Nonlinear Dynamics and Systems Theory, **24** (4) (2024) 331–339

Optimization of Hotel Y Management through Application of Occupancy Forecasting by Support Vector Machine and K-Nearest Neighbors Methods

M. Y. Anshori [1], I. H. Santoso [2], P. Katias [1], T. Herlambang [3], H. Arof [4], B. Suharto [5*] and K. Oktafianto [6]

[1] *Department of Management, Universitas Nahdlatul Ulama Surabaya, Indonesia.*
[2] *Department of Accounting Magister, University of Wijaya Kusuma, Surabaya, Indonesia.*
[3] *Department of Information System, Universitas Nahdlatul Ulama Surabaya, Indonesia.*
[4] *Department of Electrical Engineering, University of Malaya, Malaysia.*
[5*] *The Tourism and Hospitality Department, Faculty of Applied Sciences, Airlangga University, Indonesia.*
[6] *Department of Mathematics, University of PGRI Ronggolawe, Indonesia.*

Received: October 28, 2023; Revised: June 22, 2024

Abstract: Almost all countries in the world, including Indonesia [1], strive to develop their tourism potential to earn as much foreign exchange as possible. The role of tourism is very important for a country/region because it has a very broad multiplier effect. Efficiently managed hotels are able to win the competition. Therefore, making the right forecasting and estimation model is of great help for hotel managers to manage hotels effectively and efficiently. For that reason, software development is needed for forecasting and estimation systems. This research tries to synergize management discipline and mathematics so that it can be used more easily, accurately, effectively and efficiently in hotel management. Some reliable forecasting methods among others are Support Vector Machine (SVM) and K-Nearest Neighbors (K-NN). This paper optimizes hotel management through the application of occupancy forecasting by using the SVM and K-NN methods. The simulation results by the RapidMiner software and both methods using 90% of training data and 10% of testing data show that the RMSE produced by the SVM method is 0.011, while the RMSE produced by the KNN method is 0.116. Thus, the SVM method has higher accuracy than the K-NN.

Keywords: *occupancy; forecast; SVM; K-NN.*

Mathematics Subject Classification (2010): 68T45, 68T10.

* Corresponding author: mailto:bambang.suharto@vokasi.unair.ac.id

1 Introduction

Almost all countries in the world, including Indonesia [1], strive to develop their tourism potential to earn as much foreign exchange as possible. The role of tourism is very important for a country/region because it has a very broad multiplier effect. Efficiently managed hotels are able to win the competition. Therefore, making the right forecasting and estimation model is of a great help for hotel managers to manage hotels effectively and efficiently, so software development is needed for forecasting and estimation systems [2–4].

In general, forecasting can be grouped into quantitative forecasting and qualitative forecasting. Qualitative forecasting which is based on intuition and empirical experience becomes subjective. Subjective forecasting is difficult to implement due to the limitations of the human brain in analyzing information and causal relationships that affect the business. If such qualitative forecasting is done by several people separately, the results are likely to have considerable variation. On the other hand, if it is carried out jointly, it is likely that there is no similarity in the forecasting results, or an influential person in the group determines the results [5].

Some previous research used Forecast to help predict the desired number or data. In 2021, the one conducted by Rini et al. [6] used forecast to predict the number of dengue cases, in 2022, Anshori [7] forecasted the number of restaurants to be closed using ANFIS, while Susanto et al. [2] used Backpropagation Neural Network. In 2023, Anshori [1], did a hotel forecast using Linear Support Vector Machine [8, 9]. As for the estimation methods, many were applied in the fields of robotics such as AUV motion, mobile robots and finger arm robots [10, 11].

Some reliable forecasting methods among others are Support Vector Machine (SVM) and K-Nearest Neighbors (K-NN). SVM is a machine learning method able to analyze data and sort it into one of two categories [12]. SVM works to find the best hyperplane or decision boundary function to separate two or more classes in the input space. The hyperplane can be a line in two dimensions and can be a flat plane in multiple planes. Meanwhile, K-NN is an algorithm functioning to classify data based on training data sets, taken from K-Nearest Neighbors. With k, the number of nearest neighbors, the study in this paper is to optimize hotel management through the application of occupancy forecasting by the SVM and K-NN methods [13].

2 Occupancy Data of Hotel Y

Occupancy data of hotel Y are presented in Tables 1-3.

3 Algorithm of K-Nearest Neighbors

K-Nearest Neighbors is a classification algorithm for data grouping into several classes based on the closest distance or similarity of the data to the training data. K-NN performs classification based on the closest distance calculated from Euclidean. Euclidean Distance is a calculation used to find the distance between 2 points in Euclidean space. Euclidean distance calculation is shown in the equation

$$Euc = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}. \tag{1}$$

Notes:
$d(x_1, x_j)$: Euclidean distance;

	2017			2018			2019		
	Room Available	Occ (%)	Room Sold	Room Available	Occ (%)	Room Sold	Room Available	Occ (%)	Room Sold
January	7,006	48,6%	3,262	7,006	46,6%	3,262	7,006	50,5%	3,536
February	4,480	63.7%	2,853	6,328	61.3%	3,880	6,328	63,6%	4,027
March	7,006	72.3%	5,066	7,006	72.3%	5,066	7,006	66,6%	4,663
April	6,780	64.9%	4,402	6,780	64.9%	4,402	6,780	69,9%	4,736
May	7,006	63.4%	4,444	7,006	63.4%	4,444	7,006	57.5%	4,025
June	6,780	61.2%	4,150	6,780	61.2%	4,150	6,780	75,3%	5,106
July	7,006	76.4%	5,351	7,006	76.4%	5,351	7,006	74,6%	5,223
August	7,006	73.8%	5,172	7,006	73.8%	5,172	7,006	61,6%	4,316
September	6,780	70.6%	4,787	6,780	70.6%	4,787	6,780	58,7%	3,977
October	7,006	81.7%	5,723	7,006	81.7%	5,723	7,006	72,3%	5,068
November	6,780	80.3%	5,443	6,780	80.3%	5,443	6,780	79,7%	5,406
December	7,006	70.6%	4,943	7,006	70.6%	4,943	7,006	77,1%	5,401

Table 1: The occupancy data of Hotel Y for 2017–2019.

	2020			2021			2022		
	Room Available	Occ (%)	Room Sold	Room Available	Occ (%)	Room Sold	Room Available	Occ (%)	Room Sold
January	7,006	49.8%	3,487	7,006	44.5%	3,118	7,006	51,1%	3,582
February	6,554	59.8%	3,921	6,328	50.9%	3,224	6,328	49,5%	3,132
March	7,006	42.7%	2,992	7,006	65.9%	4,614	7,006	57.4%	4,023
April	6,780	16.8%	1,140	6,780	61.9%	4,200	6,780	42,6%	2,891
May	7,006	18.7%	1,308	7,006	52.8%	3,698	7,006	76,4%	5,353
June	6,780	27.0%	1,830	6,780	70.6%	4,790	6,780	65,7%	4,455
July	7,006	37.1%	2,600	7,006	32.0%	2,242	7,006	60,6%	4,246
August	7,006	46.1%	3,227	7,006	40.6%	2,845	7,006	52,1%	3,650
September	6,780	45.3%	3,070	6,780	68.7%	4,658	6,780	53,4%	3,618
October	7,006	62.0%	4,341	7,006	80.7%	5,652	7,006	62,2%	4,360
November	6,780	76.9%	5,212	6,780	71.0%	4,816	6,780	68,6%	4,652
December	7,006	66.8%	4,682	7,006	81.0%	5,672	7,006	71,2%	4,985

Table 2: The occupancy data of Hotel Y for 2020–2022.

	2023		
	Room Available	Occ (%)	Room Sold
January	7,006	48,1%	3,372
February	6,328	52,1%	3,294
March	7,006	57,4%	4,023
April	6,780	36,0%	2,441
May	7,006	62,0%	4,345

Table 3: The occupancy data of Hotel W for 2023.

x_i : record data to i;
x_j : record data to j;
a_r : data to-r from i, j.

3.1 Algorithm of Support Vector Machine

Since then, SVMs have been used in text, hypertext and image classification. SVMs can work with handwritten characters and the algorithm has been used in biology labs to perform tasks such as protein sorting. SVMs work to find the best hyperplane or decision boundary function to separate two or more classes in the input space. The hyperplane can be a line in two dimensions and can be a flat plane in multiple planes.

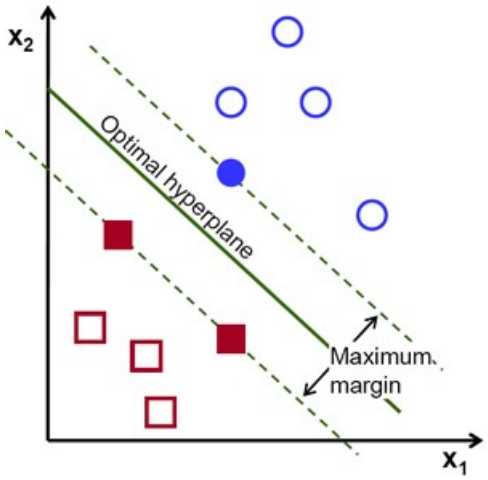

Figure 1: Support Vector Machine Model.

The algorithm of SVM is shown in Figure 2.

3.2 Implementation of the K-Nearest Neighbors Algorithm on data

The implementation of the K-NN algorithm on hotel occupancy data in the RapidMiner software is shown in Figure 3.

3.3 Implementation of Support Vector Machine Algorithm on Data

The application of the SVM method to the occupancy data model is shown in Figure 4.

4 Simulation Results

In this paper, the simulation used two algorithms, where both the SVM and K-NN methods were compared when using several types of training data and testing data. There are 3 types of training and testing data grouping cases, that is, the first case (case 1) in which the training data is 70% and the testing data is 30%, the second case in which the training data is 80% and the testing data is 20%, and the third case in which the training data is 90% and the testing data is 10%. After modeling in the RapidMiner software using the occupancy data of Hotel Y, the SVM and K-NN methods were implemented, resulting in Figures 5 – 7.

Figure 5 is as an explanation related to case 1, with 70% of training data and 30% of the testing data. It can be seen that the forecasting results by the SVM method have a

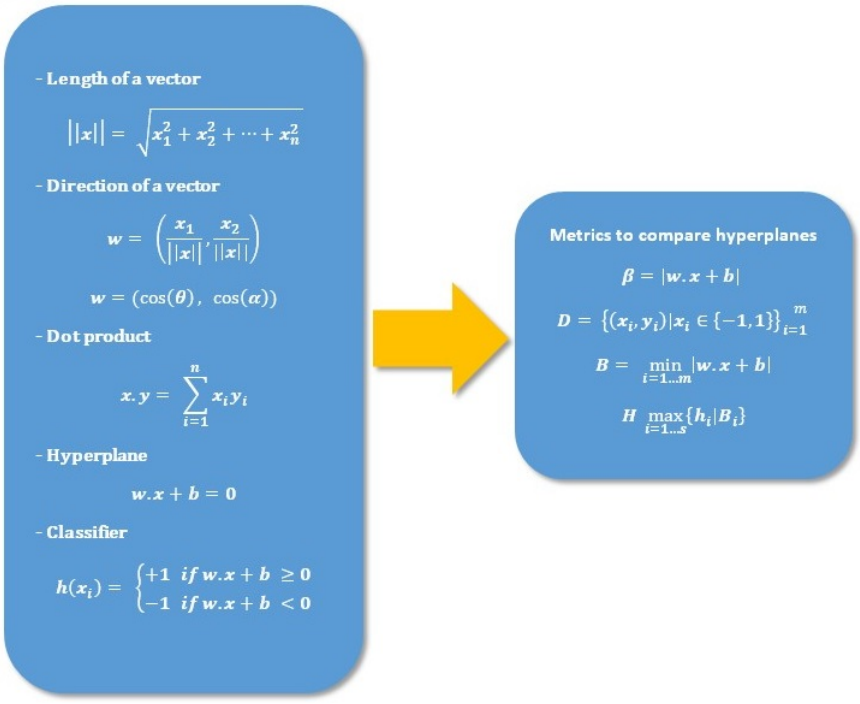

Figure 2: Algorithm of Support Vector Machine.

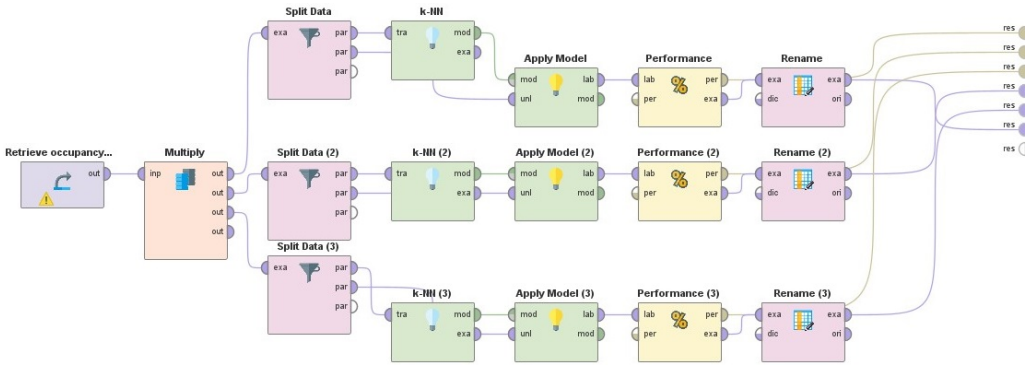

Figure 3: Implementation of the K-NN algorithm on occupancy data of hotel Y.

smaller error than those by the K-NN because the simulation-resulted graph shows that the SVM method has almost the same value as the real data. It can be seen in Table 2 that the RMSDE by the SVM method is 0.014 and that by the K-NN method is 0.137. So, the SVM method has a smaller error of about 0.1.

Figure 6 is as an explanation related to the first case (case 1) with 80% of training data and 20% of testing data, it can be seen that the forecasting results by the SVM method have a smaller error than those by the K-NN because the simulation-resulted

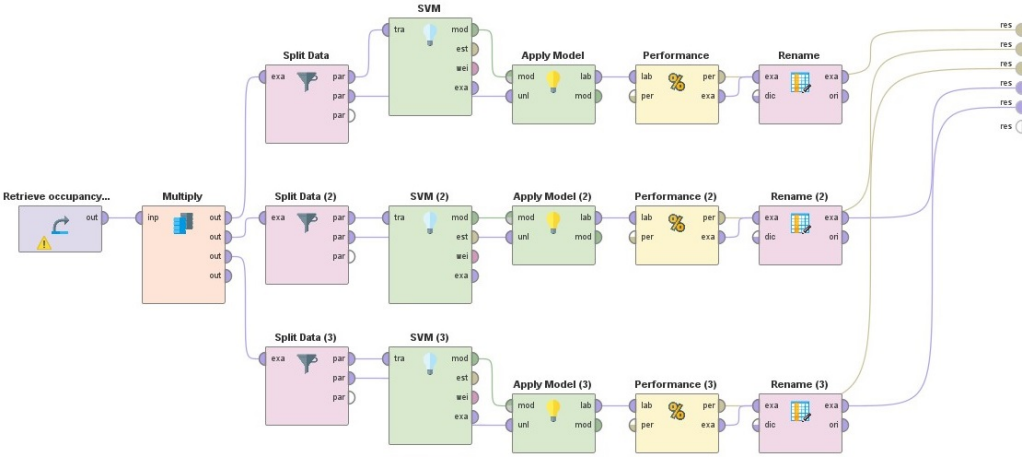

Figure 4: Implementation of the SVM Algorithm on occupancy data of Hotel Y.

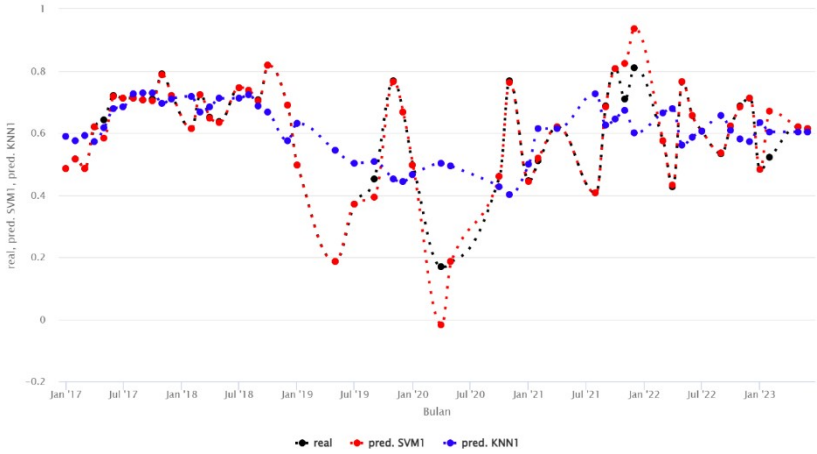

Figure 5: Simulation Results of hotel Y occupancy forecast using 70% of training data and 30% of testing data.

graph shows that the SVM method has almost the same value as the real data. It can be seen in Table 2 that the RMSE by the SVM method is 0.029 and that by the K-NN method is 0.134. So, the SVM method has a smaller error of about 0.1.

Figure 7 is as an explanation related to the first case (case 1), with 90% of training data and 10% of testing data. It can be seen that the results of forecasting by the SVM method have a smaller error than those by K-NN because the simulation-resulted graph shows that the SVM method has almost the same value as the real data. It can be seen in Table 2 that the RMSE by the SVM method is 0.011 and the RMSE by the K-NN method is 0.116. So, the SVM method has a smaller error of about 0.1.

Table 4 above shows that the SVM method has a higher accuracy and a smaller error

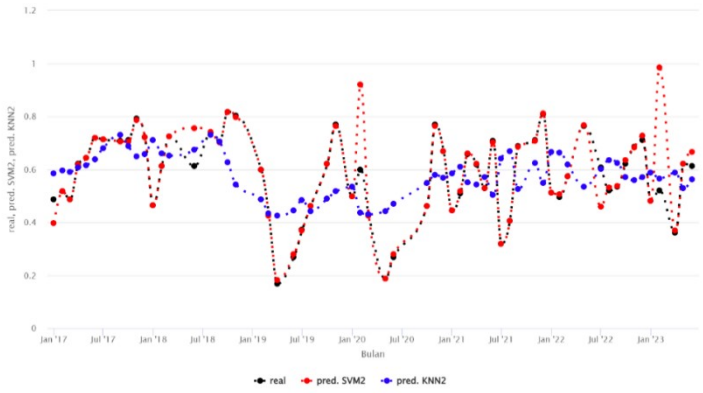

Figure 6: Simulation results of hotel Y occupancy forecast using 80% of training data and 20% of testing data.

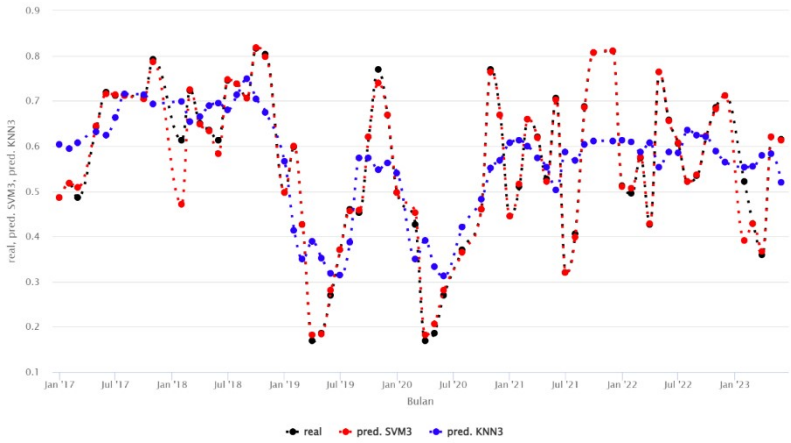

Figure 7: Simulation results of hotel Y occupancy forecast occupancy using 90% of training data and 10% of testing data.

	SVM			KNN		
	70% Training data and 30% Testing data	80% Training data and 20% Testing data	90% Training data and 10% Testing data	70% Training data and 30% Testing data	80% Training data and 20% Testing data	90% Training data and 10% Testing data
RMSE of Forecasting results	0.014	0.029	0.011	0.137	0.134	0.116

Table 4: The RMSE comparison using SVM and K-NN.

than the K-NN for all three cases. When compared, overall, it is clear that using SMV in case 3 shows the smallest RMSE value. When considering the SMV method alone, case 3 has the smallest RMSE value because the training data is larger than those of cases 1 and 2. Likewise, for the K-NN, case 3 has a smaller RMSE value than cases 1 and 2.

5 Conclusion

Based on the results of the discussion above and the forecasting results as in the graph above, it can be concluded that the SVM method has a higher accuracy and a smaller error than the K-NN method for all three cases. When compared, as a whole, using SMV in case 3 shows the smallest RMSE value. When considering the SVM method alone, case 3 has the smallest RMSE value because the training data is larger than those of cases 1 and 2. Likewise, for K-NN, case 3 has a smaller RMSE value than cases 1 and 2. Thus, the SVM method is highly reliable for hotel occupancy forecasting.

Acknowledgment

High appreciation to the Kemdikbudristek for the very fund support for the completion of the research conducted in the year of 2024 with contract number 109/E5/PG.02.00.PL/2024, 054/SP2H/PT/LL7/2024, and 1104/UNUSA-LPPM/Adm.I/VI/2024.

References

[1] M. Y. Anshori, I. H. Santoso, T. Herlambang, D. Rahmalia, K. Oktafianto and P. Katias. Forecasting of Occupied Rooms in the Hotel Using Linear Support Vector Machine. *Nonlinear Dynamics and Systems Theory* **23** (2) (2023) 129–140.

[2] F. A. Susanto, M. Y. Anshori, D. Rahmalia, K. Oktafianto, D. Adzkiya, P. Katias and T. Herlambang. Estimation of Closed Hotels and Restaurants in Jakarta as Impact of Corona Virus Disease (Covid-19) Spread Using Backpropagation Neural Network. *Nonlinear Dynamics and Systems Theory* **22** (4) (2022) 457–467.

[3] M. Y. Anshori, T. Shawyun, D. V. Madrigal, D. Rahmalia, F. A. Susanto, T. Herlambang and D. Adzkiya. Estimation of closed hotels and restaurants in Jakarta as impact of corona virus disease spread using adaptive neuro fuzzy inference system. *International Journal of Artificial Intelligence (IJAI)* **11** (2) (2022).

[4] D.Rahmalia, T. Herlambang, A. S. Kamil, R. A. Rasyid, F. Yudianto, L. Muzdalifah and E. F. Kurniawati. Profitability estimation of XYZ company using H-infinity and Ensemble Kalman Filter. In: *The Third International Conference on Combinatorics, Graph Teory and Network Topology* University of Jember-Indonesia, 26-27 Oct. 2019.

[5] D. B. Maghfira, R. Sarno, F. A. Susanto, T. Herlambang, K. Oktafianto, W. Hartawan and I. W. Farid. Electronic Nose for Classifying Civet Coffee and Non-Civet Coffee. *Nonlinear Dynamics and Systems Theory* **23** (3) (2023) 323–337.

[6] F. S. Rini, T. D. Wulan and T. Herlambang. Forecasting The Number of Demam Berdarah Dengue (DBD) Patients Using The Fuzzy Method At The Siwalankerto Public Health Center. In: *The 1st International Conference on Neuroscience and Learning Technology (ICONSATIN 2021)* Jember, Indonesia, 18-19 September 2021.

[7] M. Y. Anshori, D. Rahmalia, T. Herlambang and D. F. Karya. Optimizing Adaptive Neuro Fuzzy Inference System (ANFIS) parameters using Cuckoo Search (Case study of world crude oil price estimation). In: *The 4th International Conference on Combinatorics, Graph Theory, and Network Topology (ICCGANT) 2020* East Java, Indonesia, 22-23 August 2020.

[8] M. Y. Anshori, D. Rahmalia and T. Herlambang. Comparison Backpropagation (BP) and Learning Vector Quantification (LVQ) on classifying price range of smartphone in market. In: *The 4th International Conference on Combinatorics, Graph Theory, and Network Topology (ICCGANT) 2020* East Java, Indonesia, 22-23 August 2020.

[9] M. Y. Anshori, T. Herlambang, P. Katias, F. A. Susanto and R. R. Rasyid. Profitability estimation of XYZ company using H-infinity and Ensemble Kalman Filter. In: *The 5th International Conference of Combinatorics, Graph Theory, and Network Topology (ICCGANT 2021)* Jember, Indonesia, 21-22 August 2021.

[10] V. N. Vapnik and A. Y. Lerner. Pattern recognition using generalized portraits. *Automation and Remote Control* **24** (1963) 709–715.

[11] T. Herlambang, S. Syamsuar, F. Yudianto, A. Basuki, G. Wijiatmoko, A. Roschyntawati, H. Hendrato, K. Oktafianto and R. S. Marjianto. Trajectory Estimation of Amphibious Aircraft Using H-Infinity and Ensemble Kalman Filter Methods. *Nonlinear Dynamics and System Theory* **24** (3) (2024) 259–266.

[12] M. Y. Anshori, T. Herlambang, V. Asyari, H. Arof, A. A. Firdaus, K. Oktafianto and B. Suharto. Optimization of Hotel W Management through Performance Comparison of Support Vector Machine and Linear Regression Algorithm in Forecasting Occupancy. *Nonlinear Dynamics and System Theory* **24** (3) (2024) 228–235.

[13] A. Muhith, I. H. Susanto, D. Rahmalia, D. Adzkiya and T. Herlambang. The Analysis of Demand and Supply of Blood in Hospital in Surabaya City Using Panel Data Regression. *Nonlinear Dynamics and System Theory* **22** (5) (2022) 550–560.