support Vector Regression (SVR) model to process tidal waves prediction, and uploaded to the AWS S3 Bucket online database in the Asian region. The entire tidal waves prediction model dataset stored in the AWS Bucket. In the public access configuration the dataset is managed as “object can be public” in the sense that the bucket is not public information, but anyone with the appropriate permissions can grant public access to the object. Therefore, the public can access the tidal prediction model dataset through permission from the

account authority, but the data cant be edited or processed or managed, so the value of dataset is safe and accurate.

4. CONCLUSION

The conclusions that can be drawn, The prediction model for the occurrence of sea tides using the Support Vector Regression (SVR) method was formed with fairly good results. The best results were obtained by tuning the parameters with Gamma parameter with a gain value of 100 and RMSE value of 0.451896. This value has a relationship in determining the suitability of a model, so the Support Vector Regression (SVR) model is good enough to be used in processing tidal prediction data. The valid data are uploaded to the AWS S3 Bucket online cloud computing database service in the Asian region. Datasets are managed as “objects can be public” which means that the bucket is not public information, but anyone with the appropriate permissions can grant public access to the objects in order to keep the dataset secure and authentic.

ACKNOWLEDGEMENTS

The authors thanks to Allah SWT. who has given an infinite blessings for completing this research, and for the authors lecturers who helped and guided us until making this paper completed.

REFERENCES

Anharudin, A., Siswanto, S., & Syakira, R. M. (2022). Rancang Bangun Data Storage System berbasis Web Dengan Metode Extreme Programming. *Jurnal Tekno Kompak*, *16*(1), 123–135.

Ardhiyah, N. (2019). *Prediksi awal waktu shalat berdasarkan titik belok kecerahan langit menggunakan metode Support Vector Regression dan Restricted Cubic Spline*. UIN Sunan Ampel Surabaya.

Arifin, W. A., Ariawan, I., Rosalia, A. A., Sasongko, A. S., Apriansyah, M. R., & Satibi, A. (2021). MODEL PREDIKSI PASANG SURUT AIR LAUT PADA STASIUN PUSHIDROSAL BAKAUHENI LAMPUNG MENGGUNAKAN SUPPORT VECTOR REGRESSION. *Jurnal Kemaritiman: Indonesian Journal of Maritime*, *2*(2), 139–148.

Asyhar, A. H., Febrianti, F., & Fajriyah, N. R. (2018). AnalisisModelExponentialSmoothing Terhadap Prediksi Pasang Surut Air Laut di Wilayah Pelabuhan Tanjung Perak Surabaya. *Jurnal Sains Matematika Dan Statistika*, *4*(1), 10–15.

Khaliq, K. F. (2021). PENGAMANAN DATA AKTA DENGAN METODE AES BERBASIS CLOUD COMPUTING. *JURNAL TEKNOLOGI DAN ILMU KOMPUTER PRIMA (JUTIKOMP)*, *4*(1), 509–512.

Muttaqin, M., Sinambela, M., Andryanto, A., Parewe, A. M. A. K., Sari, A. N., Fadhillah, Y., Sari, N. R., Saputra, H., Darsin, D., & Pakpahan, A. F. (2022). *Teknologi Cloud Computing*. Yayasan Kita Menulis.

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.

Pamungkas, A. (2018). Karakteristik parameter oseanografi (pasang-surut, arus, dan gelombang) di perairan utara dan selatan Pulau Bangka. *Buletin Oseanografi Marina*, *7*(1), 51–58.

Pratama, I. P. A. E. (2021). Infrastructure as Code (IaC) Menggunakan OpenStack untuk Kemudahan Pengoperasian Jaringan Cloud Computing (Studi Kasus: Smart City di Provinsi Bali)(Infrastructure as Code (IaC) Using OpenStack for Ease of Operation of A Cloud Computing Network (A Case Study of Smart City in Bali Province). *JURNAL IPTEKKOM (Jurnal Ilmu Pengetahuan & Teknologi Informasi)*, *23*(1), 93–106.

Putri, R. A., Winahju, W. S., & Mashuri, M. (2020). Penerapan Metode Ridge Regression dan Support Vector Regression (SVR) untuk Prediksi Indeks Batubara di PT XYZ. *Jurnal Sains Dan Seni ITS*, *9*(1), D64–D71.

Suprayitno, S. (2018). KOMPUTASI AWAN DAN MODEL IMPLEMENTASI PADA ORGANISASI PEMERINTAHAN. *SEMNASTEKNOMEDIA ONLINE*, *6*(1), 2–3.

ISSN 2301-8984 (Print)

Supriyadi, E., & Siswanto, W. S. (2019). Analisis Pasang Surut di Perairan Pameungpeuk, Belitung, dan Sarmi Berdasarkan Metode Admiralty. *Jurnal Meteorologi Dan Geofisika*, *19*(1), 29–38.

Utami, T. M. N., Novitasari, D. C. R., Setiawan, F., Ulinnuha, N., & Farida, Y. (2021). Tide Prediction in Prigi Beach using Support Vector Regression (SVR) Method. *Scientific Journal of Informatics*, *8*(2), 194–201.

Wakhid, N., & Syahbuddin, H. (2018). WAKTU TANAM PADI SAWAH RAWA PASANG SURUT PULAU KALIMANTAN MENGHADAPI PERUBAHAN IKLIM. *Buletin Buletin*, *4*(1), 77–86.