

Prediction of Tidal Data Using Cloud Computing-Based Support Vector Regression in Alor, East Nusa Tenggara

Ruly Adhitya¹, Irawati Devi², Imel Novia Sari³

Marine Information Systems Study Program, University of Education Indonesia, Jl. Dr. Setiabudi No.229, Isola, Kec, Sukasirim, Bandung City, West Java 40154

ARTICLE INFO

Article history:

Received May 10, 2022

Revised Jun 12, 2022

Accepted Jul 28, 2022

Keywords:

Alor
Ups and down
Support Vector Regression
Cloud Computing

ABSTRACT

Alor Regency is an area located at 8°6' South Latitude – 8°36' South Latitude and 123°48' East Longitude – 125°48' East Longitude. Alor as an archipelagic region with wide waters and rich in various types of fish. The tides (tidal) of sea water is a phenomenon of periodic ups and downs of sea levels that can affect the activities of coastal communities. The existence of today's technological advances by predicting the height of the tides can be a control for the sustainability of community activities in coastal areas affected by the tides. Predictions are made using tidal data in time series since 2019, 2020, and 2021 obtained from the PUSHIDROSAL website using the Support Vector Regression (SVR) algorithm based on Cloud Computing. Therefore, the purpose of this research is to build tidal predictions in Alor, NTT which then integrates all datasets into the Cloud Computing environment to speed up data processing. parameter tuning is done with the Radial Basis Function (RBF) kernel using a gamma parameter of 1000 which produces an RMSE value of 0.0105 and a correlation of 1, so the model is considered quite good in predicting tides at the Alor pushidrosal station, East Nusa Tenggara. .

This is an open access article under the [CC BY-NC](#) license.



Corresponding Author:

Ruly Aditya,
Marine Information Systems Study Program,
Indonesian education university,
Jl. Dr. Setiabudi No.229, Isola, Kec, Sukasirim, Bandung City, West Java 40154.
Email: rulyadhitya@upi.edu

1. INTRODUCTION

The area of Alor Regency is located at 8° 6'LS – 8° 36' South Latitude and 123° 48' East Longitude – 125° 48' East Longitude, which is adjacent to the Flores Sea in the north, in the south side by the Ombay Strait and Timor Leste, in the east with the Northwest wider than the land area, which is 10,773.62 km² or about 79%, and has an area of 2,864.64 km² or about 21% in 2009 (Maleiku and Nurlala 2022). The total population is around 168,965 people. The diversity that exists in Alor Regency, NTT makes the majority of the residents' livelihoods as farmers and fishermen (Kusnandar 2020). As an archipelago, Alor has a wide water area and is rich in various types of fish such as grouper, shark, snapper, anchovies, mackerel, tuna, tuna, and others. As well as other types of marine products such as pearls, seaweed, sea cucumbers, jellyfish, and several types of shells and marine gardens. (Muawanah, Triyanti, and Soejarwo 2020).

Tides (tidal) are events that change the motion of the sea level in a proportional manner due to a mixture of gravitational and attractive forces from astronomical objects such as the sun, earth, and moon.(Radjawane, Saputro, and Egon 2018). Having an understanding of Indonesian waters such as tides (tidal), tsunami disaster monitoring, defense and security, tourism, and sports is very important(Harito, Fatoni, and Pranowo 2020). Tides affect physical processes such as the rushing of seawater to the shore caused by waves and the cleansing of water masses in estuaries, lagoons and bays. Tides can also affect biological activities such as plant zoning and the feeding of birds, fish and other marine organisms, as well as processes associated with the submergence and drying of land surfaces in the intertidal zone, i.e. the blood that lies between the highest and lowest tides.(Fallahiyah 2021). Tidal activity can affect coastal dynamics in some areas of East Nusa Tenggara, namely the phenomenon of coastal flooding (Rob) so that it has an impact on transportation activities around ports and coasts, activities of salt farmers, inland fisheries, and loading and unloading activities at the port are disrupted.(MULHAN 2020).

Today's technology has developed a lot in various aspects of human life, one of which is technology in the computer field, where computerized systems are currently being developed as an alternative to assist in data processing and providing accurate and more efficient and effective information.(Cholik 2021). In order to be able to predict the existence of undesirable phenomena that occur due to tidal activity, then with the advancement of technology today to predict the tides of sea water can be done by creating a system automatically. One of the fields of science that can help create automatic prediction systems is Machine Learning (ML). The computer process for learning data is ML(Id 2021). ML has several algorithms, one of which is Support Vector Regression (SVR) which is used to make predictions. There have been many objects that use SVR to predict(Yudhawan 2020). This algorithm has been successfully used to predict sugarcane crops in every island of Java based on climatic factors carried out by(Utami et al. 2021). Utilization of technology to facilitate storage and speed up processing of large amounts of data is used information technology based on Cloud Computing or cloud computing. Cloud Computing is a technology that uses the internet and a central server remotely, and has the purpose of maintaining or managing customer data(Giap et al. 2020). In addition, Cloud computing can help to use applications without installation, access personal files anywhere by using internet access(Quan et al. 2020). Cloud Computing technology can be considered as an efficient technology by centralizing data storage, processing and memory(Dhika et al. 2019)

According to research that has been done regarding the Support Vector Regression method as a good method for forecasting, as well as Cloud Computing regression which can produce significant equation models, this study uses a combination of the Support Vector Regression method to predict tides. Then proceed with storing data so that it can be processed quickly using the use of Cloud Computing technology. Therefore, this study aims to build tidal predictions in Alor, East Nusa Tenggara using daily data containing collection time and water level using Support Vector Regression. Then integrate all data sources (datasets) into the Cloud Computing environment to speed up data processing.

2. RESEARCH METHODS

2.1 Research Place

The area used in this study is the Pushidrosal station which is located in Alor Regency, East Nusa Tenggara. The map of the research location can be seen in Figure 1.



Figure 1.District Map. Alor, East Nusa Tenggara

2.2 Research Stages

This study focuses on making a prediction model for the tidal phenomenon at the Pushidrosal Station of Alor Regency and will be implemented in a cloud computing environment. There are stages to be carried out in this research process, these stages are shown in Figure 2. below:

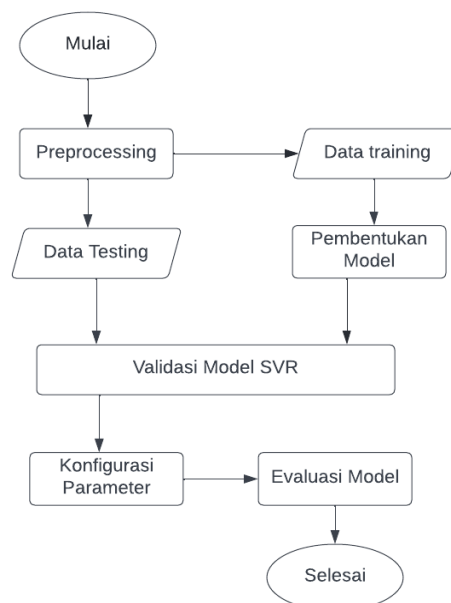


Figure 2.Research Flowchart

2.3 Data collection

The data in this study utilizes real-time tidal data which is accessed through the website of the Naval Hydrographic and Oceanographic Center (pasut-pushidrosal.luweswatersensor.com). Tidal data used is data from January 1, 2018 to May 24, 2019 in Alor, East Nusa Tenggara. Then the data is collected into a time series dataset by copying all the tidal data and then pasting it in Microsoft Excel software. In the attribute dataset, the time attribute shows the time of data collection and the water level attribute shows the sea level at the time of monitoring. This estimate uses hourly data with 11,104 data lines collected. Table 1. shows an example of tidal data from the pushidrosal website.

Table 1.Tidal Dataset Sample

Time (day/month/date hour:minute)	Water Level (m)
04/19/2018 06:00	1.438
19/04/2018 05:00	1,734
19/04/2018 04:00	2,140
19/04/2018 03:00	2,563
04/19/2018 02:00	2,770
04/19/2018 01:00	3.018
04/19/2018 00:00	2,566

2.4 Data Preprocessing or Preprocessing

This stage is the initial stage of research so that the data that has been prepared becomes a dataset, for further data processing using the utilization of the use of data mining methods. This is done in order to get tidal predictions that have good accuracy. The first thing to do is to check for missing values in the dataset. To fill in the missing value if the distance between two dates lies in the data, the median value will be used and if the missing value is not within two dates, the mean of similar data will be used. Then the next stage in the research process, the dataset will be divided into two parts, namely training data and testing data. The tidal prediction dataset contains an allocation of 80% training data, while 20% is used as testing data. In this study, the distribution of the dataset uses the k-fold cross validation method with 10 fold iterations. The overall results of the divided data are 8,333 for training data and 2,221 for testing data.

2.5 Parameter Configuration

In this stage will determine the most optimal parameters to form the best model in predicting. Parameter determination on SVR with polynomial kernel and Radial Basis Function (RBF) is shown in Table 2. Polynomial Kernel Parameters and Table 3. RBF Kernel Parameters.

Table 2. Polynomial Kernel Parameters

Parameter	Value/Status
C	1.0
(Epsilon)	1.0
	2.0

Table 3.RBF Kernel Parameters

Parameter	Value/Status
C	1.0
Gamma	0.1
	1

10

100

1000

2.6 Tools Study

The dataset has been prepared, The next step is to prepare computing with the python programming language on the Windows 11 operating system hardware.

- Programming language: python v3.9.7, this programming language is used because of its convenience, multifunction (many support modules) and can be used to perform data analysis besides that it can run on many operating systems.
- *Interface*: interface using jupyter notebook in this study.
- *Library*: module of python using Numpy, SciPy, Scikit-learn, and Matplotlib in this research.

2.7 Model Evaluation

The model evaluation step is needed in this study because in order to ensure that the model that has been built, this model must have a fairly good accuracy performance. Root Mean Squared Error (RMSE) and correlation coefficient (R) are evaluation methods that are useful in the process of developing tidal predictions in this study. The RMSE value is obtained by calculating the equation 1(Ariffin et al. 2021). The correlation coefficient value of the model is obtained by calculating the equation.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n}} \quad (1)$$

with :

- n : predicted amount of data
 l : data order
 e : error between actual and predicted values

$$R = \frac{S_{xy}}{S_x S_y} \quad (2)$$

with :

- S_{xy} : covariance between actual and predicted values
 S_x : standard deviation of actual value
 S_y : standard deviation of predicted value

3. RESULTS AND DISCUSSION

This study uses secondary data in the form of tidal time series data at the Naval Hydrographic and Oceanographic Center (PUSHIDORSAL) monitoring station in Alor, East Nusa Tenggara which was obtained from the tide-pushidorsal website. The online data used is often incorrectly entered. Therefore, preprocessing is needed which aims to process the data before it is used in the prediction model. Table 4. shows tidal data in Alor, East Nusa Tenggara on February 21, 2019.

Table 4.Missing Tidal Dataset Value

Time and Date	February 21, 2019						
	3:00	4:00	5:00	6:00	7:00	8:00	9:00
Water Level	2008	1.417	0.848	0.000	0.455	0.912	1.514

There is a missing value at 6:00 (Table 4. Marked with a highlight). In the data before and after 6 o'clock there is a water level variable, the data will be filled with the median value of the 5:00 and 7:00 o'clock data. The results of filling in the missing values can be seen in table 5.

Table 5.The Result of Filling in the Missing Tidal Dataset Value

Time and Date	February 21, 2019						
	3:00	4:00	5:00	6:00	7:00	8:00	9:00
Water Level	2008	1.417	0.848	0.6515	0.455	0.912	1.514

After the dataset is collected, the next step is to divide the data into two groups of data, namely training data and testing data. The method used to divide the data is to use k-fold cross validation with 10 fold iterations. The results for the training data from all the shared data are 8,883 and for testing data are 2,221.

In the development of tidal data prediction, the Support Vector Regression (SVR) method is tested with Improved Sequential Minimal Optimization (ISMO). Based on the scenario designed using two kernels, namely polynomial and Radial Basis Function (RBF). Based on SVR using two polynomial kernels and RBF, a training and testing process was carried out with the tidal dataset, the results of the optimization can be seen in the following table.

Table 6.Polynomial Kernel SVR Training Results

C	E	RMSE	Correlation
1.0	1.0	1.9333	0.3735
1.0	2.0	1.5836	0.64

Table 7.Polynomial Kernel SVR Test Results

C	E	RMSE	Correlation
1.0	1.0	1.8952	0.3829
1.0	2.0	1.5554	0.6429

Table 8.SVR Kernel Radial Basis Function Training Results

Trial to-	Gamma	RMSE	Correlation
1	0.1	1.5817	0.6351
2	1	1.4678	0.6993
3	10	1.1515	0.8313
4	100	0.8666	0.9095
5	1000	0.0227	0.9999

Table 9.SVR Kernel Radial Basis Function Test Results

Trial to-	Gamma	RMSE	Correlation
1	0.1	1.5514	0.6393
2	1	1.441	0.702
3	10	1.1471	0.8264

4	100	0.8134	0.91740
5	1000	0.0105	1

Based on the results of the model evaluation for the tides, the best size chart for the model can be seen with the RMSE . value(Root Mean Squared Error) and its correlation coefficient value. The evaluation graph of the RMSE value is presented in Figure 2. While the evaluation graph of the correlation coefficient value can be seen in Figure 3.

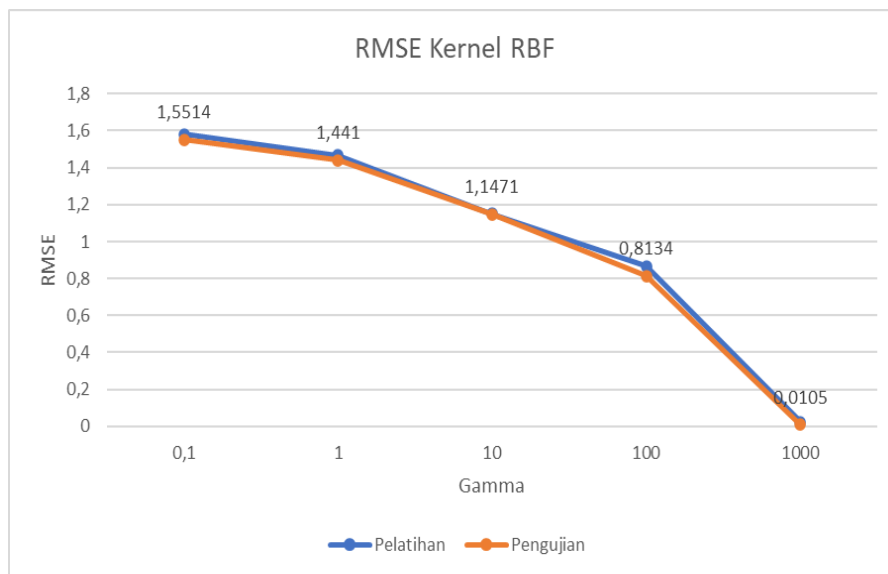


Figure 3. RBF Kernel RMSE Graph

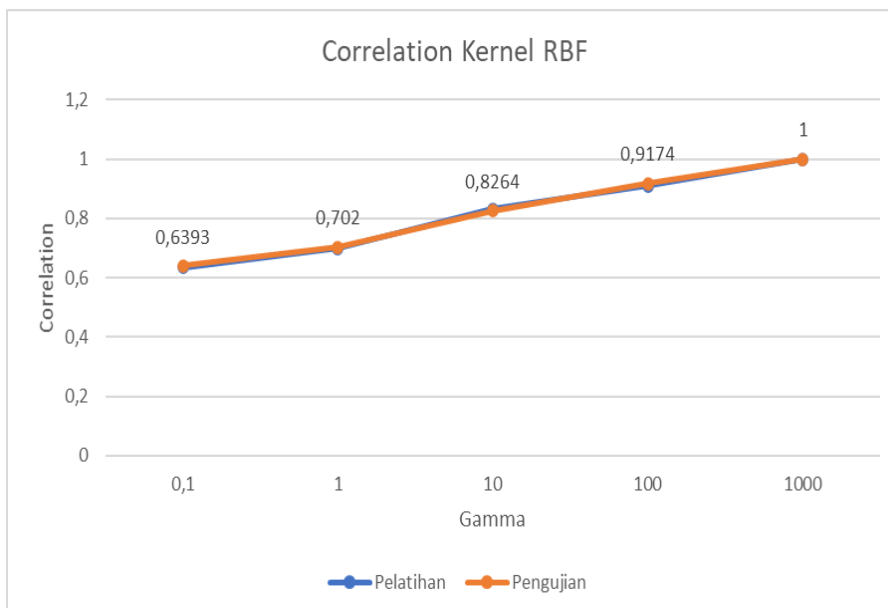


Figure 4. RBF Kernel Correlation Graph

From several experiments conducted at the pushidrosal station, the best tidal prediction model in Alor, East Nusa Tenggara was generated using the Radial Basis Function (RBF) kernel using the parameter gamma 1000. The best RMSE value obtained was 0.0105 and the best correlation value. The results obtained are 1. Based on the RMSE value and the best correlation produced, this value indicates that the prediction model made is quite good in predicting the tides at the pushidrosal station in Alor, East Nusa Tenggara.

Accuracy values and time calculations in training and testing data are needed in this study. The results of the average construction time of the prediction value using the Support Vector Regression method with polynomial kernels and RBF are shown in Table 10.

Table 10. RMSE Value and Computing Time

Method	RMSE-Train	RMSE-Test	Training Time(s)	Testing Time(s)
SVR-Poly	1.175845	1.7253	0.01	0.0055
SVR-RBF	1.05882	0.99268	8,324	2.172

In this study, the tidal data will be processed using *cloud computing* on Microsoft Azure, therefore dataset migration will be carried out into a cloud computing environment. On the personal computer, the memory used is larger, namely 8 GB, while the memory provided by the virtual machine in the cloud is only 1 GB. If the evaluation of the tidal model is carried out in a cloud computing environment, from Figure 5 it can be seen that the training time in the cloud requires a faster time. than on a personal computer.

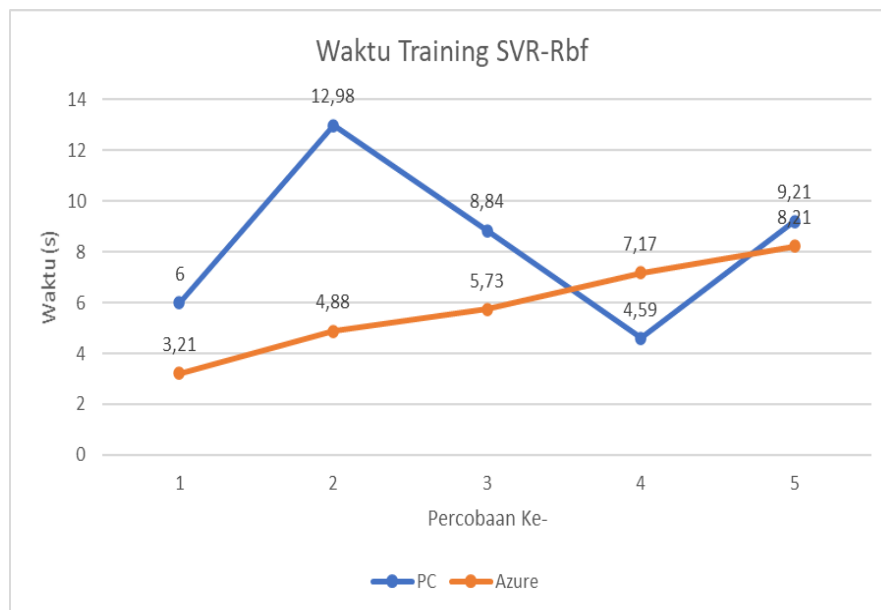


Figure 5. SVR Training Time

In the process of evaluating the model in the personal computer and azure computing environment in conducting data training with the same parameters and prediction models, it produces consistent values but in processing time with the environment. *Cloud Computing* choose a shorter time with an average of 5.84 seconds while the personal computer takes an average of 8.324 seconds. Thus the prediction model generated can be used to forecast or predict the tidal height of

the sea water at Pushidrosal Alor Station, East Nusa Tenggara in the future and processing can be optimized using Cloud Computing.

4. CONCLUSION

Based on the data processing that has been done with several methods to produce good and accurate predictions which are represented by the RMSE value and the correlation coefficient value. According to the results of this study, it can be concluded that predictions with tidal data using the Support Vector Regression method have been successfully formed with fairly good results. The best results obtained from the parameter tuning carried out are the Radial Basis Function (RBF) kernel using a gamma parameter of 1000 which produces an RMSE value of 0.0105 and a correlation of 1, so that the model is considered quite good in predicting tides at the pushidrosal station. Alor, East Nusa Tenggara.

ACKNOWLEDGEMENTS

Thank you to the supervisor who has guided and assisted the author in the learning process of data processing, and provided input during the process of making this paper. Thank you to friends for their motivation and enthusiasm so that this paper can be completed properly and on time.

REFERENCE

- Arifin, Will and Aprizal et al. 2021. "TIDE PREDICTION MODEL OF SEA WATER TIDE AT PUSHIDROSAL STATION BAKAUHENI LAMPUNG USING SUPPORT VECTOR REGRESSION." *Maritime Journal: Indonesian Journal of Maritime* 2(2): 139–48.
- Cholik, Cecep Abdul. 2021. "Development of Information and Communication Technology/ICT in Various Fields." *Journal of the Brass Engineering Faculty* 2(2): 39–46.
- Dhika, Harry, Triyani Akhirina, Dewi Mustari, and Fitriana Destiawati. 2019. "Utilization of Cloud Computing Technology as a Data Storage Media." *PKM Journal: Community Service* 2(3).
- Fallahiyah, Silvy Ana Dwi. 2021. "Mapping the Vulnerability of Coastal Areas to Climate Change in Gending District, Probolinggo Regency."
- Giap, Yo Ceng et al. 2020. *Cloud Computing: Theory And Implementation*. Our Writing Foundation.
- Harito, Rusito Retri, Khoirol Imam Fatoni, and Widodo Setyo Pranowo. 2020. "Study of Calculation of Map Sequence Based on Time of Supermoon Occurring in Kalabahi Waters, East Nusa Tenggara." *Journal Chart Datum* 6(1): 10–23.
- Id, Ibn Daqiqil. 2021. *1 Machine Learning: Theory, Case Studies and Implementation Using Python*. Unri Press.
- Kusnandar, Viva Budy. 2020. "This is the Projection of Indonesia's 2020 Population." *databox*.
- Maleiku, Mardianti Yosefina, and Mrs. Nurlela. 2022. "Sea Products and Fishermen's Life on Pura Island, Alor Regency, East Nusa Tenggara." *Journal of Social and Cultural Studies: Tebar Science* 6(2): 55–62.
- Muawanah, Umi, Riesti Triyanti, and Permana Ari Soejarwo. 2020. "Economic Impact of Maritime Tourism in Alor Regency." *Journal of Socio-Economic Marine and Fisheries* 15(1): 33–46.
- MULHAN, HADI. 2020. "LAND USE ANALYSIS OF ROB FLOOD IN WEST LOMBOK DISTRICT IN 2050."
- Quan, Quan, Zou Hao, Huang Xifeng, and Lei Jingchun. 2020. "Research on Water Temperature Prediction Based on Improved Support Vector Regression." *Neural Computing and Applications*: 1–10.
- Radjawane, Ivonne M, Bagus SC Saputro, and Andi Egon. 2018. "Tide Hydrodynamic Model in the Waters of the Bangka Belitung Islands." *Journal of Civil Engineering* 25(2): 121.
- Utami, Tri Mar'ati Nur et al. 2021. "Tide Prediction in Prigi Beach Using Support Vector Regression (SVR) Method." *Scientific Journal of Informatics* 8(2): 194–201.
- Yudhawan, Dhanukhresna Hangga. 2020. "Implementation of Support Vector Regression for Forecasting Shares of Mining Companies in Indonesia (Case Studies: PT Adaro Energy Tbk, PT Bukit Asam Tbk, and PT Indo Tambangraya Megah Tbk)."