BAREKENG: Journal of Mathematics and Its Applications
September 2024 Volume 18 Issue 3 Page 2057–2070
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss3pp2057-2070

COMPARISON OF K-NEAREST NEIGHBOR AND NEURAL NETWORK FOR PREDICTION INTERNATIONAL VISITOR IN EAST JAVA

Dina Novita[1], Teguh Herlambang[2*], Vaizal Asy'ari[3], Arasy Alimudin[4], Hamzah Arof[5]

[1]*Department of Management, Universitas Muhammadiyah Surabaya*
Jln. Raya Sutorejo No 59, Mulyorejo, Surabaya, 60113, Indonesia

[2]*Information System Department, FEBTD, Universitas Nahdlatul Ulama Surabaya*
Jln. Raya Jemursari 51-57, Surabaya, 60237, Indonesia

[3]*Postgraduate of Department of Information Technology, Universitas Pembangunan Nasional Veteran*
Jln. Rungkut Madya No. 1, Gunung Anyar, Surabaya, 60294, Indonesia

[4]*Department of Management, Universitas Narotama Surabaya*
Jln. Arief Rahman hakim No 51, Sukolilo, Surabaya, 60117, Indonesia

[5] *Department of Electrical Engineering, University of Malaya, Malaysia*
Malaysia, 50603, Kuala Lumpur

*Corresponding author's e-mail: * **teguh@unusa.ac.id***

Article History:

Received: 31st May 2024
Revised: 25th June 2024
Accepted: 15th July 2024
Published: 1st September 2024

Keywords:

Forecasting;
Tourism;
International Visitor;
K-NN;
Neural Network.

ABSTRACT

Tourism is one of the government's priority sectors for economic growth. East Java is one of Indonesia's provinces and is attractive to international visitors. International visitors will appreciate the natural beauty and multiculturalism offered by East Java. In this study, predictions of international visitor visits in East Java from the entrance of Juanda International Airport were carried out using k-NN (k-Nearest Neighbor) and a neural network. The dataset used is based on BPS statistics of Jawa Timur Province in the form of the number of international visitor arrivals from January 2000 to February 2024. The datasets were distributed by dividing the data into 70% for training data and 30% for testing data. The creation of the k-NN model is carried out using k-values 2 to 7. The creation of a modern neural network using hidden layers 1 to 3. The prediction results that were made using k-NN obtained optimal RMSE at k-values 2, resulting in an RMSE of 1594,674 or an error of 3,98%. Meanwhile, the prediction results that have been made using neural networks obtained optimal RMSE at two hidden layers, which resulted in an RMSE of 1873, 355 or an error of 4,68%. So, it is recommended that the k-NN algorithm be used to predict the number of international visitors in East Java. The results of this study can be used to provide quantitative information for the government and stakeholders in adjusting the program to the development of international visitors visiting East Java.

How to cite this article:

First D. Novita, T. Herlambang, V. Asy'ari, A. Alimudin and H. Arof., " COMPARISON OF K-NEAREST NEIGBOR AND NEURAL NETWORK FOR PREDICTION INTERNATIONAL VISITOR IN EAST JAVA," *Barekeng: J. Math. &; App.*, vol. 18, iss. 3, pp. 2057-2070, September, 2024.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article • Open Access

1. INTRODUCTION

Tourism is one of the priority sectors the government has identified that can contribute to economic growth, job creation, community development, and regional development [1]. So, the development of the tourism sector must continue to be improved. Increasing the added value of tourism is also one of the government's priorities to strengthen economic resilience for quality growth [2]. The progress of the tourism sector is also used as a measuring tool in assessing a region's financial stability and security. If an area receives relatively many international visitors, then in general, the area can have relatively good economic stability and security.

An international visitor according to the United Nations World Tourism Organization (UNWTO) and the International Union of Office Travel Organization (IUOTO) is any person who visits a country outside his country of residence, less than one year (12 months), driven by one or several needs without the intention of earning income in the place visited [1], [3]. International visitors indirectly contribute to growth in various sectors of the economy, with links ranging from the creation of added value to the labor needed. The increase in international visitors impacts the increasing demand for accommodation provision [4]. There needs to be an increase in the provision of hotel rooms and other accommodations so as to avoid causing a gap between demand and supply for these accommodations.

East Java is one of Indonesia's provinces and is attractive to international visitors. International visitors will appreciate the natural beauty and multiculturalism offered by East Java. In East Java, there is a tourist destination Bromo-Tengger-Semeru as one of the National Tourism Strategic Areas (KSPN), which is a government program through Presidential Regulation Number 3 of 2016 concerning the acceleration of the implementation of national strategic projects, which has a strategic nature to increase growth and equitable development to improve community welfare and regional development [5]. In the Annex to the Presidential Regulation of the Republic of Indonesia Number 18 of 2020 concerning the 2020-2024 National Medium-Term Development Plan (RPJMN) [6], Bromo-Tengger-Semeru is one of the ten priority tourism destinations, where the benefits of significant projects for the ten tourism destinations of the 2020-2024 RPJMN project are to increase international visitors by 22.3 million arrivals (2024). The policies carried out by the government to support the tourism sector directly increase international visitor visits to East Java. Based on BPS-Statistics of East Java Province, international visitor visits in February 2024 increased by 135.78 percent compared to the number of international visitors in February 2023, which amounted to 11,882 visits; this also had an impact on the occupancy rate of star hotels in East Java in February 2024 reaching an average of 50.53 percent or an increase of 5.51 points compared to the previous month, The occupancy rate of non-star classification hotel rooms in February 2024 reached an average of 24.15 percent, an increase of 0.76 points compared to the last month [7].

With the increasing number of international visitors in East Java, the government and stakeholders must implement proper planning and decision-making strategies to determine policy direction. One of the important things in planning is prediction. Prediction is a forecast of an event that will occur in the future [8], [9]. We will benefit significantly based on adequate quantitative data for predictions about international visitors in East Java, especially when adjusting existing programs to the development of international visitors visiting East Java. Based on this, this study will conduct forecasting of international visitors in East Java using k-Nearest Neighbor (k-NN) and neural network. The RMSE results from each method are then compared to see which RMSE works best.

The k-nearest neighbor (k-NN) algorithm can be applied to regression and classification. The k-closest training sample of a data set is the input in both cases [10]. K-NN is referred to as case-based reasoning, a methodology based on the reasoning of cases based on training data of a case stored, trained, and accessed to solve new problems [11]. K-NN makes predictions based on comparisons of the nearest neighbors obtained from past data that has been tested. The close or far of neighbors is usually calculated based on Euclidian distance. The best k-value for the k-NN algorithm depends on the data used; in some data, a high k-value will usually reduce the noise effect [12].

Since 1943, when Warren McCulloch and Walter Pitts presented the first neural network model computations, neural networks have been around. This model explains the binary information processing capabilities of artificial neurons. Frank Rosenblatt conducted more studies and discovered a two-layer network known as a perceptron in 1950. Since neural networks have flexible function structures, they are rapidly evolving and have been used extensively in various fields. In situations where standard methods prove difficult or regularly fail, neural networks can be utilized to identify solutions [13].

Previous research used an adaptive neural fuzzy inference system to estimate the number of shuttered hotels and restaurants in Jakarta as a result of the corona virus sickness [14], forecasting using the fuzzy method to prediction the number of Demam Berdarah Dengue (DBD) patients[15], prediction of the number of visitors per period to beach attractions using triple exponential smoothing [16], classifying price range of smartphone in market using backpropagation and Learning Vector Quantification (LVQ) [17], estimation of stock price using Unscented Kalman Filter (UKF) [18], forecasting using linear Support Vector Machine (SVM) of occupied hotel rooms [19], H-Infinity and Ensemble Kalman Filter (EnKF) algorithm to profitability estimation [20], panel data regression analysis of blood supply and demand in a Surabaya City hospital [21], prediction of sunlight intensity using neural network and Adaptive Neuro Fuzzy Inference System (ANFIS) [22], using a backpropagation neural network to estimate the number of closed hotels and restaurants as a result of the Covid-19 wave [23], breast cancer diagnosis using neural network [24], SVM, k-NN, and decision tree algorithm for classifying civet coffee [25], k-NN algorithm to forecasting agricultural products in Malang, East Java [26], k-NN algorithm to forecasting average room rate in hotel [27], utilizing neural networks to forecast occupancy rate [28]. In this study, a comparison of K-NN and Neural Networks will be carried out to predict international visitors in East Java. The authors chose to compare the K-NN and Neural Network methods because both methods have high accuracy and provide quantitative information for the government and stakeholders in adjusting the program to the development of international visitors visiting East Java.

2. RESEARCH METHODS

2.1 Exploratory Data Analysis

In this study, the dataset used is the number of international visitor arrivals in East Java through the Juanda entrance based on BPS-Statistics of Jawa Timur Province from January 2000 to February 2024. The data obtained is univariate data with a total of 290 records, so it is necessary to do an Exploratory Data Analysis (EDA). EDA is a data exploration method using simple arithmetic and graphical techniques to summarize observational data [29], [30]. Then, this univariate data will be processed into multivariate data and changed from period 1 to period 10. The pattern model of changing univariate data into multivariate data can be seen in Table 1 [9], [31].

Table 1. Univariate to multivariate data pattern

Pattern	Input lag	Output/Target
1	$x_1, x_2, x_3, x_4, \ldots, x_p$	x_{p+1}
2	$x_2, x_3, x_4, x_5, \ldots, x_p$	x_{p+2}
3	$x_3, x_4, x_5, x_6, \ldots, x_p$	x_{p+3}
…	…	…
x_{m-p}	$x_{m-p}, x_{m-p+1}, x_{m-p+2}, \ldots, x_{m-1}$	x_m

2.2 K-Nearest Neigbor

K-nearest neighbor (k-NN) is a method that makes forecasting using classification techniques on test data strictly based on the comparison of the K-value of the nearest neighbor [12]. The closest neighbor is the trained object with the greatest similarity value or the least similarity value based on past data. The value of k expresses the nearest number of ladders. K-NN works on the principle that each data point adjacent to each other will be in the same class. In general, a high k value will reduce the effect of noise on the classification. However, it will blur the boundaries between each classification [32].

The k-NN algorithm is used to identify the nearest neighbor of a given query point so that a class label can be assigned to that point. In k-NN, the classification of new objects is based on attributes and training samples. The prediction result will determine which data point is closest to a particular query point, rarely

between the query point and other data that needs to be calculated. The distance calculation used is Euclidean Distance with the **Equations (1) and (2).**

$$d_i = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \quad (1)$$

$$d_i = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2} \quad (2)$$

p_i-q_i : Euclidean vectors, starting from the origin of the space (initial poin)
d_i : distance
n : n-space

K-NN works based on the principle of finding the closest distance, that each point adjacent to each other will be in the same class. In other words, K-NN classifies new data points based on similarity. The stages of the K-NN algorithm work process include **[32]**, **[33]**:

1. Defines the k parameter, which stands for the number of nearest neighbors.
2. Using the provided sample data as a guide, compute the square of each object's Euclid jar (quei instance).
3. Then, the items will be categorized according to the shortest Euclidean distance.
4. Ascertain which classification category the closest neighbor belongs to.
5. Predicting the value of the calculated query instance is possible by utilizing the nearest neighbor category with the highest majority.

2.3 Neural Network

Neural network fundamentals include inputs, weights, processing units, and outputs. To classify patterns, transfer patterns from inputs into new patterns in outputs, store patterns for later recall, map comparable patterns, optimize issues, and predict, neural networks can be used. Preparing data for training and learning, identifying neural network design, training and learning procedures, and testing procedures are the first steps in creating neural networks **[34]**. Neural networks can alter their structural composition in response to information that enters and leaves the network. Data patterns can be found using neural networks to represent the link between input and output. A neural network's fundamental processing unit is its neuronal architecture. A neuron's basic form can be seen in **Figure 1**.

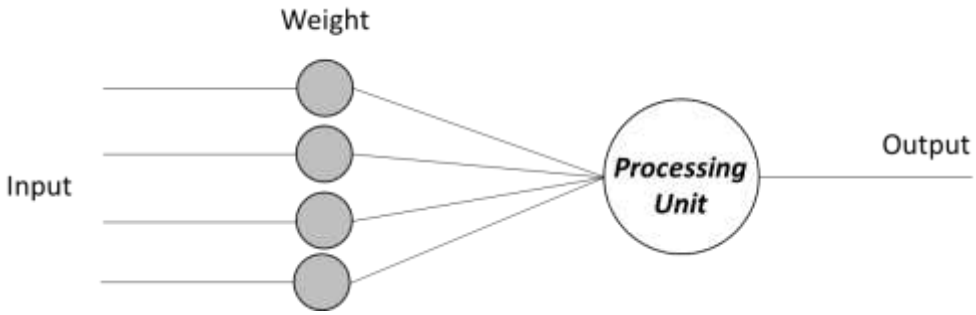

Figure 1. Design neural network model

Weights that join the various network components are contained in the weight vector (w). In the language of neural networks, the letter "w" denotes the expression of a connection between two neurons or the amount of information that is sent from one neuron to another inside the neural network. The procedure of adding inputs $x_1, x_2, \ldots, x_n$ and multiplying the result by its weight $w_1, w_2, \ldots, w_n$ represented in the equation constitutes the first stage (3):

$$net = (w_1x_1 + w_2x_2 + w_3x_3 + \ldots + w_nx_n) \quad (3)$$

Perceptron model of input, weight, and bias so that the output can be written with vector notations as follows:

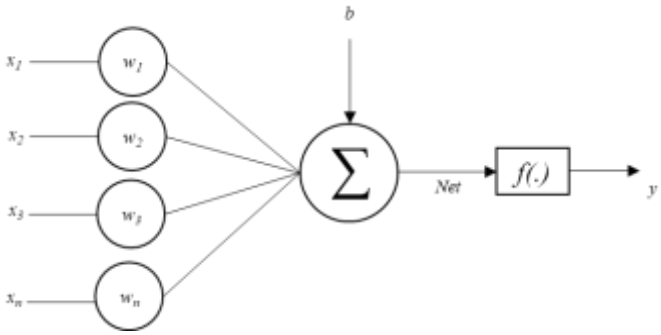

Figure 2. Perceptron model

Referred to as a separate neuron model parameter, a bias is a threshold value of b that is significant for certain neuron models. Different input conditions and output influences are needed to incorporate a nonlinear activation function f(.) in the neuron design [35]. The objective is to prevent the output from exceeding limits that are not appropriate and to reach a sufficient level if the input signal is modest. Like the perceptron model in Figure 2 the output of a neuron can be written in an Equation (4):

$$y = f(net) \tag{4}$$

3. RESULTS AND DISCUSSION

This study uses a dataset of international visitor arrivals in East Java through the Juanda International Airport entrance from January 2000 to February 2024 (290 months). The dataset is divided into 70% for training data and 30% for test data. Then, a test analysis was carried out with the k-nearest neighbor algorithm with k-values 2 to 7 and a neural network with 1 to 3 hidden layers. Each of these forecasts will be compared for the most optimal RMSE.

3.1 Exploratory Data Analysis

The dataset used in this study is the number of international visitor arrivals based on statistics of East Java Province. Datasets that are still in univariate form will be processed into multivariate. Univariate data determined variable x_t in the period to be created. The variable x_t is the dependent variable, and the variables x_{t-1} to x_{t-10} are the dependent variables. The following is univariate data obtained from the number of international visitor arrivals in East Java from January 2000 to February 2024, which can be seen in Table 2.

Table 2. Univariate data on the number of international visitors in East Java

Year	Moon	Sum
2000	January	7879
2000	February	7268
2000	March	8532
2000	April	7376
2000	May	7582
2000	June	8696
2000	July	9005
2000	August	8947
2000	September	7854

Year	Moon	Sum
2000	October	8949
2000	November	10926
2000	December	12357
2001	January	7488
2001	February	7051
...	...	...
2023	September	25593

Next, univariate data will be converted to multivariate data from 1 period to 10 periods; the result is shown in **Table 3** below.

Table 3. Multivariate data training at input layer 1 to 10 periods

x_t	x_{t-1}	x_{t-2}	x_{t-3}	...	x_{t-10}
7879	7268	8532	7376	...	10926
7268	8532	7376	7582	...	12357
8532	7376	7582	8696	...	7488
7376	7582	8696	9005	...	7051
582	8696	9005	8947	...	8904
8696	9005	8947	7854	...	8189
9005	8947	7854	8949	...	8717
8947	7854	8949	10926	...	10161
7854	8949	10926	12357	...	10177
8949	10926	12357	7488	...	10380
10926	12357	7488	7051	...	9186
12357	7488	7051	8904	...	9753
7488	7051	8904	8189	...	11521
...	...	...	...	...	...
12361	15734	18025	20303	...	28015

3.2 K-Nearest Neighbor Algorithm Test Analysis

The dataset of the number of international visitors has been converted into multivariate data; then, the dataset is divided into 70% for training data and 30% for test data. Furthermore, using the k-nearest neighbor algorithm with k-values 2 to 7, an application model was carried out to determine the performance of each k-value. The RMSE results from using the k-NN algorithm with k-values 2 to 7 are shown in **Table 4**.

Table 4. Test Results Using k-NN

K-Values	RMSE	Error
2	1594,674	3,98%
3	1622,701	4,05%
4	1688,320	4,22%
5	1636,585	4,09%
6	1713,065	4,28%
7	1795,068	4,48%

From the forecasting of international visitors using k-NN with k-values 2 to 7, RMSE is best found in k-values 2 with an RMSE of 1,594,674. Meanwhile, the highest RMSE is found in k-values 7 with an RMSE of 1,795,068, a difference of 200,394 compared to RMSE in k-values 2. An accurate comparison of international visitor data with prediction results using the K-NN algorithm can be seen in **Figure 3**.

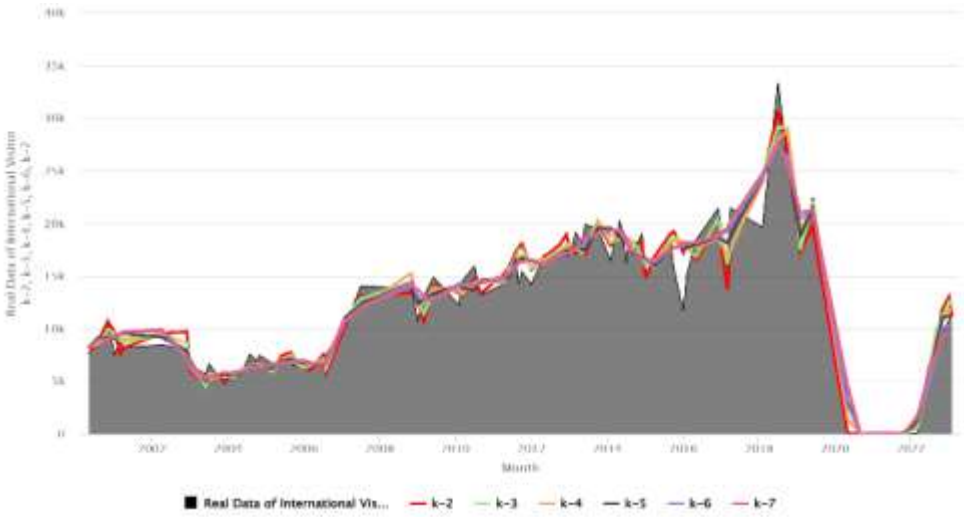

Figure 3. Prediction of international visitors using k-NN

The graph above shows the actual data of international visitors in the black plot type area. At the same time, the results of international visitor predictions using the k-NN algorithm are shown in a line plot type that displays each k-value from 2 to 7. The highest international visitor data was found in July 2018, with 33,390 international visitors on accurate data and 30,691 international visitors on prediction data with k-values 2 (a difference of 2,699 compared to actual data), and international visitors with 27,685 on prediction data with k-values 7 (a difference of 5,705 compared to real data). A detailed graph comparing the highest international visitor data can be seen in **Figure 4** which shows the prediction results from May 2018 to September 2018 for algorithm k-NN. The highest international visitor data in July 2018 for the data closest to accurate data equals the optimal RMSE result at k-values 2, with accurate international visitor data of 33,390 and prediction data of 30,691. However, the prediction data that is furthest from the actual data, in contrast to the highest RMSE result at k-value 7, for the prediction data with the farthest distance is found at k-value 5 with a prediction of international visitors of 27,389 (a difference of 6,001 compared to the actual data).

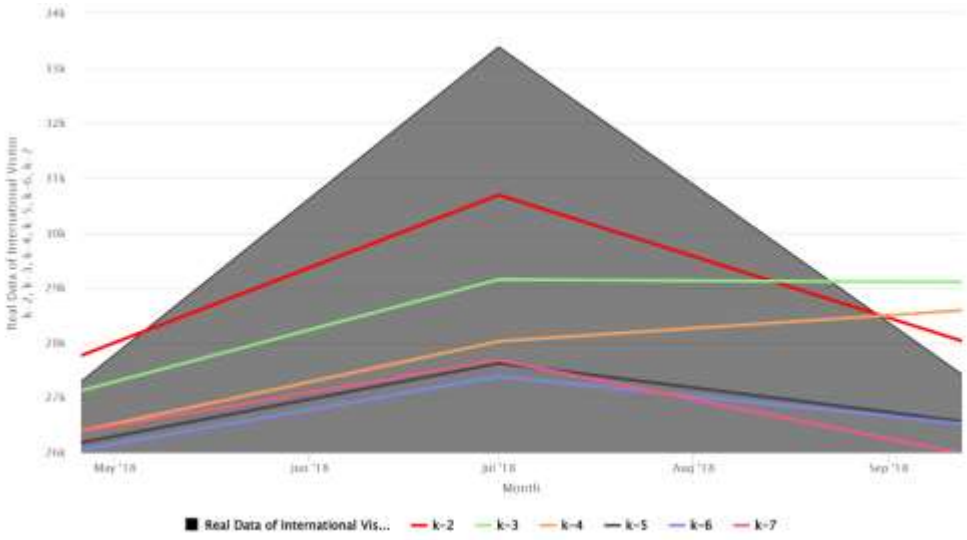

Figure 4. Prediction of international visitors using k-NN from May 2018 to September 2018

The comparison of the lowest data on international visitors in East Java was in November 2021, with accurate data at 0 for international visitors; at that time, the COVID-19 virus was soaring. Details of the comparison of the lowest data for international visitors can be seen in **Figure 5** which shows the prediction results from September 2021 to December 2021. The lowest international visitor data in November 2021 for the data closest to actual data is at k-value 5 with a prediction data of 10, in contrast to the optimal RMSE at k-value 2. Meanwhile, the prediction data farthest from the actual data is the same as the highest RMSE result at k-values 7, with prediction data in November 2021 amounting to 195.

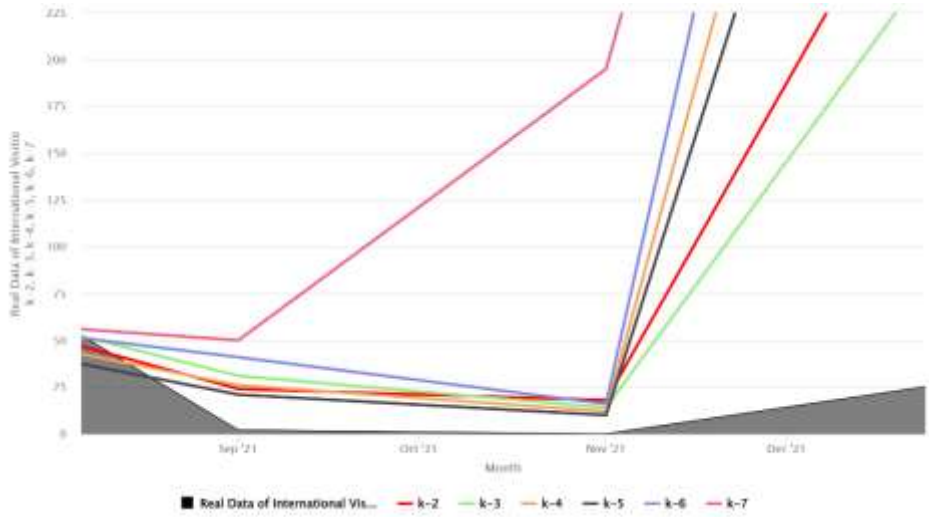

Figure 5. Prediction of international visitors using k-NN from September 2021 to December 2021

RMSE is best found in k-values 2 with an RMSE of 1,594,674. An actual comparison of international visitor data with prediction results using k-NN with k-values two can be seen in **Figure 6**. The graph compares the actual data of international visitors and forecasts with the line; the black line shows the exact data of international visitors, and the red line shows the predicted data of international visitors. Accurate data of international visitors in May 2020 experienced a significant decrease with 12 international visitors until November 2021, international visitors showed 0, which was the impact of the Covid-19 pandemic, while prediction data using k-NN with k-values 2 in May 2020 showed 58 international visitors and in November 2021 showed 18 international visitors. The development of international visitors began to increase again in July 2022, with actual data of international visitors of 6,087 and international visitors of 6,524 in the predicted data with k-values 2 (difference of 437).

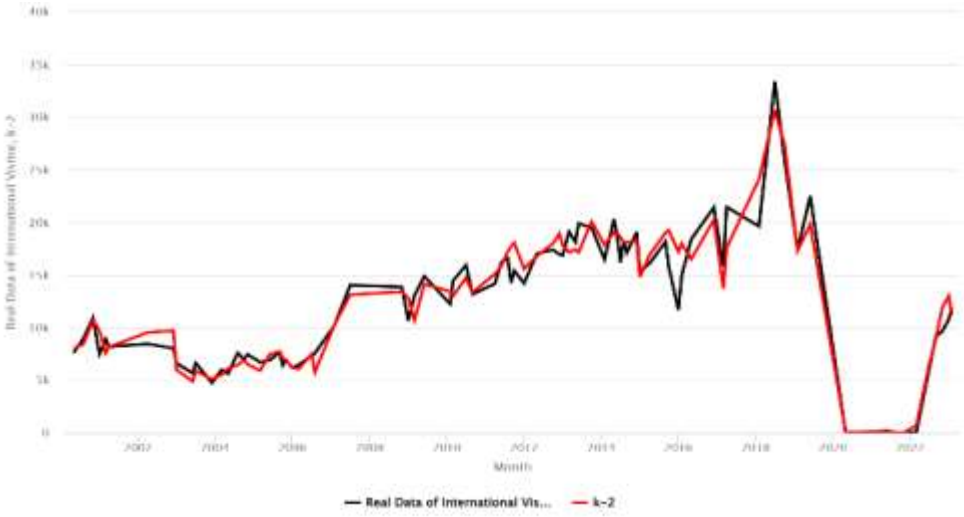

Figure 6. Prediction of international visitors using k-NN

3.3 Neural Network Algorithm Test Analysis

The dataset of the number of international visitors has been converted into multivariate data, then divided into training data and testing data using a ratio of 70:30. Furthermore, using a neural network algorithm with hidden layers as many as 1 to 3, apply a model to determine the performance of each hidden layer. The RMSE results from using the neural network algorithm with hidden layers 1 to 3 are shown in **Table 5**.

Table 5. Test results using neural network

Hidden Layer	RMSE	Error
1	1985.373	4.96%
2	1873.355	4.68%
3	1933.628	4.83%

From the forecasting that has been carried out to generate RMSE using a neural network algorithm with hidden layers 1 to 3, the best RMSE result on two hidden layers is obtained at 1,873,355. Meanwhile, the highest RMSE is found in the three hidden layers with an RMSE of 1,985,373, a difference of 112,018 compared to the RMSE in the two hidden layers. The model of the data that has been predicted using a neural network with each layer size of 8 is shown in **Figure 7** i.e., figure (a) for testing using one hidden layer, figure (b) for testing using two hidden layers, figure (c) for testing using three hidden layers.

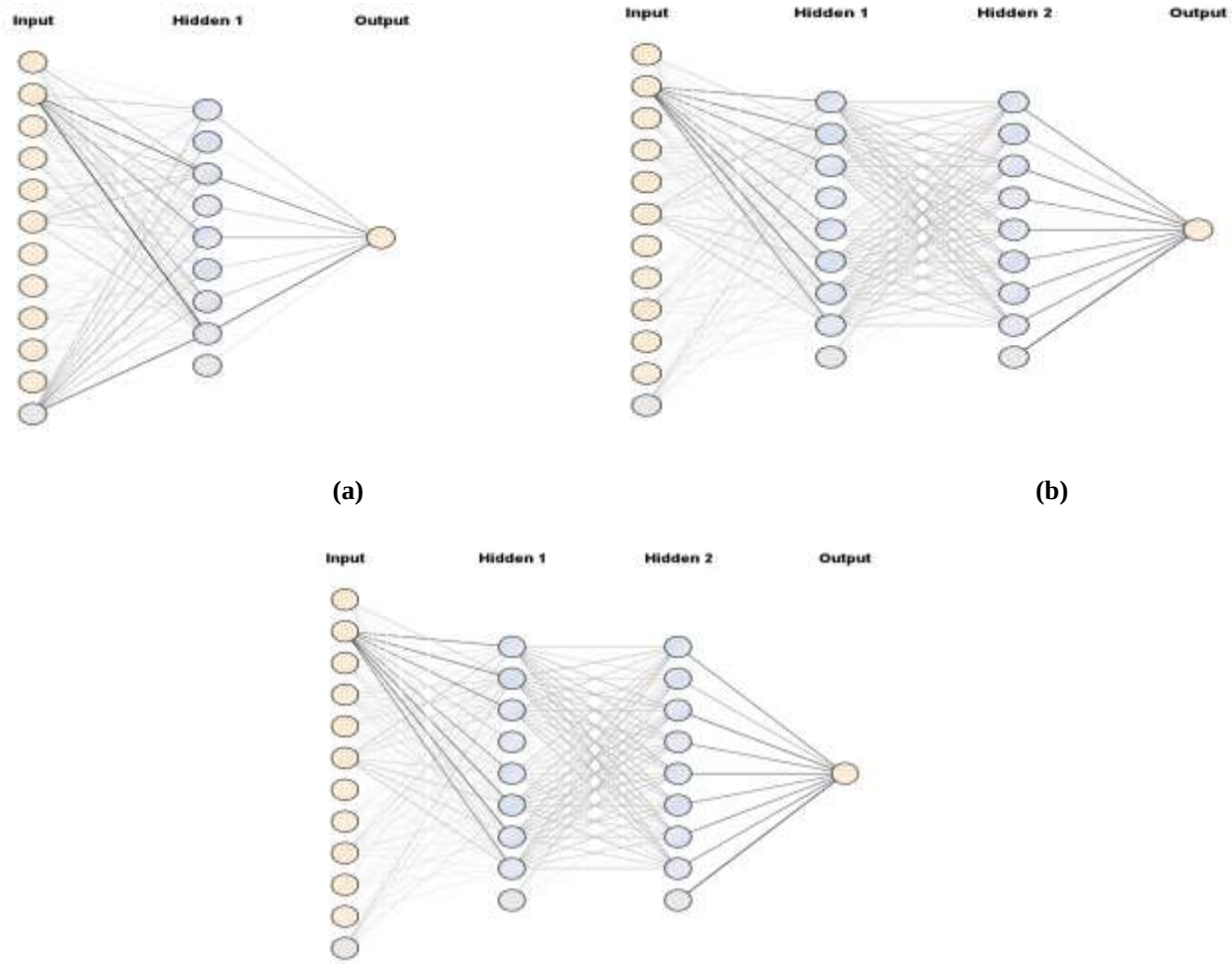

Figure 7. Neural Network model, (a) One hidden layer, (b) Two hidden layer, (c) Three hidden layer

Accurate data comparison of international visitors with prediction results using neural network algorithms with hidden layers as many as 1 to 3 can be seen in **Figure 8**. The actual data of international visitors is shown in the black plot type area. At the same time, the results of international visitor predictions using neural network algorithms are shown in line-type plots that display each hidden layer from 1 to 3. The highest international visitor data was found in July 2018, with 33,390 international visitors on actual data and 29,306 international visitors on prediction data with two hidden layers (a difference of 4,084 compared to actual data), and international visitors with 31,431 on prediction data with one hidden layer (a difference of 1,959 compared to actual data).

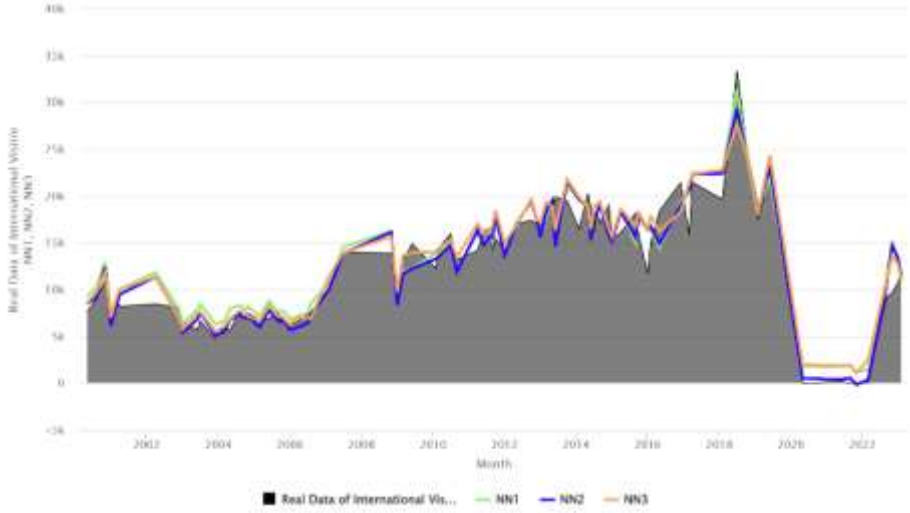

Figure 8. Predict international visitors using neural networks

Detailed graphs for comparing the highest data of international visitors can be seen in **Figure 9**, which shows the prediction results for the neural network algorithm from May 2018 to September 2018. The highest international visitor data in July 2018 for the data closest to actual data is in one hidden layer, with actual international visitor data of 33,390 and predicted data of 31,431, in contrast to the optimal RMSE results, which are in two hidden layers. For the prediction data that is the farthest distance from the actual data, it is also different from the highest RMSE result in one hidden layer, for the prediction data with the farthest distance is found in the three hidden layers with a prediction of international visitors of 27,750 (a difference of 5,640 compared to actual data).

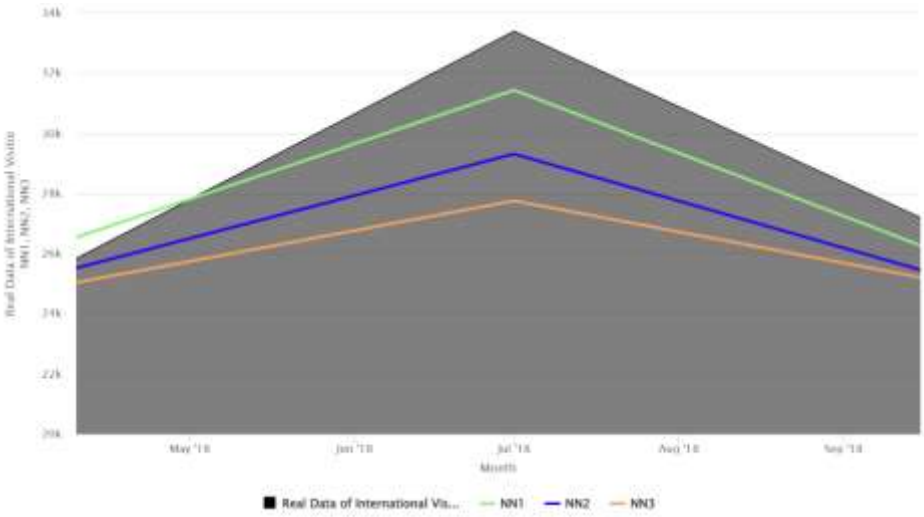

Figure 9. Prediction of international visitors using neural network from May 2018 to September 2018

The lowest comparison of international visitor data in East Java for artificial neural network algorithms with real data at 0 for international visitors was in November 2021, when the Covid-19 virus surged. Details of the comparison of the lowest international visitor data can be seen in **Figure 10** which shows the prediction results from September 2021 to December 2021. In the lowest international visitor data in November 2021, the data closest to actual data is in the three hidden layers with prediction data of -174, in contrast to the

optimal RMSE in the two hidden layers. Meanwhile, the prediction data furthest from the actual data is the same as the highest RMSE result in one hidden layer, with the November 2021 prediction data of 1205.

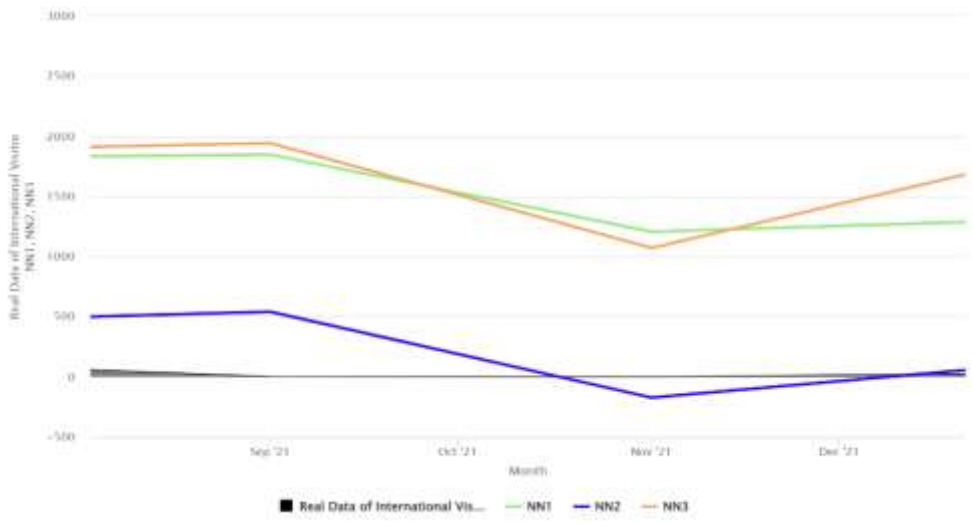

Figure 10. Prediction of international visitors using Neural Network from September 2021 to December 2021

RMSE is best found in two hidden layers with an RMSE of 1,873,355. A comparison of actual data of international visitors with prediction results using neural networks with two hidden layers can be seen in **Figure 11**. A comparison of actual data of international visitors and forecast is shown with the line; the black line shows actual data of international visitors, and the blue line shows prediction data of international visitors. Exact data of international visitors in May 2020 experienced a significant decrease with 12 international visitors, until November 2021, international visitors showed 0, which was the impact of the Covid-19 pandemic, while prediction data using neural networks with two hidden layers in May 2020 showed 258 international visitors and in November 2021 showed 174 international visitors. The development of international visitors began to increase again in July 2022, with actual data of international visitors of 6,087 and international visitors of 6,630 in the predicted data with two hidden layers (difference of 543).

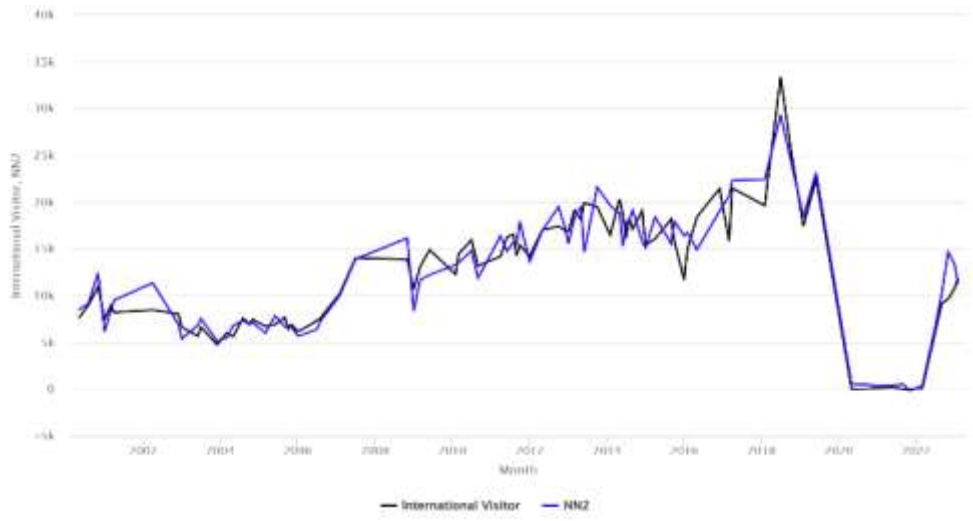

Figure 11. Predict international visitors using Neural Networks with Two Hidden Layers

3.4 Algorithm Testing Comparison

A comparison of the optimal RMSE results from testing international visitor data in East Java for each algorithm is shown in **Table 6**.

Table 6. Comparison of algorithm testing

Algorithm	RMSE	Remarks
k-NN	1594.674	k-values = 2
Neural Netwok	1873.355	Two hidden layers

The results of the RMSE comparison for the k-NN algorithm with k-value 2 produced a value of 1,594,674. In contrast, the neural network with two hidden layers produced a value of 1,873,355, a difference of 278,681. If we can see in **Table 4** and **Table 5**, it shows that the simulation results with KNN with K-value 2 produce the smallest error. Comparison of actual data of international visitors, prediction of international visitor split data 70%:30% using k-NN algorithm with k-values two and neural network algorithm with two hidden layers shown in **Figure 12**. In the graph, a comparison of actual data of international visitors and forecast is shown with the line: the black line shows actual data of international visitors, the red line shows prediction data using K-NN with k-values 2, and the blue line shows prediction data using neural networks with two hidden layers.

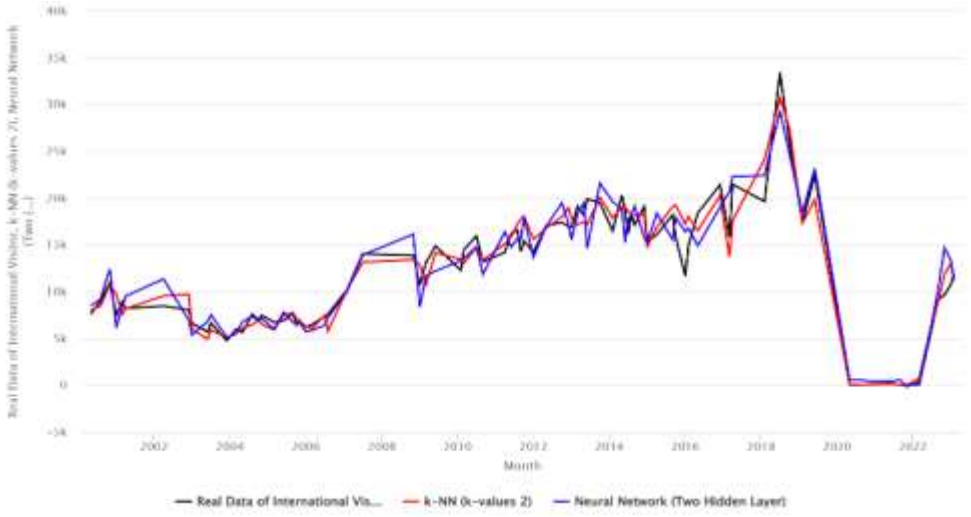

Figure 12. Comparison of international visitor predictions using k-NN with k-values 7 and neural networks with two hidden layers

4. CONCLUSIONS

Based on a comparison of tests using the k-NN algorithm and neural network to determine the forecasting of the number of international visitor arrivals in East Java through the Juanda entrance with a data split of 70%:30%, the most optimal RMSE is produced using the k-NN algorithm with a value of k 7 so that it makes an RMSE of 1594,674 or error of 3,98%. So, it is recommended that the K-NN algorithm be used to predict the number of international visitors in East Java; the results of this study can provide quantitative information for the government and stakeholders in adjusting the program to the development of international visitors visiting East Java. This research can be a reference to increase knowledge and further research and develop with other methods.

REFERENCES

[1] Badan Pusat Statistik Provinsi Jawa Timur, *Statistik Pariwisata Provinsi Jawa Timur 2022*. BPS Provinsi Jawa Timur, 2023.

[2] A. Mun'im, "Improvement on the Measurement of Tourism Contribution: An Alternative to Accelerating Indonesia's Economic Growth," *Jurnal Keprawisataan Indonesia*, vol. 16, no. 1, 2022.
[3] BPS-Statistics of Jawa Timur Province, *Jawa Timur Province in Figures 2023*. BPS-Statistics of Jawa Timur Province, 2023.
[4] BPS Provinsi Jawa Timur, *Statistik Daerah Provinsi Jawa Timur 2023*. BPS Provinsi Jawa Timur, 2023.
[5] T. A. Rihaksa and H. Susanti, "Penyuluhan Pentingnya Peran Infrastruktur Dalam Permintaan Pariwisata Internasional Indonesia," *Jurnal Abdimas Bina Bangsa*, vol. 4, no. 1, pp. 731–744, 2023.
[6] Republik Indonesia, *Lampiran Peraturan Presiden Republik Indonesia Nomor 18 Tahun 2020 Tentang Rencana Pembangunan Jangka Menengah Nasional 2020-2024*. 2020.
[7] Badan Pusat Statistik Provinsi Jawa Timur, "Perkembangan Pariwisata Provinsi Jawa Timur Februari 2024," Apr. 2024.
[8] R. A. Pratama, Suhud, and A. D. Jubaedi, "Rancang Bangun Sistem Ramalan Penjualan Menggunakan Metode Regresi Linear," *ProTekInfo(Pengembangan Riset dan Observasi Teknik Informatika)*, vol. 7, pp. 11–16, Sep. 2020, doi: 10.30656/protekinfo.v8i1.5017.
[9] H. Nisyak, N. Fithriyah, and Fatimatuzzahra, "Optimasi Neural Network Menggunakan Algoritma Genetika Untuk Memprediksi Jumlah Wisatawan Berdasarkan Hunian Hotel," *JoMI: Journal of Millennial Informatics*, vol. 2, no. 1, pp. 23–30, Jan. 2024.
[10] F. Sabry, *K Nearest Neighbor Algorithm: Fundamentals and Applications*. One Billion Knowledgeable, 2022.
[11] M. Arhami and M. Nasir, *Data Mining - Algoritma dan Implementasi*. Penerbit Andi, 2020.
[12] J. Indriyanto, *Algoritma K-Nearest Neighbor untuk Prediksi Nasabah Asuransi*. NEM, 2021.
[13] S. Chakraverty and S. K. Jeswal, *Applied Artificial Neural Network Methods for Engineers and Scientists*. WORLD SCIENTIFIC, 2021.
[14] M. Y. Anshori *et al.*, "Estimation of closed hotels and restaurants in Jakarta as impact of corona virus disease spread using adaptive neuro fuzzy inference system," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 11, no. 2, pp. 462–472, Jun. 2022.
[15] 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.
[16] A. Rahim, A. D. Rahajoe, and M. Mahaputra, "Prediksi Jumlah Pengunjung Perperiode Terhadap Tempat Wisata Pantai Menggunakan Triple Exponential Smoothing (Studi Kasus Pantai Gili Labak Sumenep)," *JIFTI-Jurnal Ilmiah Teknologi Informasi dan Robotika*, vol. 3, no. 2, pp. 39–43, Dec. 2021.
[17] M. Y. Anshori, D. Rahmalia, and T. Herlambang, "Comparison Backpropagation (BP) and Learning Vector Quantification (LVQ) on classifying price range of smartphone in market," *J Phys Conf Ser*, vol. 1836, no. 1, p. 012040, Mar. 2021.
[18] D. F. Karya, P. Katias, T. Herlambang, and D. Rahmalia, "Development of Unscented Kalman Filter Algorithm for stock price estimation," *J Phys Conf Ser*, vol. 1211, p. 012031, Apr. 2019.
[19] 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," 2023.
[20] 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)*, IOP Publishing Ltd, Jan. 2021.
[21] 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 Systems Theory*, vol. 22, no. 5, pp. 550–560, 2022.
[22] 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*, vol. 1538, no. 1, p. 012044, May 2020.
[23] 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 Systems Theory*, vol. 22, no. 4, pp. 457–467, 2022.
[24] F. S. Nugraha, M. J. Shidiq, and S. Rahayu, "Analisis Algoritma Klasifikasi Neural Network Untuk Diagnosis Penyakit Kanker Payudara," *Jurnal Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 149–156, Aug. 2019.
[25] D. B. Magfira *et al.*, "Electronic Nose for Classifying Civet Coffee and Non-Civet Coffee," 2023.
[26] P. Andrean, "Penerapan Metode K-NN Untuk Memprediksi Hasil Pertanian di Kabupaten Malang," *Jurnal Mahasiswa Teknik Informatika*, vol. 3, no. 1, pp. 235–242, Mar. 2019.
[27] 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 *2023 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA)*, IEEE, Nov. 2023, pp. 496–500.
[28] M. Y. Anshori, V. Asy'Ari, T. Herlambang, and I. W. Farid, "Forecasting occupancy rate using neural network at Hotel R," in *2023 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA)*, IEEE, Nov. 2023, pp. 347–351.
[29] F. V. P. Samosir, L. P. Mustamu, E. D. Anggara, A. I. Wiyogo, and A. Widjaja, "Exploratory Data Analysis terhadap Kepadatan Penumpang Kereta Rel Listrik," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 7, no. 2, pp. 449–467, Aug. 2021, doi: 10.28932/jutisi.v7i2.3700.
[30] G. Leinhardt and S. Leinhardt, "Exploratory Data Analysis: New Tools for the Analysis of Empirical Data," *Review of Research in Education*, vol. 8, pp. 85–157, 1980, doi: 10.2307/1167124.
[31] Chairani, J. Harahap, and U. Zein, "Hubungan Pola Asuh Ibu dengan Karies Gigi Anak Balita di TK Perkebunan Nusantara I Kota Langsa Tahun 2019," *Journal of Healthcare Technology and Medicine*, vol. 9, no. 1, pp. 416–424, Apr. 2023, doi: 10.33143/jhtm.v9i1.2834.
[32] U. Hidayah and A. Sifaunajah, *Cara Mudah Memahami Algortima K-Nearest Neighbor Studi Kasus Visual Basic 6.0*. Jombang: LPPM Universitas KH. A. Wahab Hasbullah, 2019.
[33] F. Gorunescu, *Data Mining: Concepts, Models and Techniques*, vol. 12. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.
[34] Imam Robandi, *ARTIFICIAL INTELLIGENCE-Mengupas Rekayasa Kecerdasan Tiruan*. Yogyakarta: Andi Offset, 2021.
[35] D. Kriesel, *A Brief Introduction to Neural Networks*. Germany: ZETA2-EN, 2005. [Online]. Available: http://www.dkriesel.com/en/science/neural_networkshttp://www.dkriesel.com/en/tech/snipe