Nonlinear Dynamics and Systems Theory, **25** (4) (2025) 373–385

Comparison of K-Nearest Neighbor and Neural Network for Forecasting Occupancy Rate at Hotel XYZ

M. Y. Anshori [1]*, P. Katias [1], T. Herlambang [2], M. S. Azmi [3], Z. B. Othman [4] and K. Oktafianto [5]

[1] *Department of Management, Universitas Nahdlatul Ulama Surabaya, Indonesia.*
[2] *Department of Information Systems, Universitas Nahdlatul Ulama Surabaya, Indonesia.*
[3] *Department of Software Engineering, Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia.*
[4] *Department of Diploma Studies, Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia.*
[5] *Department of Mathematics, University of PGRI Ronggolawe, Indonesia.*

Received: July 12, 2024; Revised: July 4, 2025

Abstract: The occupancy rate of a hotel is an important factor to see the development of providers business performance. By forecasting occupancy rate, the hotel can identify business opportunities or adjust room prices, determine hotel operations, and take this into consideration for strategic decision making. In this study, occupancy rate forecasting for Hotel XYZ was carried out by comparing the k-nearest neighbor (k-NN) and neural network methods. The dataset used in this study included rooms available, rooms sold out, and available occupancy percentage data in Hotel XYZ from April 2018 to June 2023. The simulation was carried out by dividing the data into training data and testing data with a ratio of 70:30 and 80:20. Model creation was carried out by applying the k-NN and neural network methods to the Hotel XYZ data set. Forecasting results that were obtained using k-NN showed an optimal RMSE at 70%:30% split of data with an RMSE of 0.080 at k-value 3, while forecasting results obtained using the neural network showed an optimal RMSE at 70%:30% data split with an RMSE of 0.007 for two hidden layers. The comparison of results of forecasting by k-NN and neural network showed an optimal RMSE when using neural network method with an RMSE of 0.004, a GAP of 0.076 compared to using k-NN. The results of this study can be used by Hotel XYZ to make better decisions in determining hotel policies in the future and goals set by the hotel.

Keywords: *hotel; occupancy rate; forecasting; k-nearest neoghbor; neural network.*

Mathematics Subject Classification (2020): 68T45, 68T10.

* Corresponding author: mailto:yusak.anshori@unusa.ac.id

1 Introduction

Business development in the hotel sector in Indonesia has shown a positive trend, as can be seen from the increasing occupancy rate. The occupancy rate of star classified hotels in September 2023 reached an average of 53.02%, an increase of 3.00 points compared to the occupancy rate in September 2022 of 50.02% and YTD 2023 reached an average of 49.43%, an increase of 4.04 compared to the YTD 2022 occupancy rate of 45.40% [1]. The occupancy rate of non-star classified hotels in September 2023 reached an average of 24.82%, an increase of 1.43 points compared to the occupancy rate in September 2022 of 23.39% and YTD 2023 reached an average of 23.72%, an increase of 1.52 points compared to the YTD 2022 occupancy rate of 22.21% [2]. Along with this growth, competition among hoteliers is also increasing. In-depth knowledge of the level of competition is crucial to identify opportunities, face challenges, and develop effective strategies to compete competitively.

Occupancy rate is the percentage of sold room occupancy rate [3] and is one of the indicators that can be used to see the development of business efficiency of hotel service providers in a certain period [4]. This occupancy can be calculated before the current date, which is commonly called a forecast, or after the date passed [5]. By forecasting occupancy rate, the hotel can identify business opportunities or adjust room prices, determine hotel operations, and take this into consideration for strategic decision making.

In 1951, Evelyn Fix and Joseph Hodges created the k-nearest neighbor algorithm (k-NN) in statistics as a non-parametric supervised learning technique. Regression and classification are two uses for it. The input in both situations consists of a data set's k closest training samples [6]. K-NN is referred to as case-based reasoning, which is a methodology based on reasoning of cases in terms of training data of a case stored, trained, and accessed to solve new problems [7]. K-NN makes firm predictions on test data based on k-nearest neighbor comparisons. The near or far of neighbors is usually calculated based on Euclidian distance. The best k value for this algorithm depends on the data, usually a high k value will reduce the effect of noise on the application [8].

Neural networks have been around since 1943, when Warren McCulloch and Walter Pitts introduced the first neural network model calculations. This model describes the way artificial neurons can be used to process information binary. In 1950, Frank Ronseblatt continued his research by discovering a two-layer network called a perceptron. A neural network is a model with a flexible function structure, so the neural network model is rapidly developing and has been widely applied in various fields. Neural networks can be used to find solutions to problems when classical methods prove difficult or fail frequently [9].

In the previous studies, several forecasting methods were applied for the estimation of closed hotels and restaurants in Jakarta because of corona virus disease spread using an adaptive neuro fuzzy inference system [10], forecasting the number of Demam Berdarah Dengue (DBD) patients using the fuzzy method [11], predicting the number of visitors per period to beach attractions using triple exponential smoothing [12], classifying the price range of smartphones in the market using backpropagation and Learning Vector Quantification (LVQ) [13], for stock price estimation using Unscented Kalman Filter (UKF) [14], forecasting of occupied rooms in the hotel using linear support vector machine [15], profitability estimation using H-Infinity and Ensemble Kalman Filter (EnKF) [16], the analysis of demand and supply of blood in hospital in Surabaya city using panel data regression [17], prediction of sunlight intensity using neural network and Adaptive Neuro Fuzzy Inference System (ANFIS) [18], estimation of closed hotels and restaurants because of Covid-19 spread using backpropagation neural network [19]. See also the neural network algorithm for breast cancer diagnosis [20], electronic nose for classifying civet coffee using Support Vector Machine (SVM), k-nearest neighbors (k-NN), and decision tree [21], forecasting agricultural products in Malang Regency using k-NN [22], forecasting average room rate using k-NN [23], forecasting occupancy rate using neural network [24]. In this study, the k-NN and neural network methods with data ratios of 70%: 30% and 80%: 20% are applied for forecasting occupancy rate at Hotel XYZ so that the results can be used in identifying opportunities, operational implementation, and strategic decision making for management.

2 Research Method

2.1 Occupancy rate

Occupancy rate is the level of occupancy of hotel rooms calculated based on the number of rooms rented by guests and compared to the number of rooms available in a certain period [5]. The calculation of occupancy rate is shown in (1) with the result of the calculation being a percentage index measured from 0% to 100%.

$$Occupancy\ Rate\ \% = \frac{RoomSold}{RoomAvailable} \times 100\%. \tag{1}$$

The occupancy shows the number of rooms sold out of the number of available rooms. This ratio can fluctuate every day, the ratio in one month or one year is the average percentage of rooms sold. The highest occupancy rate is the best indicator for the hotel and a measure of the success of hotel operations.

2.2 K-Nearest neighbor

K-nearest neighbor (k-NN) is a classification technique that makes firm predictions on test data based on the comparison of K nearest neighbors [8]. The nearest neighbor is the trained object that has the greatest similarity value or the smallest dissimilarity with the previous data. The number of nearest neighbors is expressed by the value of k. The best k value depends on the data. In general, a high k value will reduce the effect of noise on classifications but make the boundaries between each classification even more blurred [25].

The purpose of the k-NN algorithm is to classify new objects based on attributes and training samples. The results of the new test sample are classified based on the majority of categories in k-NN using the neighborhood classification as the predictive value of the new test data sample. The distance used is Euclidean Distance with the following formula:

$$d_i = \sqrt{\Sigma_{i=1}^{n}(p_i - q_i)^2}, \tag{2}$$

$$d_i = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \ldots + (p_n - q_n)^2}, \tag{3}$$

where p_i is the sample data; q_i is the test data; i is the data variable; d is the distance; n is the data dimension.

The working principle of k-NN is to find the closest distance between the data to be evaluated and the nearest k (neighbor) in the training data. The sequence of k-NN work processes [25], [26] is as follows. Specify the parameter k (the number of the closest neighbors). Calculate the square of each object's Euclid jar (query instance) against the given sample data using an equation. Then sort the objects into groups that have the smallest Euclidean distance. Collect the nearest neighbor classification category.

By using the nearest neighbor category, that is, the most majority, one can predict the value of the query instance that has been calculated.

2.3 Neural network

The basis of neural networks consists of inputs, weights, processing units, and outputs. Neural networks can be applied to classifying patterns, mapping patterns obtained from inputs into new patterns in outputs, storing patterns to be recalled, mapping similar patterns, optimizing problems, and predicting. Neural networks start from preparing data for training and learning, finding neural network architecture, training and learning processes, and testing processes [27]. Neural networks can be divided into three parts called layers. Input layer, responsible for receiving information, signals, features, or measurements from the external environment. Hidden layers, responsible for extracting patterns related to the process or system being analyzed. The output layer, responsible for producing and presenting the final tissue, results from processing by neurons in the previous layer.

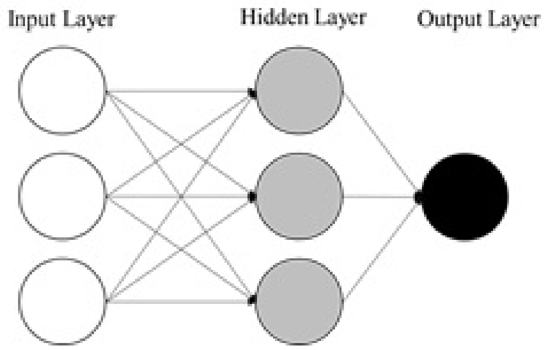

Figure 1: Arrangement of neural networks in layers.

Neural networks can change structures to solve problems based on internal and external information flowing through the network. Neural networks can be used to model the relationship between input and output to find patterns of data. Neurons are a basic part of the processing of a neural network. The basic shape of a neuron can be seen in Figure 2 below.

The weight vector (w) contains weights that connect the various parts of the network. The term "w" is used in the terminology of neural networks and is a suggestion of the expression of connections between two neurons, that is, the weight of information flowing from neuron to other neurons in the neural network. The first stage is the process of summing inputs $x_1, x_2, \ldots, x_n$, which is multiplied by its weight $w_1, w_2, \ldots, w_n$ and is expressed as

$$Net = (w_1x_1 + w_2x_2 + w_3x_3 + \ldots + w_nx_n)\,. \tag{4}$$

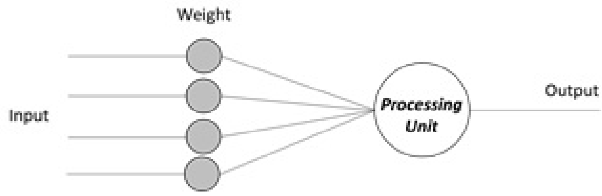

Figure 2: Basic form of neurons.

This concept can be written in vector notation as follows:

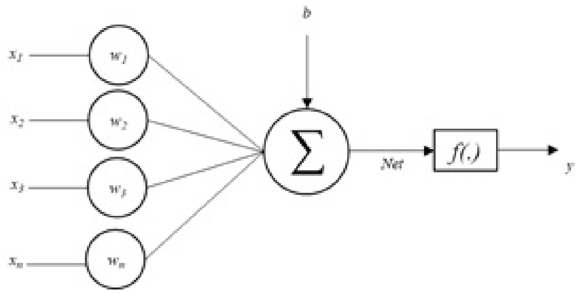

Figure 3: Perceptron model.

A threshold value of b is called a bias, which plays an important role for some neuron models and is referred to as a separate neuron model parameter. Various input conditions and influences on output are required to include a nonlinear activation function $f(\cdot)$ in the arrangement of neurons [28]. This aims to achieve an adequate level if the input signal is small and avoid the risk of output going to inappropriate limits. Like the perceptron model in Figure 2, the output of the neuron can be expressed in terms of $y = f(net)$.

3 Result and Discussion

The dataset used in this study includes rooms available, rooms sold out, and available occupancy percentage data in Hotel XYZ. The dataset used is from April 2018 to June 2023 (63 months). Furthermore, the data is split into training data and testing data. Then a test analysis is carried out with k-nearest neighbor and neural network. After that, a comparison of RMSE results from several tests of the algorithm is carried out.

3.1 Making a comparison model of K-Nearest neighbor and neural network

At this stage, the Hotel XYZ dataset testing process is carried out using k-NN and neural network. The Hotel XYZ dataset that has been entered is then selected for the rooms available, rooms sold out, and available occupancy percentage attributes. The Hotel XYZ dataset is further divided into training data and testing data using a ratio of 70 : 30 and 80 : 20. The model is designed using k-NN with k-values 3 to 7 and using the neural network with one hidden layer and two hidden layers.

3.2 K-Nearest neighbor algorithm test analysis

At this stage, a comparison of the results of the test was carried out using the k-NN algorithm with k-values 3 to 7 at the split of data with a percentage ratio of 70 : 30 and 80 : 20. The results of the k-NN algorithm testing can be seen in Table 1.

k-Values	Training Data	Testing Data	RMSE	k-Values Data	Training Data	Testing	RMSE
3	70%	30%	0.080	3	80%	20%	0.102
4	70%	30%	0.079	4	80%	20%	0.101
5	70%	30%	0.082	5	80%	20%	0.098
6	70%	30%	0.086	6	80%	20%	0.106
7	70%	30%	0.090	7	80%	20%	0.120

Table 1: Test Results Using k-NN.

From forecasting carried out using k-NN at 70% : 30% split of data, it is clear that the best RSME is for k-value 3 with an RMSE of 0.080. The comparison of the simulation with k-values 3 to 7 at 70% : 30% data split can be seen in Figure 4.

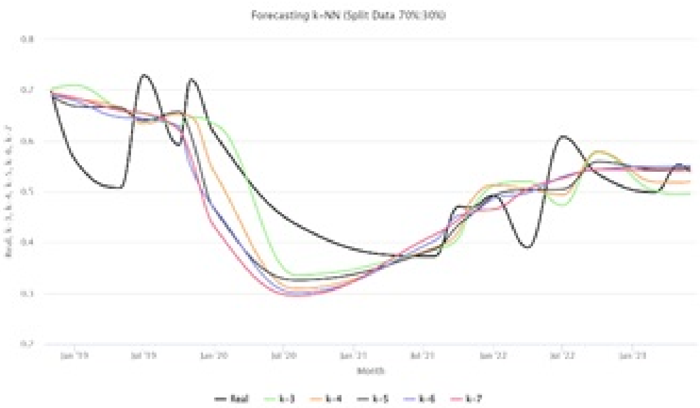

Figure 4: Occupancy rate prediction using k-NN (split of data 70% : 30%).

Figure 5 shows the comparison of real occupancy rate data and occupancy rate prediction at 70% : 30% data split using the k-NN algorithm with k-value 3. In the graph, the black line shows real occupancy data, and the green line shows occupancy forecasting data with k-value 3. There was a significant decrease in August 2021 with an occupancy rate of 0.37 in real data and an occupancy rate of 0.39 in forecasting data. The highest ARR was in July 2019 with an occupancy rate of 0.73 in real data and an occupancy rate of 0.65 in forecasting data.

From forecasting carried out using k-NN at 80% : 20% split of data, it is clear that the best RSME is for k-value 5 with an RMSE of 0.098. The comparison of the simulation with k-values 3 to 7 at 80% : 20% data split can be seen in Figure 6.

Figure 7 shows the comparison of real occupancy rate data and occupancy rate prediction at 80% : 20% data split using the k-NN algorithm with k-value 5. In the graph,

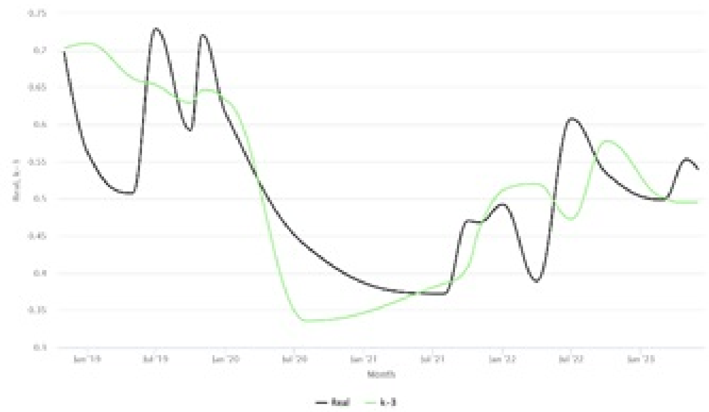

Figure 5: Occupancy rate prediction using k-NN with k-value 3 (split of data 70% : 30%).

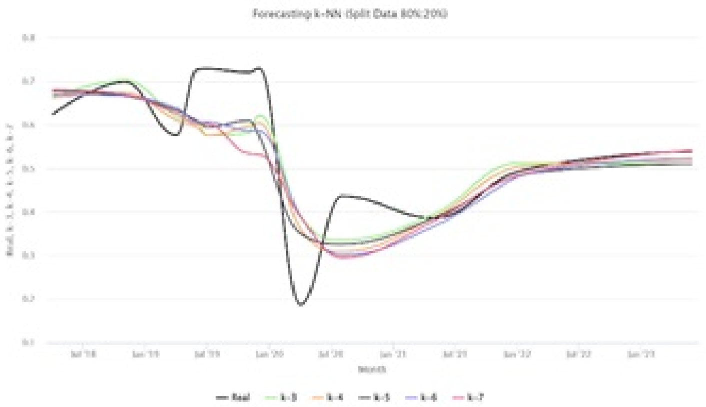

Figure 6: Occupancy rate prediction using k-NN (split of data 80% : 20%).

the black line shows real occupancy data, and the green line shows occupancy forecasting data using k-NN with k-value 5. There was a significant decrease in April 2020 with an occupancy rate of 0.19 in real data and an occupancy rate of 0.35 in forecasting data. The highest ARR was December 2019 with an occupancy rate of 0.73 in real data and an occupancy rate of 0.57 in forecasting data.

3.3 Neural network algorithm test analysis

At this stage, a comparison of the results of the test was carried out using a neural network algorithm at the split of data with a percentage ratio of 70:30 and 80:20. In testing, the neural network algorithms for one hidden layer and two hidden layers were used. The results of the neural network algorithm testing can be seen in Table 2.

From forecasting carried out using neural network at 70% : 30% split of data, it is clear that the best RSME is found for one hidden layer with an RMSE of 0.004. The

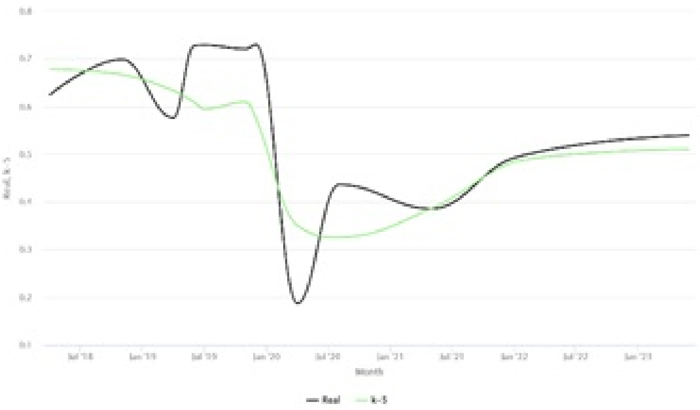

Figure 7: Occupancy rate prediction using k-NN with k-value 5 (split of data 80% : 20%).

Hidden Layer	Training Data	Testing Data	RMSE	Hidden Layer	Training Data	Testing Data	RMSE
1	70%	30%	0.004	1	80%	20%	0.005
2	70%	30%	0.007	2	80%	20%	0.015

Table 2: Test Results Using Neural Network.

comparison of the simulation with one hidden layer and two hidden layers at 70% : 30% data split can be seen in Figure 8.

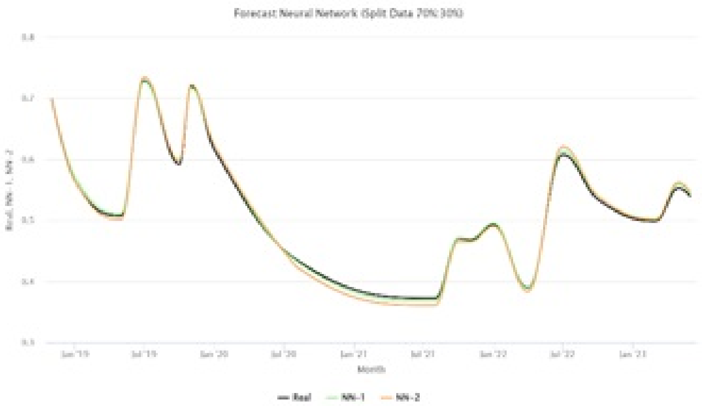

Figure 8: Occupancy rate prediction using neural network (split of data 70% : 30%).

Figure 9 shows the comparison of real occupancy rate data and occupancy rate prediction at 70% : 30% data split using the neural network algorithm with one hidden layer. In the graph, the black line shows real occupancy data, and the green line shows occupancy forecasting data using the neural network with one hidden layer. There was a significant decrease in August 2021 with an occupancy rate of 0.367 in real data and an occupancy rate of 0.372 in forecasting data. The highest ARR was in July 2020 with an occupancy rate of 0.729 in real data and an occupancy rate of 0.27 in forecasting data.

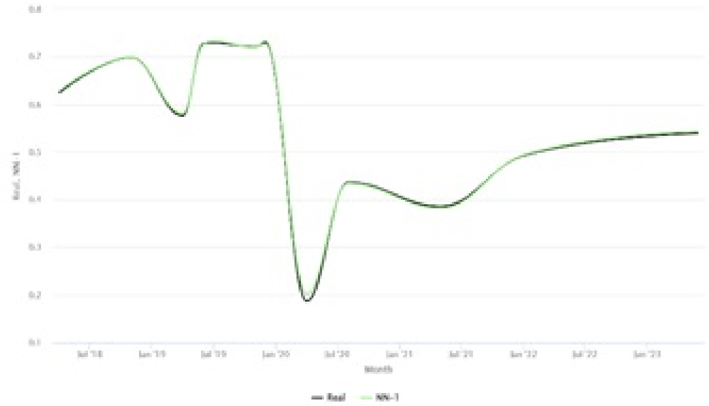

Figure 9: Occupancy rate prediction using neural network with one hidden layer (split of data 70% : 30%).

From forecasting carried out using neural network at 80% : 20% split of data, it is clear that the best RSME is found for one hidden layer with an RMSE of 0.005. The comparison of the simulation with one hidden layer and two hidden layers at 80% : 20% data split can be seen in Figure 10.

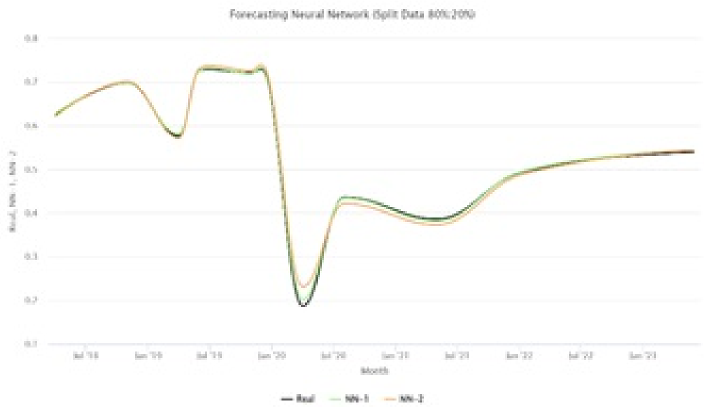

Figure 10: Occupancy rate prediction using neural network (split of data 80% : 20%).

Figure 11 shows the comparison of real occupancy rate data and occupancy rate prediction at 80% : 20% data split using the neural network algorithm with one hidden layer. In the graph, the black line shows real occupancy data, and the green line shows occupancy forecasting data using the neural network with one hidden layer. There was a significant decrease in August 2021 with an occupancy rate of 0.367 in real data and an occupancy rate of 0.372 in forecasting data. The highest ARR was July 2020 with an occupancy rate of 0.729 in real data and an occupancy rate of 0.27 in forecasting data.

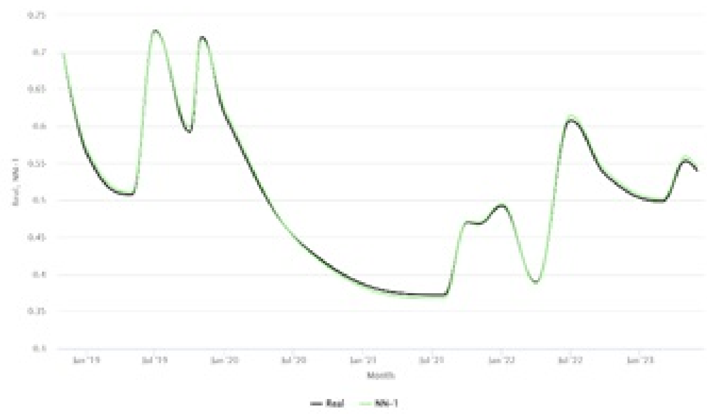

Figure 11: Occupancy rate prediction using neural network with one hidden layer (split of data 80% : 20%).

3.4 Algorithm testing comparison

At this stage, the most optimal RMSE results from each test with k-NN and neural network are compared. The results of the comparison of these algorithm tests are shown in Table 3.

Algorithm	Split data (%)	RSME	Remarks
k-NN	70:30	0.080	k-value = 3
	80:20	0.098	k-value = 5
Neural network	70:30	0.004	one hidden layer
	80:20	0.005	one hidden layer

Table 3: Comparison of Algorithm Testing.

From the comparison results, the most optimal RMSE result for k-NN at 70% : 30% data split with k-value 3 is an RMSE of 0.080, while for neural networks also at 70% : 30% data split with one hidden layer, an RMSE is 0.004. The results of comparison of occupancy rate forecasting using k-NN and neural network at 70% : 30% split of data can be seen in Figure 12. In the graph, the black line shows real occupancy rate data, the green line shows occupancy rate prediction data using the k-NN algorithm with k-value 3, and the orange line shows occupancy rate prediction data using the neural network algorithm with one hidden layer.

The comparison results for occupancy rate forecasting using k-NN and neural network resulted in the most optimal RMSE for the neural network algorithm with one hidden layer resulting in an RMSE of 0.004, a difference of 0.076 compared to using the k-NN algorithm. The evaluation of RMSE results for the k-NN and neural network algorithms reveals that the neural network, specifically with one hidden layer at 70% : 30% data

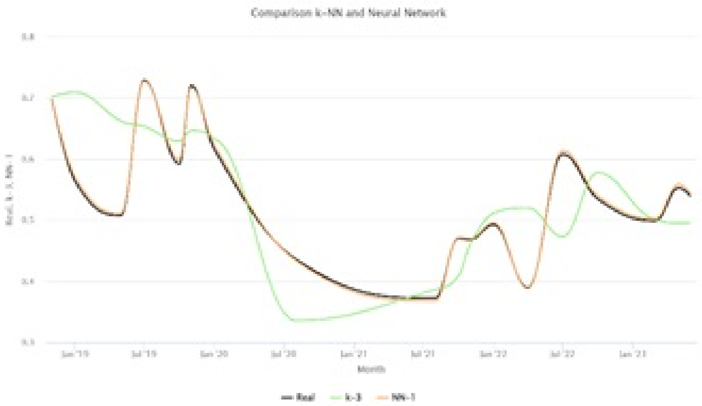

Figure 12: Comparison of occupancy rate predictions using k-NN and neural network (split of data 70% : 30%).

split, outperforms k-NN with a noticeable margin. The optimal RMSE of 0.004 for the neural network demonstrates its effectiveness in occupancy rate forecasting compared to the best-performing k-NN configuration, which achieved an RMSE of 0.080 at 70% : 30% data split and k-value of 3. These findings underscore the superiority of the neural network approach in this specific context.

4 Conclusion

Based on a comparison of tests using k-nearest neighbor and neural network to determine occupancy rate forecasting at Hotel XYZ with split of data 70% : 30% and 80% : 20%, the most optimal RMSE results were obtained using the neural network with one hidden layer at 70% : 30% split of data with an RMSE of 0.004. So, it is recommended for Hotel XYZ to use a neural network with one hidden layers for occupancy rate forecasting, which can later support better decisions in determining hotel policies in the future and goals set by the hotel. For other research, this study can be a reference to increase knowledge and conduct further investigations and develop other methods.

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] Anonim. *Star Hotel Room Occupancy Rate 2023.* November 2023.

[2] Anonim. *Non-Star Hotel Room Occupancy Rate in 2023.* Sep. 03, 2023. [Online]. Available: https://kemenparekraf.go.id/direktori-statistik/tingkat-penghunian-kamar-hotel-non-bintang-tahun-2023.

[3] B. Hermawan. *Hotel Operational Analysis*. NEM, Pekalongan, Indonesia, 2022.

[4] BPS-Statistics Indonesia. *Occupancy Rate of Hotel Room 2022*. Jakarta, May 2023.

[5] S. Sofiani and I. S. Djunaid. *Hotel Front Office Management*. Yogyakarta: Star Universe Media, 2023.

[6] F. Sabry. *K Nearest Neighbor Algorithm: Fundamentals and Applications*. One Billion Knowledgeable, 2022.

[7] M. Arhami and M. Nasir. *Data Mining – Algorithms and Implementation*. Penerbit Andi, 2020.

[8] J. Indriyanto. *K-Nearest Neighbor Algorithm for Insurance Customer Prediction*. NEM, 2021.

[9] S. Chakraverty and S. K. Jeswal. *Applied Artificial Neural Network Methods for Engineers and Scientists*. World Scientific, 2021.

[10] D. Novita et al. Comparison of K-Nearest Neighbor and Neural Network for Prediction International Visitor in East Java. *BAREKENG: Journal of Mathematics and Its Applications* **18** (3) (2024) 2057–2070.

[11] 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: *AIP Conference Proceedings*, American Institute of Physics Inc., Jan. 2023.

[12] A. Rahim, A. D. Rahajoe and M. Mahaputra. Prediction of the Number of Visitors Per Period to Beach Tourism Sites Using Triple Exponential Smoothing (Case Study of Gili Labak Beach, Sumenep). *JIFTI-Scientific Journal of Information Technology and Robotics* **3** (2) (2021) 39–43.

[13] T. Herlambang, V. Asy'ari, R. P. Rahayu, A. A. Firdaus and N. Juniarta. Comparison of Naïve Bayes and K-Nearest Neighbor Models For Identifying The Highest Prevalence Of Stunting Cases In East Java. *BAREKENG: Journal of Mathematics and Its Applications* **18** (4) (2024) 2153–2164.

[14] D. F. Karya, P. Katias, T. Herlambang and D. Rahmalia. Development of Unscented Kalman Filter Algorithm for stock price estimation. *J. Phys. Conf. Ser.* **1211** (1) (2019) 012031.

[15] I. Indasah, A. Y. P. Asih, T. Herlambang, P. Triwinanto and K. Oktafianto. Forecasting Air Pollution Levels Using Support Vector Regression and K-Nearest Neighbor Algorithm. *Nonlinear Dynamics and System Theory* **25** (2) (2025) 153–160.

[16] A. Muhith. 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.

[17] A. Muhith, I. H. Susanto, D. Rahmalia, D. Adzkiya and T. Herlambang. Profitability estimation of XYZ company using H-infinity and Ensemble Kalman Filter. In: *The Analysis of Demand and Supply of Blood in Hospital in Surabaya City Using Panel Data Regression*, 2022.

[18] D. Rahmalia et al. Comparison between Neural Network (NN) and Adaptive Neuro Fuzzy Inference System (ANFIS) on sunlight intensity prediction based on air temperature and humidity. *J. Phys. Conf. Ser.* **1538** (1) (2020) 012044.

[19] F. A. Susanto, et al. Estimation of Closed Hotels and Restaurants in Jakarta as Impact of Corona Virus Disease (Covid-19) Spread Using Backpropagation Neural Network. *Nonlinear Dynamics and System Theory* **22** (4) (2022) 457–467.

[20] F. S. Nugraha, M. J. Shidiq, and S. Rahayu. Analysis of Neural Network Classification Algorithm for Breast Cancer Disease Diagnosis). *Nusa Mandiri Pillar Journal* **15** (2) (2019) 149–156.

[21] D. B. Maghfira et al. Electronic Nose for Classifying Civet Coffee and Non-Civet Coffee. *Nonlinear Dynamics and System Theory* **23** (3) (2023) 323–337.

[22] P. Andrean. Application of K-NN Method to Predict Agricultural Results in Malang Regency. *Journal of Informatics Engineering Students* **3** (1) (2019) 235–242.

[23] V. Asy'ari, M. Y. Anshori, T. Herlambang, I. W. Farid, D. Fidita Karya, and M. Adinugroho. Forecasting average room rate using k-nearest neighbor at Hotel S. In: *International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA)*), IEEE, Nov. 2023, 496–500.

[24] V. Asy'Ari, et al. Forecasting average room rate using k-nearest neighbor at Hotel S. In: *International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA)*), 2023.

[25] U. Hidayah and A. Sifaunajah. *Easy Way to Understand K-Nearest Neighbor Algorithm Visual Basic 6.0 Case Study.* Jombang: LPPM KH. A. Wahab Hasbullah University, 2019.

[26] F. Gorunescu. *Data Mining: Concepts, Models and Techniques.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.