Nonlinear Dynamics and Systems Theory, **24** (3) (2024) 228–235

Optimization of Hotel W Management through Performance Comparison of Support Vector Machine and Linear Regression Algorithm in Forecasting Occupancy

M. Y. Anshori [1], T. Herlambang [2], V. Asy'ari [2], H. Arof [3], A. A. Firdaus [4], K. Oktafianto [5] and B. Suharto [6*]

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

Received: September 5, 2023; Revised: April 19, 2024

Abstract: The hospitality industry continues to grow globally. More and more people are traveling for various purposes such as leisure, business, or special events. This growth has created a huge opportunity for hotels to increase revenue and profits. Due to the growth of the industry, competition among hotels has also intensified. So, this condition encourages hotels to look for efficient ways of managing hotel resources. One way to efficiently manage a hotel is by forecasting the hotel occupancy. The study in this paper is aimed to optimize hotel management through the application of occupancy forecasting by the SVM and linear regression methods. The results indicated that the linear regression method had a higher accuracy and a smaller error than the SNM for two cases. When compared, as a whole, using linear regression in case 2 had the smallest RMSE value, in which the difference in RMSE by the linear regression method is around 0.3 -0.9 smaller than that by the SVM method.

Keywords: *hotel management; SVM; linear regression algorithm; forecasting occupancy.*

Mathematics Subject Classification (2010): 62J05, 70-10, 90Bxx.

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

1 Introduction

Hospitality industry continues to grow globally. More and more people are traveling for various purposes such as leisure, business, or special events. This growth has created a huge opportunity for hotels to increase revenues and profits. Due to the growth of the industry, competition among hotels has also intensified [1]. Hotels must compete to attract the attention of potential customers and retain existing guests. This encourages hotels to find innovative ways to improve their services and operational efficiency. On the other hand, hotel operating costs, including employee salaries, raw materials, energy, and property maintenance, continue to increase. Such condition compels them to find any possible efficient ways of managing hotel resources.

One way to efficiently manage a hotel is by forecasting the hotel occupancy [2]. Before the 1980s, forecasting was still seen as a technical activity in the western world. While in the 1990s, the same view was still felt among Indonesian business people and those in other developing countries. Whereas in its place of origin (the United States), the scope of forecasting has grown rapidly beyond its technical nature, encompassing an extensive use in planning, decision-making, and other managerial disciplines [3].

The main purpose of the forecasting process in general is to make estimates or predictions about future events or data based on available historical information. The function of forecasting is among others to assist organizations in a long-term and short-term strategic planning by providing a view of future demand, sales, production, or other needs and to support the development of business strategies by providing a better understanding of market trends, customer behavior, and several other business development and research studies used forecast methods to help predict the desired figures or data. Several publications related to forecasting were applied in the health sector for forecasting the spread of Dengue Fever (DHF) [4]. Forecasting was also used to detect coffee aroma [5]. Forecasting researches using the Backpropagation Neural Network [6, 7], ANFIS and Linear Support Vector Machine methods were also applied for forecasting in the economic [8, 9] and tourism fields to support government policies related to tourism development.

Some reliable forecasting methods include Support Vector Machine (SVM) and linear regression. SVM is a machine learning method which can be used to analyze data and sort it into one of two categories. SVM was originally developed for classification problems. It can be used to separate data into two classes. However, SVM can also be applied to regression problems whose goal is to predict numerical values. In the context of data forecasting, SVM is used for both classification (e.g., predicting whether an event will occur or not) and regression (e.g., predicting stock prices). Linear regression analysis is an approach for modeling the relationship between one dependent variable and one independent variable. The main purpose of the Linear Regression method is to model the linear relationship between one or more independent variables (also referred to as predictor variables or features) and the dependent variable (also referred to as the target variable) so as to make predictions or estimates. And, in this paper, it is for optimizing hotel management through the application of occupancy forecasting by the SVM and linear regression methods.

2 Occupancy Data of Hotel W

The following tables present the occupancy data of Hotel W.

	2017			2018			2019		
	Room Available	Occ (%)	Room Sold	Room Available	Occ (%)	Room Sold	Room Available	Occ (%)	Room Sold
Jan				2,201	27.1%	596	2,201	50.3%	1,108
Feb				1,988	32.3%	643	1,988	76.3%	1,516
Mar				2,201	51.7%	1,139	2,201	82.8%	1,822
Apr				2,130	43.0%	916	2,130	71.3%	1,518
May				2,201	52.8%	1,162	2,201	74.6%	1,641
Jun				2,130	40.7%	866	2,130	72.0%	1,533
Jul				2,201	41.7%	918	2,201	63.1%	1,389
Aug				2,201	57.5%	1,265	2,201	69.8%	1,536
Sep				2,130	49.4%	1,052	2,130	68.9%	1,468
Oct	1,988	36.8%	732	2,201	58.6%	1,290	2,201	68.3%	1,504
Nov	2,130	53.1%	1,132	2,130	53.4%	1,137	2,130	66.5%	1,416
Dec	2,201	52.9%	1,164	2,201	47.7%	1,050	2,201	66.4%	1,461

Table 1: The occupancy data of Hotel W for 2017-2019.

	2020			2021			2022		
	Room Available	Occ (%)	Room Sold	Room Available	Occ (%)	Room Sold	Room Available	Occ (%)	Room Sold
Jan	2,201	55.8%	1,229	2,201	37.3%	820	2,201	31.2%	687
Feb	2,059	65.4%	1,346	1,988	36.6%	728	1,988	35.0%	695
Mar	2,201	46.5%	1,023	2,201	39.6%	871	2,201	43.0%	947
Apr	2,130	9.9%	210	2,130	41.0%	874	2,130	37.5%	798
May	2,201	0.0%	0	2,201	37.6%	827	2,201	47.2%	1,038
Jun	2,130	12.2%	259	2,130	44.8%	955	2,130	34.8%	741
Jul	2,201	28.7%	632	2,201	36.2%	797	2,201	53.2%	1,170
Aug	2,201	25.4%	560	2,201	38.0%	836	2,201	48.0%	1,056
Sep	2,130	24.1%	514	2,130	46.9%	1,000	2,130	44.0%	938
Oct	2,201	27.1%	597	2,201	48.6%	1,070	2,201	52.9%	1,165
Nov	2,130	33.4%	711	2,219	34.1%	756	2,130	52.5%	1,118
Dec	2,201	45.0%	990	2,201	39.5%	870	2,201	61.4%	1,352

Table 2: The occupancy data of Hotel W for 2020-2022.

	2023		
	Room Available	Occ (%)	Room Sold
Jan	2,201	58.4%	1,286
Feb	1,988	52.2%	1,038
Mar	2,201	43.0%	947
Apr	2,130	60.1%	1,280
May	2,201	52.6%	1,158

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

3 Algorithm of Support Vector Machine

SVM was invented in [11]. 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.

The algorithm of SVM is shown in Figure 2.

3.1 Linear Regression method

Simple linear regression analysis is an approach method for modeling the relationship between one dependent variable and one independent variable. In regression, the independent variable explains the dependent variable. In a simple regression analysis, the relationship between variables is linear, in which the changes in the variable X are followed by those in the variable Y in a fixed manner. While in a non-linear relationship, the changes in the variable X are not followed by those in the variable Y proportionally [10].

A simple linear regression analysis model is

$$Y = a + bX + \varepsilon, \tag{1}$$

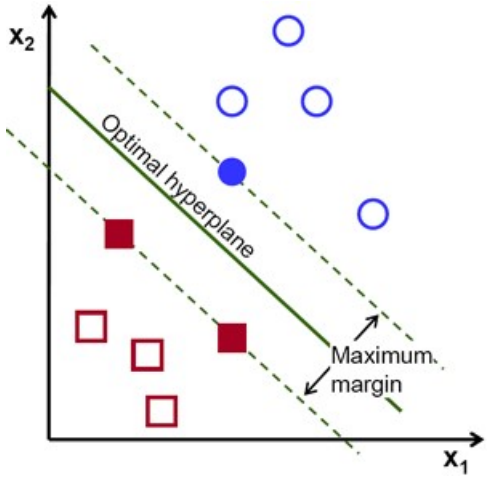

Figure 1: Support Vector Machine Model [1].

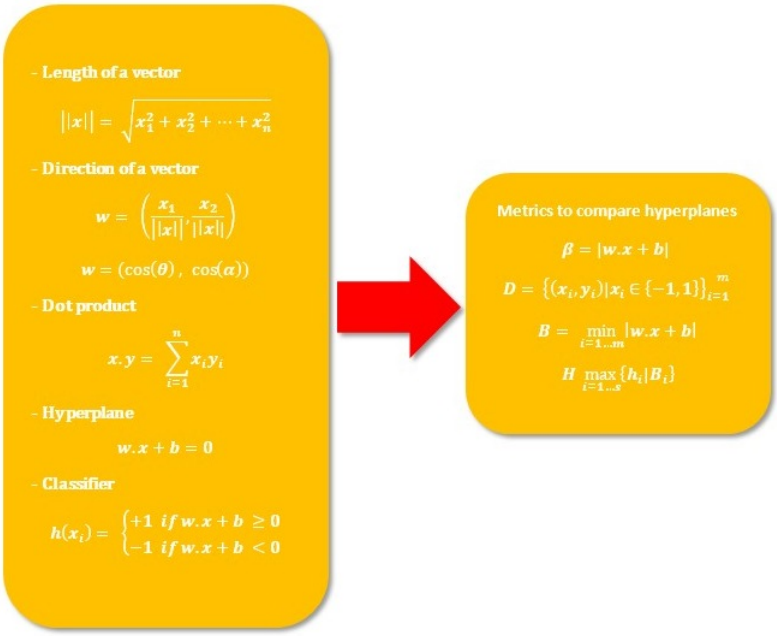

Figure 2: Algorithm of Support Vector Machine.

where Y is the predicted value, a is the constant, b is the regression coefficient, X is the independent variable, ε is the residual value.

The value $a = \frac{\Sigma Y - b(\Sigma X)}{n}, b = \frac{n(\Sigma XY) - (\Sigma X)(\Sigma Y)}{n(\Sigma X^2) - (\Sigma X)^2}$.

4 Simulation Results

4.1 Implementation of SVM and linear regression algorithms on data

Below is the implementation of the SVM and linear regression algorithms for hotel occupancy data in the RapidMiner software.

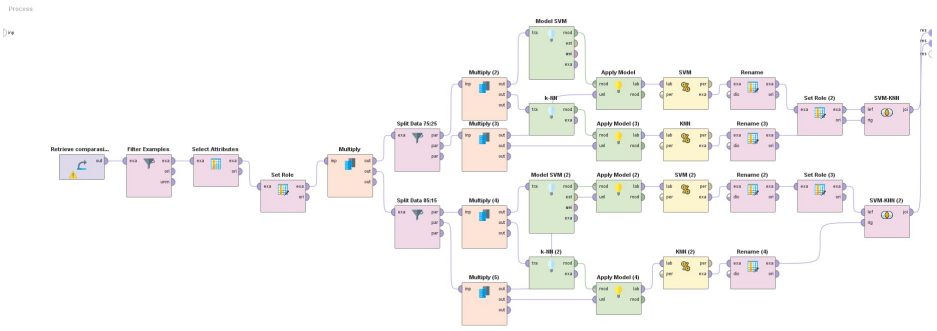

Figure 3: Implementation of the SVM and linear regression algorithms for hotel occupancy.

4.2 Simulation results

In the simulation, this study used two algorithms, the SVM and linear regression algorithms, the results of which were compared for several types of training data and testing data. There were 2 types of training and testing data grouping cases, that is, case 1 with 75% of training data and 25% of testing data and case 2 (the second case) with 85% of training data and 15% of testing data. After modeling in the RapidMiner software, using Hotel W occupancy data, the SVM and linear regression methods were then implemented, resulting in Figures 4 and 5.

Figure 4 is an explanation related to case 1 with 70% of training data and 30% of testing data, indicating that the H-infinity forecasting results by the SVM method have a bigger error than those by the linear regression method because the simulation results graph shows that the SVM method still has a difference from the real data. It can be seen in Table 4 that the RMSE produced by the SVM method is 0.946 and that by the linear regression method is 0.005. Thus, the linear regression method has a smaller error of about 0.9.

Figure 5 is an explanation related to case 2 with 85% of training data and 15% of testing data. It shows that the forecasting results by the SVM method have a bigger error than those by the linear regression method because the simulation results graph shows that the SVM method still has a difference from the real data. It can be seen in Table 4 that the RMSE produced by the SVM method is 0.371 and that by the linear regression method is 0.004. Thus, the linear regression method has a smaller error of about 0.3.

Based on Table 4, it can be seen that the linear regression method has a higher accuracy and a smaller error than the SVM for both cases. When compared, as a whole, the linear regression method for case 2 has the smallest RMSE value. When considering the SVM method alone, case 2 has the smallest RMSE value because the training data

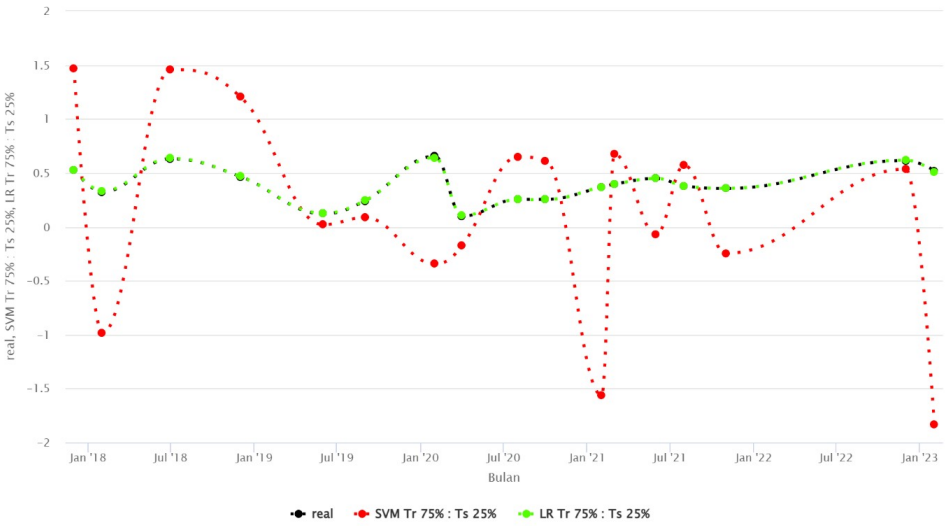

Figure 4: Hotel Y occupancy forecast simulation results with 70% of training data and 30% of testing data.

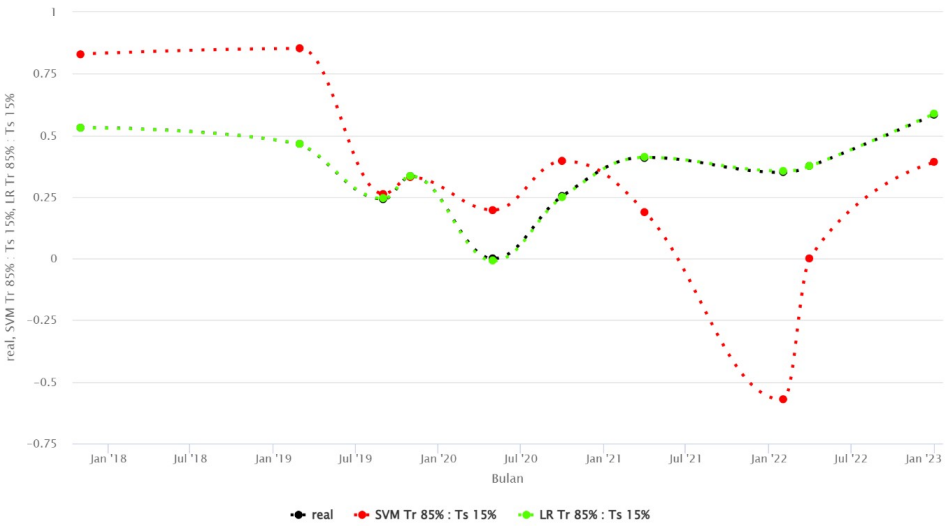

Figure 5: Simulation results of Hotel Y occupancy forecast with 80% of training data and 20% of testing data.

is larger than that in case 1. Likewise, the linear regression method in case 2 has a smaller RMSE value than that in case 1. The Linear Regression method is often used in various fields such as economics, social science, natural science, finance, and engineering. However, it is important to remember that the Linear Regression has certain assumptions

	SVM		Linear Regression	
	75% Training data and 30% Testing data	85% Training data and 15% Testing data	75% Training data and 30% Testing data	85% Training data and 15% Testing data
RMSE of Forecasting results	0.946	0.371	0.005	0.004

Table 4: Comparison of RMSE generated by the SVM and linear regression methods.

to be met for valid results, for example, the assumption that the relationship between the variables is linear and normally distributed. At the same time, the main advantage of the SVM is its ability to handle non-linear data, in this case, using hotel occupancy data which is classified as linear data. So, the linear regression method has a higher accuracy than the SVM method.

5 Conclusion

Based on the results of the discussion above and the forecasting results graph above, it can be concluded that the linear regression method has a higher accuracy and a smaller error than the SNM for two cases. When compared, as a whole, the linear regression in case 2 has the smallest RMSE value. When considering the SVM method alone, in case 2, it has the smallest RMSE value because the training data is larger than that in case 1. Likewise, the linear regression for case 2 has a smaller RMSE value than it has for case 1. The Linear Regression is often used in various fields such as economics, social science, natural science, finance, and engineering. However, it is important to remember that Linear Regression has certain assumptions to be met for the results to be valid, for example, the assumption that the relationship between the variables is linear and normally distributed. At the same time, the main advantage of the SVM is its ability to handle non-linear data, in this case, using hotel occupancy data which are classified as linear data, therefore the linear regression method is more accurate than the SVM. However, the SVM method is very reliable for forecasting hotel occupancy.

Acknowledgment

High appreciation to the Kemdikbudristek for the very fund support for the research conducted in the year of 2023 with contract number 183/E5/PG.02.00.PL/2023,049/SP2H/PT/LL7/2023 and 1287/UNUSA-LPPM/Adm-I/VII/2023.

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] 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.

[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] 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.

[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, 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.

[9] 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.

[10] 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. In: *Nonlinear Dynamics and System Theory* **22** (5) (2022) 550–560.

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