

Method of Application of Support Vector Regression In Predicting The Number of Visits of Foreign Tourists To The Province of Maluku

¹Wahyuni Aprilya, ²Marlon S. N. Van Delsena, ³M.Y. Matdoana

^{1,2,3}Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Pattimura, Indonesia

¹wahyuniapriily@gmail.com, ²marlonnvd@gmail.com, ³keepyahya@gmail.com

Article Info

Article history:

Received Nov 23th, 2022

Revised Feb 14th, 2023

Accepted Jun 13th, 2023

Keyword:

Application

Covid-19

Support Vector Regression

Tourist

Traveler

ABSTRACT

Maluku Province is one of the areas in Indonesia that has many tourist attractions to visit, both natural and cultural heritage. The high interest of foreign tourists who want to visit various tourist objects, makes the tourism industry able to bring benefits and profits for most of the people of Maluku. However, in the last two years, 2020-2021, all countries were faced with the Covid-19 pandemic. The Covid-19 pandemic has resulted in a decrease in the number of foreign tourist visits to Indonesia. To increase marketing activities in the midst of the Covid-19 pandemic that has hit Indonesia since 2020, foreign tourist information is increasingly needed as material for evaluation and planning for future development. One of the methods used to predict the number of foreign tourist visits to Maluku Province is Support Vector Regression (SVR). Based on forecasting using test data, the RMSE value is 1.334985 and the MAPE obtained is 1.256346%, so the prediction of the number of foreign tourist visits to Maluku Province in 2022 (January-June) states that in January the number of tourist visits was 999 hundred visits. and increased until June as many as 1121 thousand visits.

Copyright © 2023 Puzzle Research Data Technology

Corresponding Author:

M.Y. Matdoana,

Program Studi Statistika,

Fakultas Matematika dan Ilmu Pengetahuan Alam,

Universitas Pattimura, Maluku Indonesia,

Email: keepyahya@gmail.com

DOI: <http://dx.doi.org/10.24014/ijaidm.v6i1.19803>

1. INTRODUCTION

Indonesia is one of the countries that has the most islands in the world, so it is dubbed the country of the Thousand Islands or the country of a thousand islands with 17,508 islands inhabited by more than 360 ethnic groups. The abundance of natural beauty and the diversity of culture, customs and historical heritage that stretches from Sabang to Merauke. This causes Indonesia to become one of the most popular tourist destinations for foreign tourists so that it has a good influence on being a source of foreign exchange for the country and makes tourism one of the leading sectors in Indonesia. In general, tourism is a planned trip, carried out for a while from one place to another with the intention of getting a form of satisfaction or pleasure [1-3]. For Indonesia, the role of the national tourism sector is increasingly important in line with the contribution made by the tourism sector through foreign exchange earnings, regional income, regional development, as well as in the absorption of investment and labor and business development spread across various parts of Indonesia [4]. Maluku Province is one of the regions in Indonesia which has a group of small islands with a diversity of ethnic groups that grow and develop and have many features. As a province that has most of the sea compared to the mainland, Maluku has a lot of potential for coastal tourism objects, a wealth of arts and culture, marine tourism, history, culinary arts and is rich in spices and abundant marine products. So many potentials for natural beauty and cultural heritage and historical heritage make the tourism

sector in Maluku one of the supporters of tourist attraction, both local and foreign tourists. This can encourage and improve the economy in Maluku Province, especially the people [5-7].

The development of the tourism sector in the Maluku Province is quite developed. In the development of the number of foreign tourist visits to Maluku Province before 2020, it shows a positive growth rate[8-9]. According to the Maluku Central Statistics Agency (BPS), in 2019 the number of foreign tourist visits to Maluku Province fluctuated from year to year (BPS, 2019). It was recorded that the number of tourist visits in 2016 was 8,991 thousand tourist visits, then in 2017 there were 18,075 thousand tourist visits, then decreased in 2018 to 14,253 thousand tourist visits. Then again, it decreased again in 2019, there were 13,153 thousand tourist visits, followed by 2020 there were 2,010 thousand tourist visits and in 2021 there were 1,450 thousand tourist visits. In the last two years, 2020-2021, the number of visits has decreased very sharply[10-12]. This is allegedly influenced by the 2019 Corona Virus (Covid -19) pandemic which spread throughout the country. Covid-19 is an infectious disease that attacks the respiratory system, causing minor disturbances such as flu, severe lung infections, and even death [13]. This disease outbreak is now a trending topic around the world originating from Wuhan, China. The Covid-19 pandemic has changed many things, including presenting the impact of the global economic crisis which is the most serious compared to previous crises. On the other hand, the Covid-19 pandemic has also had an impact on the decline in the number of foreign tourist visits to the country[14]. Around 60 countries in the world apply entry restrictions for visits from Indonesian citizens. With the flow of tourist borders between countries carried out by the Indonesian government, it affects the number of foreign tourist visits to Indonesia. On the other hand, these countries will also prohibit their citizens from visiting Indonesia because of the high risk of a pandemic. This has resulted in a decrease in the number of foreign tourist arrivals [15].

In order to increase marketing activities in the midst of the Covid-19 pandemic that has hit Indonesia since 2020, foreign tourist information is increasingly needed as material for evaluation and development planning in the future. Based on these conditions, it is very important to make accurate predictions to see the predicted rate of visits. In optimizing the facilities, we need a method that can predict the number of foreign tourists visiting Maluku Province in the future. One of the methods used to predict the number of foreign tourist visits to Maluku Province is Support Vector Regression (SVR) [16]. Support Vector Regression (SVR) is a development method of the Support Vector Machine (SVM) with a regression method so that it can be used for forecasting like a regression method. The concept of SVR is to maximize the hyperplane so as to minimize errors. To help overcome non-linear problems in the SVR model, the kernel is used. The advantage of the SVR method is its ability to utilize non-linear data implicitly through the application of kernel functions [16]. Based on the description above, the authors are interested in taking the title of the study, namely the application of the Support Vector Regression method in predicting the number of tourist visits to Maluku Province.

2. RESEARCH METHOD

2.1 Research Type

The type used in this study is quantitative by applying the Support Vector Regression (SVR) Method in Predicting the Number of Tourist Visits to Maluku Province.

2.2 Research Materials

The materials or materials used in this research are secondary data, which are obtained from other parties and are readily available. The dataset used in this study is monthly data on the number of foreign tourist visits to Maluku by nationality for the 2016-2021 period obtained from the Central Statistics Agency (BPS). Data collection was carried out in March 2022 through the official website <https://maluku.BPS.go.id/> Maluku Province in Figures 2017-2022. The variable used in this study is data on the number of foreign tourist visits who come to Maluku by nationality for the 2016-2021 period.

2.3 Research procedure

This study uses the help of software R, Minitab 19 and Microsoft Excel 2016 to support other calculations. The stages of the research used are as follows:

1. Description of data on the number of foreign tourist visits to Maluku Province in 2016-2021 as initial information on the diversity of observational data.
2. Testing the normality of data on foreign tourist visits to Maluku.
3. Determination of the predictor variable (x) using the PACF plot.
4. Divide the data into two, namely the training and testing stages.
5. Building SVR models with R software.
6. Selection of the best model through grid-search.
7. In this research, the kernel used is Radial Basic Function.

8. Forecasting the training data. Forecasting the testing data.
9. Evaluation of the forecasting model by looking at the results of the MAPE and RMSE error values generated between the actual data and the predicted data on the testing data.
10. Forecasting the next period. Forecasting the number of foreign tourist visits to Maluku Province for the next six months or 6 months in 2022, from January to June.
11. Conclusion explanation.

3. EXPERIMENT RESULT AND DISCUSSION

3.1 Overview of the Development of Macanegara Tourists to Maluku Province

Maluku Province is located in the eastern part of Indonesia, with an area consisting of a group of islands stretching between Sulawesi Island and Papua Island. As a province that has a wide sea, Maluku has a lot of natural wealth that has the potential as a place for tourism. In addition to the great marine tourism potential, Maluku also has historical and cultural heritage such as various colonial forts, peace gong museum, cultural festivals, and others. The development of the tourism industry in Maluku Province is finally getting a breath of fresh air, after struggling for a long time with the problem of decreasing the number of tourist arrivals since 2010. Data from the Ministry of Tourism notes that the number of domestic and foreign tourists visiting Maluku during 2017 has increased by 20%. This positive trend was repeated in 2018 and had an impact on the economy of the people in Maluku as well as an impact on the increase in PAD from the Maluku Provincial Tourism Office from three tourist locations (namely Hunimua Beach, Namalatu Beach and the World Peace Gong Monument). The development of the tourism sector in the Maluku Province is quite developed. In the development of the number of foreign tourist visits to Maluku Province before 2020, it shows a positive growth rate. According to the Maluku Central Statistics Agency in 2017 the number of tourist visits increased rapidly.

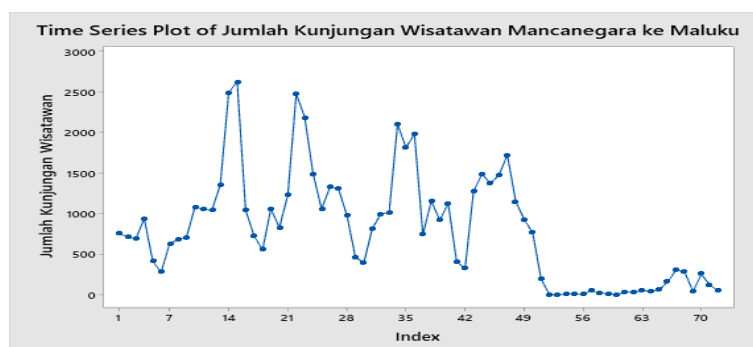


Figure 1. Number of Foreign Tourist Visits to Maluku Province

It was recorded that the number of tourist visits in 2016 was 8,991 thousand tourist visits, then in 2017 there were 18,075 thousand tourist visits, then decreased in 2018 to 14,253 thousand tourist visits. Then again, it decreased again in 2019, there were 13,153 thousand tourist visits, followed by 2020 there were 2,010 thousand tourist visits and in 2021 there were 1,450 thousand tourist visits. The decline in the number of foreign tourist visits to Maluku Province in the last two years 2020-2021 was influenced by the fact that the whole world was faced with the Covid-19 pandemic. The first confirmed case of Covid-19 in Indonesia in early March 2020 was confirmed in early March 2020. This global pandemic has not only attacked the public health sector but has also devastated the economic sector and the tourism sector. The decline in the demand side and the decline in the supply side at the same time as well as social restrictions have made the tourism industry worse off. The high number of positive confirmed cases of Covid-19 in Indonesia has an impact on the tourism industry in Indonesia. The existence of large-scale social restrictions and the closure of access and entry gates resulted in a decrease in state revenue in the tourism sector by 20.7 billion. However, with the pandemic, the tourism sector has been the worst hit and it will take a long time to recover.

3.2 Data Normalization

Before building a forecasting model using SVR, the data will be normalized first. Normalization is done to reduce the difference in the range of data that is far between the data of the dependent variable and the data of the independent variable. The first step in estimating the model parameters in the SVR method is to perform a Z-score or standard score on the x data.

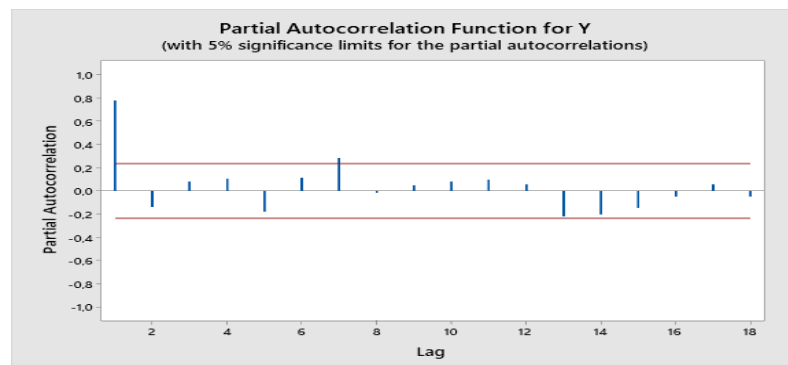
Table 1. Results of Normalization of Data on the Number of Foreign Tourist Visits to Maluku Province

Number	Visit
1	-0,196398993
2	-0,161075433
3	0,365952081
⋮	⋮
60	-0,799725397
71	-0,994711448
72	-1,093617416

Based on Table 1, the normalization process using the Z-Score, the value obtained in the table above is the data minus the average value (mean) then divided by the standard deviation, then the value of each normalization is obtained in each data set. Berdasarkan Tabel 4.1 mengalami proses normalisasi dengan menggunakan *Z-Score*, nilai yang diperoleh pada tabel diatas adalah data dikurangi dengan nilai rata-rata (*mean*) lalu dibagi dengan *standar deviasi*, maka diperoleh nilai masing-masing normalisasi pada setiap data set.

3.3 Determination of Predictor Variables

Before entering into modeling using SVR, the first step is to convert the data on the number of foreign tourist visits to Maluku Province into the form of time lag which will be used as input data in the SVR model. At this stage the step taken is to convert the data on the number of foreign tourist arrivals who come to Maluku Province into the form of time lag. Determination of the amount of time lag is based on the PACF plot. Significant lag in the PACF plot will be used as a predictor variable (x). The PACF plot of data on the number of foreign tourist visits to Maluku Province can be seen in Figure 2.

**Figure 2.** Plot of PACF Data on Number of International Tourist Visits to Maluku Province

Based on the PACF plot in Fig 2, it can be seen that the significant lag in the PACF plot is lag 1 and lag 7 (attachment 3) so that the total time lag for SVR data input is 2 lags, namely lag 1 and 7. So there are two predictor variables (x) used, namely $x_1 = y_{t-1}$ and $x_2 = y_{t-7}$ in the data set.

3.4 Determination of Training Data and Testing Data

To perform analysis of Support Vector Regression, the data will be divided into training data and testing data. The creation of training data is carried out to improve the performance of Support Vector Regression on data testing and determine the best parameters for model formation. The distribution of data in this study is 70% of the total data used as training data and the remaining 30% of the total data as testing data. Thus, the training data is 46 datasets and the testing data is 19 datasets. The amount of training data has a greater percentage than testing data because machine learning in forming models and models that are formed are trained using data testing can provide more optimal forecasting on testing data.

3.5 Implementation of Support Vector Regression

3.5.1 Building a Support Vector Regression Model

In this study, it is limited to using the Radial Basic Function (RBF) kernel function only. In the RBF kernel there are parameters C , and that must be specified. Hyperparameter value optimization is done by using training data by using Grid Search on Radial Basic Function (RBF) kernel with parameter value $cost = 10^{-1}, 10^0, 10^1, 10^2, 10^3$ combined with parameter $gamma = 10^{-1}, 10^0, 10^1, 10^2, 10^3$. Based on the

lower and upper limit values of the cost and gamma hyperparameters that have been determined previously, the error results from each combination of parameter values performed on the training data will be shown in Table 2.

Table 2. SVR Model Error for Optimal Parameters with Grid Search Method

Gamma	Cost	Error
0.1	0.1	0.7421845
1	0.1	0.7504382
10	0.1	0.8153147
100	0.1	0.8163433
1000	0.1	0.8171383
⋮	⋮	⋮
100	1000	0.8736868
1000	1000	0.8209905

From these results, it is known that from the parameter selection, the best model is selected with the Radial Basic Functiona function where the smallest error value is in the 1st combination with an error value of 0.7421845 where the hyperparameter values obtained are cost = 0.1 and gamma = 0.1. Furthermore, data prediction is carried out and the results are shown in the Figure 3.

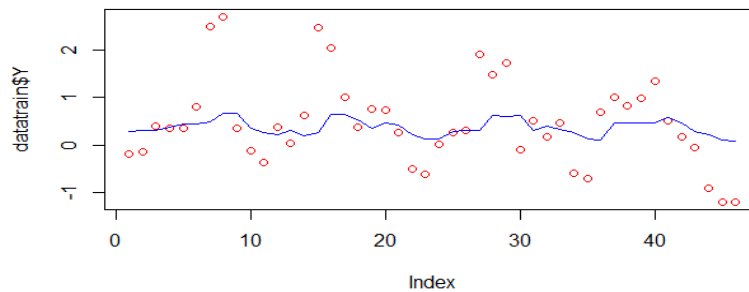


Figure 3. Plot Prediction of Training Data Model

Figure 3 is a visualization of the calculation prediction results using the Radial Basis Function kernel. The blue line represents the actual data, while the red dot represents the predicted data. From the figure, the prediction results are quite good following the trend of the data.

The next step is to see how far the accuracy or accuracy of the SVR model in the training data is for predicting the number of foreign tourists to Maluky Province by looking at the accuracy of the model's accuracy by looking at the Mean Absolute Percentage Error (MAPE) value. Based on Fig 3 shows the prediction results of the training data based on the best model obtained. The optimal predictive value is obtained from the model performance test results.

Table 3. Data Model Training Performance Test Results

Model Performance Test	Score
RMSE	0,8221864
MAPE	1,478026%

Based on Table 3, attached the calculation results of the RMSE value in the training data obtained a value of 0.8221864 and the calculation results of the MAPE value in the training data obtained a value of 1.478026%. According to the MAPE criteria, the MAPE value below 10% indicates that the model has very good forecasting ability. After obtaining the optimal SVR model, the parameter values are then implemented in the testing data.

3.5.2 Implementation of Data Testing

After getting the best model, then enter the best model and its parameters in the test data. The results of the best model predictions and their parameters can be seen in the test data, which can be seen in the table in Figure 4.

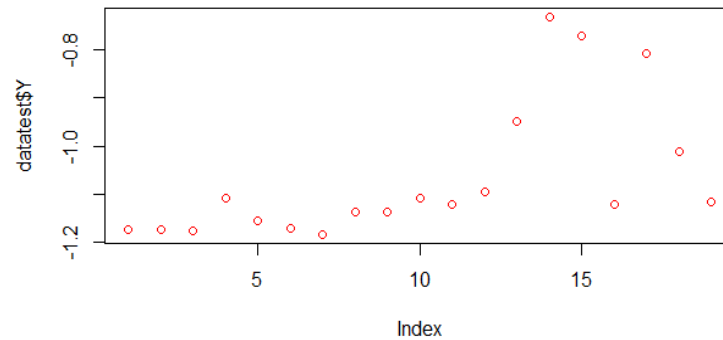


Figure 4. Prediction of Plot Data Testing

Based on Figure 4, it can be seen that the prediction results on the test data are based on the best model obtained from the SVR model. Prediction results are shown by following the red dot with an RMSE error value of 0.8892287 and the resulting MAPE, which is 0.8383393%. The following are the results of data predictions on testing data.

Table 4. Testing Data Prediction Results

Month	Prediction Data
47	0,08317261
48	0,12649694
49	0,15558588
50	0,17970040
51	0,27034793
52	0,29252820
53	0,29159560
54	0,28978481
55	0,29291135
56	0,29293728
57	0,28922976
58	0,29229102
59	0,29521732
60	0,30725205
61	0,32269305
62	0,31913187
63	0,28832495
64	0,31444540
65	0,29518979

3.6 Evaluation of Forecasting Models

After obtaining the best hyperparameter C and gamma values, this hyperparameter value can then be applied to the SVR forecasting model to predict testing data and predict future periods. Forecasting on testing data is carried out to evaluate the SVR forecasting model with the best hyperparameter C and gamma values. The evaluation of the forecasting model is carried out by looking at the results of the MAPE and RMSE error values generated between the actual data and the predicted data on the testing data. At this stage, an evaluation of the forecasting model generated through data testing forecasting will be carried out using the SVR forecasting model with the RBF kernel.

Table 5. Evaluation Results of Forecasting Models on Data Testing

Forecasting Model Evaluation	Score
RMSE	1,334985
MAPE	1,256346%

Based on the evaluation results of the best forecasting model in Table 5, the RMSE error value is 0.8892287 and the resulting MAPE is 0.8383393%. Based on the standard MAPE error value, the forecasting model is a forecasting model with very good criteria because the resulting MAPE error value is < 10%.

3.7 Forecasting the Next Period

Forecasting data in the future period is carried out using a forecasting model that has been built and tested in the previous stages, using the SVR forecasting model with the RBF kernel and the optimal

hyperparameter C and gamma values. At this stage, the number of foreign tourist visits to Maluku Province is forecasted for the next six months or 6 months in 2022.

Forecasting for the next period is carried out using testing data with the SVR forecasting model with the best hyperparameter values that have been searched for in the previous process. The forecasting results obtained are as follows.

Table 6. Predicted Number of Tourist Visits

Month	Number of Visits
Jan-22	999
Feb-22	1104
Mar-22	1115
Apr-22	1118
Mei-22	1125
Jun-22	1121

Based on Table 6, shows the predicted number of foreign tourist visits to Maluku Province for a six-month period in 2022. Based on these results, it is known that in January the number of tourist visits was 999 visits and increased until June as many as 1121 visits. Therefore, it can be concluded that the number of tourist visits has increased when compared to the previous year.

4. CONCLUSION

Based on the results and discussion, it is concluded that the development of the number of foreign tourists was recorded in 2016 and 2017 experienced an increase in tourist visits, then decreased in 2018 as many as 14,253 thousand tourist visits. Then again, it decreased again in 2019, there were 13,153 thousand tourist visits, followed by 2020, there were 2,010 thousand tourist visits and in 2021 there were 1,450 thousand tourist visits. In addition, based on forecasting using testing data, the RMSE value obtained is 1.334985 and the MAPE obtained is 1.256346%, so that the prediction results of the number of foreign tourist visits to Maluku Province in 2022 (January-June) states that it is known in January the number of tourist visits as many as 999 visits and increased until June as many as 1121 thousand visits. The number of tourist visits has increased when compared to the previous year.

REFERENCES

- [1]. Atmojo, joko tri, Akbar, P. S., Kuntari, S., Yulianti, I., & Darmayanti, A. T. Definition and Path of Transmission of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) or COVID-19. *Journal of Health Education (e-Journal)*, 9(1), 57–64. 2020.
- [2]. Suwena, I. K., & Widyatmaja, I. G. N. *Basic Knowledge of Tourism Science*. 2010.
- [3]. Fitriani. *Autoregressive (AR) Modeling on Missing Data*. Thesis. 2013
- [4]. Hendayanti, N. P. N., Suniantara, I. K. P., & Nurhidayati, M. Application of Support Vector Regression (Svr) in Predicting the Number of Domestic Tourist Visits to Bali. *Journal of Variants*, 3(1), 43–50. 2019.
- [5]. Martha, S., & Sulistianingsih, E. Support vector regression (SVR). 08(1), 1–10. Of, p., using, e., vector, s., & method, r. Prediction of electricity using support vector regression prediction of electricity using support vector regression method. 8(6), 12135–12144. 2019.
- [6]. Purnama, D. I., & Hendarsin, O. P. Forecasting the Number of Passengers Departing Through Air Transportation in Central Sulawesi Using Support Vector Regression (SVR). *Jambura Journal of Mathematics*, 2(2), 49–59. 2020.
- [7]. Purnama, D. I., & Setianingsih, S. Support vector regression (SVR) model for forecasting number of passengers on domestic flights at Sultan Hasanudin airport Makassar. *Journal of Mathematics, Statistics and Computing*, 16(3), 391. 2020.
- [8]. Suwena, I. K., & Widyatmaja, I. G. N. *Basic Knowledge of Tourism Science*. 2010.
- [9]. Wildana, Iqbal Gillang. *Digital Repository Repository University of Jember Jember Digital Digital Repository Repository University of Jember Jember University*. In *Digital Repository, University of Jember*. 2021.
- [10]. Atmojo, joko tri, Akbar, P. S., Kuntari, S., Yulianti, I., & Darmayanti, A. T. Definition and Path of Transmission of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) or COVID-19. *Journal of Health Education (e-Journal)*, 9(1), 57–64. 2020.
- [11]. Hendayanti, N. P. N., Suniantara, I. K. P., & Nurhidayati, M. (2019). Application of Support Vector Regression (Svr) in Predicting the Number of Domestic Tourist Visits to Bali. *Journal of Variants*, 3(1), 43–50. <https://doi.org/10.30812/varian.v3i1.506>
- [12]. Martha, S., & Sulistianingsih, E. Support Vector Regression (SVR). 08(1), 1–10. 2019.

- [13]. Vector, S., & Method, R. Prediction Of Electricity Using Support Vector Regression Prediction Of Electricity Using Support Vector Regression Method. 8(6), 12135–12144. 2021.
- [14]. Purnama, D. I., & Hendarsin, O. P. Forecasting the Number of Passengers Departing Through Air Transportation in Central Sulawesi Using Support Vector Regression (SVR). *Jambura Journal of Mathematics*, 2(2), 49–59. 2020. <https://doi.org/10.34312/jjom.v2i2.4458>
- [15]. Purnama, D. I., & Setianingsih, S. Support vector regression (SVR) model for forecasting number of passengers on domestic flights at Sultan Hasanudin airport Makassar. *Journal of Mathematics, Statistics and Computing*, 16(3), 391. 2020. <https://doi.org/10.20956/jmsk.v16i3.9176>
- [16]. Wildana, Iqbal Gillang. Digital Repository Repository University of Jember Jember Digital Digital Digital Repository Repository University of Jember Jember University. In *Digital Repository, University of Jember*. 2021.

BIBLIOGRAPHY OF AUTHORS



Wahyuni Aprily is an undergraduate form Statistics study program at Pattimura University in 2022 and is currently working as a data scientist.



Marlon Stivo Noya Van Delsen, obtained an M.Si (Master of Science) degree in the Master Program Statistics, Institut Sepuluh Nopember Surabaya. In 2015 until now, he has been lecture in the Statistics Study Programme at Pattimura University. Focus on research on Classification Method and Industry Statistics.



Muhammd Yahya Matdoan is a lecturer in the Statistics Study Program at Pattimura University. Complited Master degree from Insitut Teknologi Sepuluh Nopember with research topics occupied namely Classification Method, Data Mining and Statistical Cuality Control.