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

By Dina Novita, dkk.

WORD COUNT 5011
TIME SUBMITTED 03-JUN-2024 03:08PM
PAPER ID 109392989

BAREKENG: Journal of Mathematics and Its Applications
March 2022 Volume xx Issue xx Pages xxx–xxx
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengxxxxxxxxxxxxx

COMPARISON OF K-NEAREST NEIGBOR 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, Indonesia
Jl. Raya Sutorejo No 59, Mulyorejo, Surabaya, 60113, Indonesia
[2]Information System Department, FEBTD, Universitas Nahdlatul Ulama Surabaya
Jl. Raya Jemursari 51-57, Surabaya, 60237, Indonesia
[3]Postgraduate of Department of Information Technology, Universitas Pembangunan Nasional "Veteran" Jatim
Jl. Rungkut Madya No. 1, Gunung Anyar, Surabaya, 60294, Indonesia
[4]Department of Management, Universitas Narotama Surabaya, Indonesia
Jl. Arief Rahman hakim No 51, Sukolilo, Surabaya, 60117, Indonesia
[5] Department of Electrical Engineering, University of Malaya, Malaysia

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

ABSTRACT

Tourism is one of the government's priority sectors in contributing to economic growth. East Java is one of the provinces in Indonesia that has its own attraction for 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 were carried out using the k-NN (k-Nearest Neighbor) method and 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 simulation is carried out by dividing the data into training data and testing data in a ratio of 70:30. The creation of the k-NN model is carried out using k-values 2 to 7. The creation of a modern neural network is done using hidden layers 1 to 3. The prediction results that have been made using k-NN obtained optimal RMSE at k-values 2 which resulted in an RMSE of 1,594,674. Meanwhile, the prediction results that have been made using neural networks obtained optimal RMSE at two hidden layers which resulted in an RMSE of 1,873,355. So it is recommended to be able to use the k-NN algorithm in predicting 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.

Article History:

Received: date, month year
Revised: date, month yearAccepted: date, month year

Keywords:

Forecasting, Tourism, International Visitor, k-NN, Neural Network

How to cite this article:

First author, second author, etc., "TITLE OF ARTICLE," *together: j. Math. &; App.*, vol. xx, iss. xx, pp. xxx-xxx, Month, Year.

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 that has been identified by the government as one of the sectors that can contribute to economic growth, job creation, community development and encourage regional development [1]. So that 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 aimed at strengthening economic resilience for quality growth [2]. The progress of the tourism sector is also used as a measuring tool in assessing the economic stability and security of a region. If an area receives relatively many international visitors, then in general the area can be indicated to 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 visitor 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 the number of international visitors has an impact on increasing demand for accommodation provision [4]. There needs to be an increase in the provision of hotel rooms and other accommodation, so as not to cause a gap between demand and supply for these accommodations.

East Java is one of the provinces in Indonesia that has its own attraction for 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 in order 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 10 priority tourism destinations, where the benefits of major projects for the 10 tourism destinations of the 2020-2024 RPJMN project are to increase international visitors by 22.3 million arrivals (2024). With the policies carried out by the government to support the tourism sector, it directly increases international visitor visits in 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 classification hotel rooms 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 previous month [7].

With the increasing number of international visitors in East Java, the government and stakeholders need the right strategy in planning and decision making 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]. Based on adequate quantitative data for predictions about international visitors in East Java will be able to get great benefits, 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 Neigbor (k-NN) and neural network. The RMSE results from each method are then compared to see which RMSE works best.

Evelyn Fix and Joseph Hodges created the k-nearest neighbor algorithm (k-NN) in statistics as a non-parametric supervised learning technique in 1951. Regression and classification are two uses for it. The input in both situations consists of a data set's k closest training sample [10]. K-NN is referred to as case-based reasoning which is a methodology based on reasoning of cases based on training data of a case stored, trained, and accessed to solve new problems [11]. 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 [12].

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. 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 [13].

In previous studies, several forecasting methods were used estimation of closed hotels and restaurants in Jakarta as impact of corona virus disease spread using adaptive neuro fuzzy inference system [14], forecasting the number of Demam Berdarah Dengue (DBD) patients using the fuzzy method [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], stock price estimation using Unscented Kalman Filter (UKF) [18], forecasting of occupied rooms in the hotel using linear support vector machine [19], profitability estimation using H-Infinity and Ensemble Kalman Filter (EnKF) [20], analysis of demand and supply blood in hospital in Surabaya city using panel data regression [21], prediction of sunlight intensity using neural network and Adaptive Neuro Fuzzy Inference System (ANFIS) [22], estimation of closed hotels and restaurants as impact of Covid-19 spread using backpropagation neural network [23], neural network algorithm for breast cancer diagnosis [24], electronic nose for classifying civet coffee using Support Vector Machine (SVM), k-nearest neighbor (k-NN), and decision tree [25], forecasting agricultural products in Malang Regency using k-NN [26], forecasting average room rate using k-NN [27], forecasting occupancy rate using neural network [28]. In this study, a comparison of k-NN and neural network will be carried out to predict international visitors in East Java, so as to 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. Writing Mathematical Equations

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 2010 to February 2024. The data obtained is univariate data with a total of 290 records, so it is necessary to do Exploratory Data Analysis (EDA). EDA is a method of data exploration using simple arithmetic techniques and graphical techniques in summarizing observational data [29], [30]. Then this univariate data will be processed into multivariate, changed from period 1 to period 10, the pattern model of changing univariate data into multivariate can be seen in Table 1 [9], [31].

Table 1. Univariate to Multivariate Data Pattern

Pattern	Input lag	Output/Target
1	x1, x2, x3, x4, …, xp	xp+1
2	x2, x3, x4, x5, …, xp	xp+2
3	x3, x4, x5, x6, …, xp	xp+3
…	…	…
m-p	xm-p, xm-p+1 , xm-p+2, …, xm-1	xm

2.2. K-Nearest Neigbor

K-nearest neighbor (k-NN) is a classification technique that makes firm predictions on test data based on the comparison of K nearest neighbor [12]. The nearest neighbor is the trained object that has the greatest similarity value or the smallest dissimilarity of the old 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 [32].

The purpose of the k-NN algorithm is to classify new objects based on attributes and training samples. Where the results of the new test sample are classified based on the majority of categories in k-NN using the neighborliness classification as the predictive value of the new test data sample. The distance used is Euclidean Distance with the following formula (1) (2).

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

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

pi = sample data
qi = test data
i = data variable
d = distance
n = 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 [32], [33]:

1. Specifies the parameter k (number of closest neighbors).
2. Calculates the square of each object's Euclid jar (quei instance) against the given sample data using an equation.
3. Then sort the objects into groups that have the smallest Euclidean distance.
4. Collect the nearest neighbor classification category.

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

2.3. Neural Network

The basics of neural networks consist 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 [34]. 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 1.

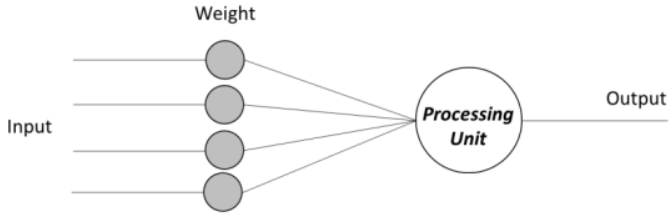

Figure 1. Design random forest model

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 x1, x2,... xn which is multiplied by its weight w1, w2 ,... WN expressed in equation (3):

$$Net = (w1.x1 + w2.x2 + w3.x3 + \ldots + wn.xn) \quad (3)$$

This concept can be written in vector notation as follows:

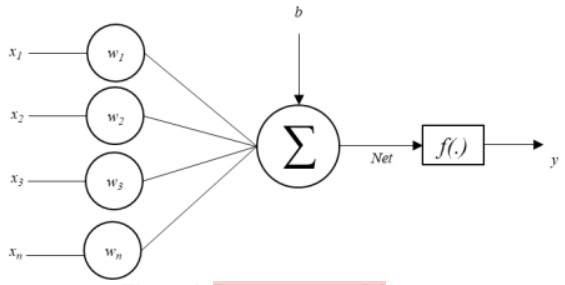

Figure 2. Perceptron model

A threshold value of b is called a bias, which plays an important role for some neuron models and needs to be referred to as a separate neuron model parameter. Various input conditions and influences on output are required to include a nonlinear activation function f(.) in the arrangement of neurons [35]. 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 (4):

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

3. RESULTS AND DISCUSSION

The dataset used in this study is the number of international visitor arrivals in East Java through the Juanda entrance from January 2000 to February 2024 (290 months). Furthermore, the data is split into training data and testing data with a ratio of 70:30. Then a test analysis was carried out with the k-nearest neighbor algorithm with k-values 2 to 7 and neural netrwok with 1 to 3 hidden layers. After that, a comparison of RMSE results from the test is carried out to obtain the optimal RMSE.

3.1. Exploratory Data Analysis

The dataset used in this study is the number of international visitor arrivals in East Java through the Juanda entrance from January 2000 to February 2024 (290 months). Furthermore, the data is split into training data and testing data with a ratio of 70:30. Then a test analysis was carried out with the k-nearest neighbor algorithm with k-values 2 to 7 and neural netrwok with 1 to 3 hidden layers. After that, a comparison of RMSE results from the test is carried out to obtain the optimal RMSE.

The dataset used in this study is the number of international visitor arrivals based on BPS-Statistics of East Java Province. Datasets that are still in univariate form will be processed into multivariate. Univariate data determined variable xt in the period to be created. Where the variable xt is the dependent variable and the variables xt-1 to xt-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

Year	Moon	Sum
2000	September	7854
2000	October	8949
2000	November	10926
2000	December	12357
2001	January	7488
2001	February	7051
...	...	...
2023	September	25593

Next is to convert univariate data to multivariate from 1 period to 10 periods, the result is as in Table 3 below.

Table 3. Multivariate Data Training at Input Layer 1 to 10 Periods

Xt	XT-1	XT-2	XT-3	...	XT-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 Neigbor Algorithm Test Analysis

The dataset of the number of international visitors has been converted into multivariate data, then divided into training data and teting data using a ratio of 70:30. 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 results of testing the k-nearest neighbor algorithm can be seen in Table 4.

Table 4. Test Results Using k-NN

K-Values	RMSE
2	1.594,674
3	1.622,701
4	1.688,320
5	1.636,585
6	1.713,065
7	1.795,068

From the forecasting carried out using k-NN with k-values 2 to 7, it is known that 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. Real comparison of international visitor data with prediction results using the k-NN algorithm can be seen in **Figure 3**.

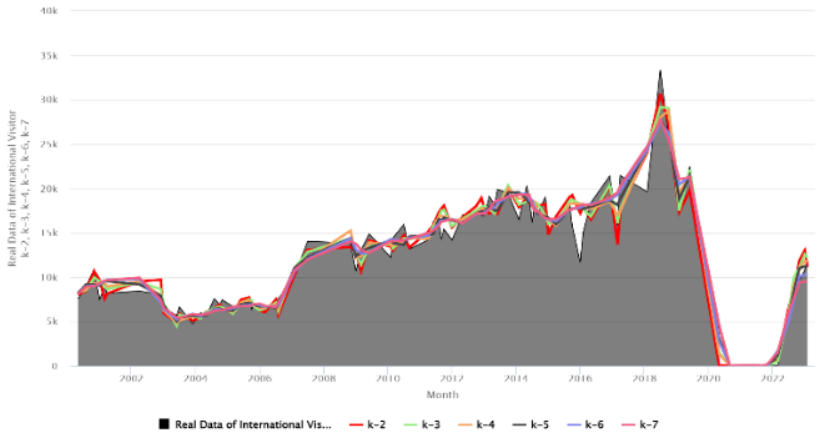

Figure 3. Prediction of international visitors using k-NN

In the graph above, the real data of international visitors is shown in the black plot type area. While 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 real data and 30,691 international visitors on prediction data with k-values 2 (a difference of 2,699 compared to real data), and international visitors with 27,685 on prediction data with k-values 7 (a difference of 5,705 compared to real data).

RMSE is best found in k-values 2 with an RMSE of 1,594,674. Real comparison of international visitor data with prediction results using k-NN with k-values 2 can be seen in Figure 4. In the graph, the black line shows real data of international visitors and the red line shows the predicted data of international visitors. Real 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 real 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 4. 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 teting 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 results of testing the neural network algorithm can be seen in Table 5.

Table 5. Test Results Using Neural Network

Hidden Layer	RMSE
1	1.985,373
2	1.873,355
3	1.933,628

From the forecasting carried out using a neural network with hidden layers as many as 1 to 3, it is known that RMSE is best found in two hidden layers with an RMSE of 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 test model using neural networks with each hidden layer size 8 can be seen in Figure 5 (a) for neural network algorithms with one hidden layer, Figure 5 (b) for neural network algorithms with two hidden layers, and Figure 5 (c) for neural network algorithms with three hidden layers.

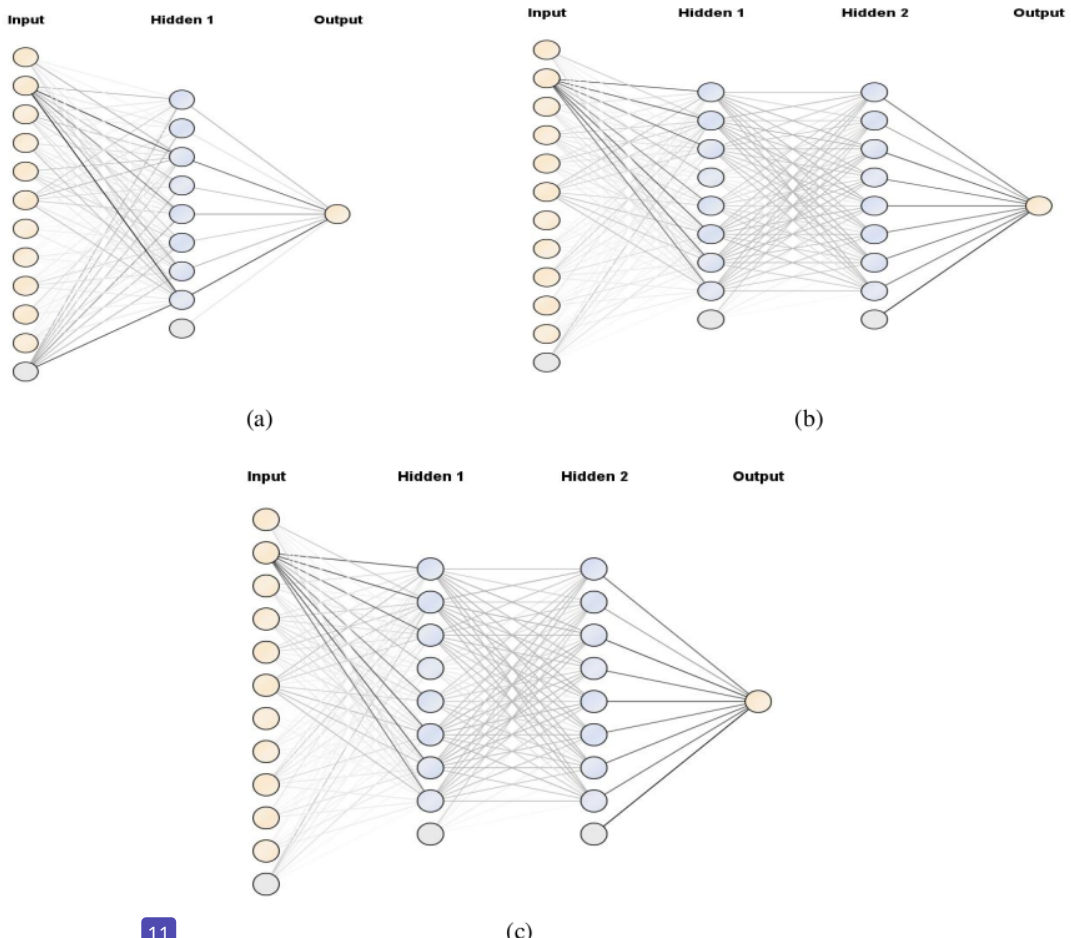

Figure 5. Neural network model, (a) one hidden layer, (b) two hidden layer, (c) three hidden layer

Real 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 6. In the graph, the real data of international visitors is shown in the black plot type area. While 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 real data and 29,306 international visitors on prediction data with two hidden layers (a difference of 4,084 compared to real data), and international visitors with 31,431 on prediction data with one hidden layer (a difference of 1,959 compared to real data).

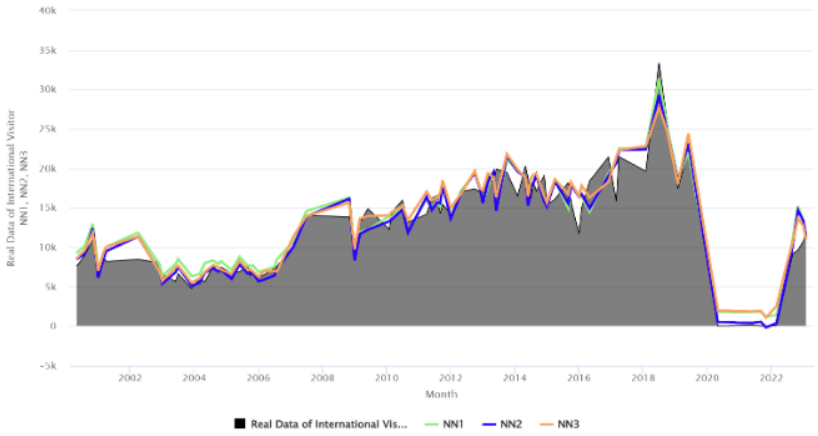

Figure 6. Predict international visitors using neural networks

RMSE is best found in two hidden layers with an RMSE of 1,873,355. A comparison of real data of international visitors with prediction results using neural networks with two hidden layers can be seen in Figure 7. In the graph, the black line shows real data of international visitors and the blue line shows prediction data of international visitors. Real 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 real 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 7. Predict international visitors using neural networks with two hidden layers

3.4. Algortihm Testing Comparison

Comparison of RMSE results using k-NN algorithm on k-values 2 and neural network on two hidden layers. The results of the comparison of these algorithm tests are shown in Table 6.

Table 6. Comparison of Algorithm Testing

Algorithm	RMSE	Remarks
k-NN	1.594,674	k-values = 2
Neural Netwok	1.873,355	Two hidden layers

The most optimal RMSE result is using k-NN with k-values 2 with an RMSE of 1,594,674, a difference of 278,681 compared to using a neural network algorithm with two hidden layers. Comparison of real data of international visitors, prediction of international visitor split data 70%:30% using k-NN algorithm with k-values 2 and neural network algorithm with two hidden layers shown in Figure 8. In the graph, the black line shows real 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.

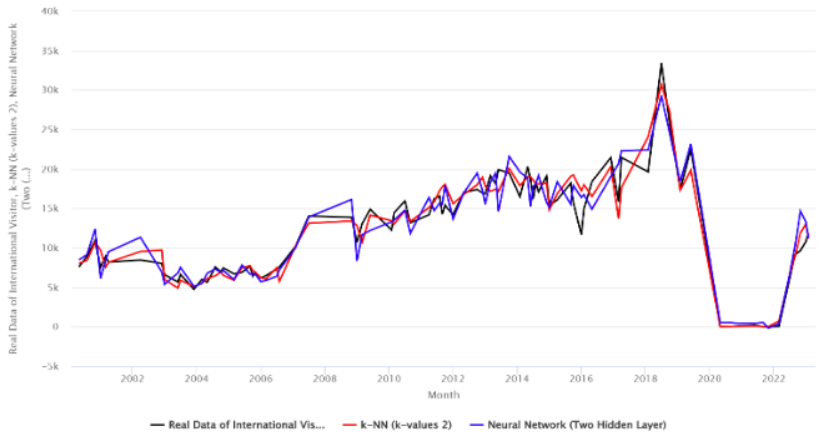

Figure 8. Comparison of international visitor predictions using k-NN with k-values 7 and neural networks with two hidden layers

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 results were obtained using k-NN with k-values 7 resulting in an RMSE of 1,594,674. So it is recommended to be able to use the k-NN algorithm as a prediction of 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. For other research, 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, p. 462, 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," 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," 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

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

ORIGINALITY REPORT

29%
SIMILARITY INDEX

PRIMARY SOURCES

1	Mohamad Yusak Anshori, Vaizal Asy'Ari, Teguh Herlambang, Imam Wahyudi Farid. "Forecasting occupancy rate using neural network at Hotel R", 2023 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA), 2023 Crossref	506 words — 12%
2	Vaizal Asy'ari, Mohamad Yusak Anshori, Teguh Herlambang, Imam Wahyudi Farid, Denis Fidita Karya, Mukhtar Adinugroho. "Forecasting average room rate using k-nearest neighbor at Hotel S", 2023 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA), 2023 Crossref	430 words — 10%
3	www.coursehero.com Internet	93 words — 2%
4	ojs3.unpatti.ac.id Internet	49 words — 1%
5	www.researchgate.net Internet	39 words — 1%

6	jurnal.unikal.ac.id Internet	34 words — 1%
7	ukinstitute.org Internet	31 words — 1%
8	publikasi.dinus.ac.id Internet	23 words — 1%
9	iptek.its.ac.id Internet	21 words — < 1%
10	Ali Khumaidi, Ika Ayu Nirmala, Herwanto Herwanto. "Forecasting of Sales Based on Long Short Term Memory Algorithm with Hyperparameter", 2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE), 2022 Crossref	17 words — < 1%
11	Nzar Shakr Piro, Ahmed Salih, Samir M. Hamad, Rawaz Kurda. "Comprehensive multiscale techniques to estimate the compressive strength of concrete Incorporated with carbon nanotubes at various curing times and mix proportions", Journal of Materials Research and Technology, 2021 Crossref	15 words — < 1%
12	docplayer.info Internet	15 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	< 15 WORDS