THE IMPACT OF DATA SPLITTING ON ANN PERFORMANCE IN PREDICTING FOREIGN TOURIST VISITS TO INDONESIA

Muhammad Dzakwan Alifi[1)]
Haidar Ramdhani[2)]
Lilis Indra Purnama[3)]
Shalma Kaisya Candradewi[4)]
Farid Yafi Suwandi[5)]
Adelia Putri Pangestika[6)]
Akbar Rizki[7)]
[1,2,3,4,5,6,7)]Department of Statistics, IPB University, Bogor, Indonesia
e-mail: akbar.ritzki@apps.ipb.ac.id

ABSTRACT

The data sharing stage is an important step in model building using Artificial Neural Network (ANN) methods to avoid the risk of overfitting and underfitting that can affect model performance. Proper data division aims to ensure that the model can generalize well to data that has never been seen before. Generally, data sharing is done by dividing the dataset into two main parts, namely training and testing data. However, to better address overfitting, there are also those who divide the data into three parts, namely training, testing, and validation. This study aims to evaluate the performance of ANN modelling using these two ways of dividing data. The model is evaluated using Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) metrics to measure prediction error. The data used is data on foreign tourist arrivals to Indonesia, which has a fluctuating pattern and is influenced by calendar effects. The results show that the data division type with two groups generally produces a smaller MAPE value than the data division into three groups. However, the model with two parts of data is not able to capture the seasonal pattern in the data. On the other hand, the model with three parts of data can overcome this problem better. The best model was obtained with the proportion of training data, validation data, and test data of 80%, 10%, and 10%, respectively, which resulted in a MAPE value of 24.45%.

Keywords: Artificial Neural Network (ANN), data splitting, prediction, tourism.

INTRODUCTION

Machine learning models have evolved into a very powerful tool in handling complex data analysis. One of the leading methods in machine learning is Artificial Neural Network (ANN), which is designed to capture non-linear patterns in data. ANN is used because of its pattern recognition ability to distinguish patterns in data (Amir et al., 2024). In addition, ANN also has the advantage of reducing noise in the data and it does not require any a priori assumption of the underlying process of the functional form of the dependency (Saikia et al., 2020). However, one of the advantages of ANN is the occurrence of overfitting (Srivastava et.al.,2014). These advantages make ANN an effective method for modeling data with high fluctuations such as tourist visit data, which mitigates splitting techniques to overcome the overfitting problem.

The data division stage plays an important role in building ANN models. Improper data partitioning can have a significant impact on the accuracy and stability of the forecasting model. Proper data sharing allows the model to learn from historical patterns without the risk of overfitting and

underfitting (Géron, 2022). In addition, it is also mentioned that data sharing can affect the accuracy of the resulting model (Birba, 2020).

Data division in machine learning is generally done into training data and testing data for model building. However, models are often biased and have unstable performance when the data used is complex and has many parameters (Srivastava et al., 2014)). Therefore, some studies add validation data to the data division type. Data division can be divided into 3 parts, namely train, test, and validation (Birba, 2020). This is done to avoid overfitting the hyperparameters in the test data and ensure that the test data is completely new and not used during model building. During the training process, validation data is used to avoid overfitting, validate the formation of the most optimal model, and select hyperparameters (Russell & Norvig, 2003). In addition, validation data is also used to ensure that the test data accurately demonstrates the model's ability to deal with real-world data.

Data sharing is a very important factor to ensure that the resulting model is accurate with minimum bias (Muraina, 2022). This becomes even more important when the model works on data with high volatility. Tourist visit data is one example of volatile data. This is because visitation patterns are often influenced by external factors such as pandemics, policy changes, or global tourism trends. This data is crucial for forecasting because it provides strategic information for governments and stakeholders in designing adaptive policies. Moreover, the COVID-19 pandemic has shown how vulnerable the tourism sector is to sudden external changes, making data-based forecasting a necessity that cannot be ignored.

Modeling tourist visitation data using ANN supported by appropriate data sharing strategies can improve prediction accuracy. Accurate forecasting results not only help understand visitation patterns in depth, but also provide a strong basis for strategic decision-making, especially in formulating post-pandemic tourism sector recovery policies (Niazkar & Niazkar, 2020). This study aims to look at the performance of ANN with different types of data division so that the resulting model can be measured more accurately which can be a solution to overcome volatility and maintain seasonal patterns in tourist visit data.

METHOD

Data Source

This study uses monthly data on the number of foreign visitors to Indonesia from January 1979 to August 2024. This data is obtained from the Central Bureau of Statistics (BPS) and consists of 548 observations in total. This data is able to provide a fairly comprehensive picture of the pattern of tourist visits for more than four decades.

Methodology

The data fed into the ANN model at each input includes arrival information for the past 6 months arranged sequentially to predict arrivals in the following month. The ANN model was trained with two types of data splits, i.e. without validation data and with validation data. The split with validation data follows a common approach in time series modeling, where training data is used to build the model and validation data to adjust hyperparameters without affecting results on test data (Kamilia & Yeni, 2023). The test data is then used to measure the generalization ability of the model on new data.

Steps of the ANN Method:

Artificial Neural Networks (ANN) are computational methods designed to mimic the way the human brain efficiently processes information. ANNs have the ability to approximate complex functions in situations where the relationship between input and output is highly complicated and non-linear (Hassoun, 2003). The ANN method was chosen because of its ability to capture non-linear patterns and complex time relationships, especially in time series data that show seasonal patterns or trends. The network structure used is the Multilayer Perceptron (MLP) model which consists of one input layer, two hidden layers, and one output layer. In this research, the ANN method is applied through several systematic stages to ensure that the resulting model has a high level of accuracy. The stages of the ANN method in this research are as follows:

a. Divide the data into two types. The first type into two parts: training data and test data, and the second type into three parts: training data, validation data, and test data. Training data is used for model building, validation data for hyperparameter tuning, and test data for measuring model generalization (Goodfellow et al., 2016).

Table 1. Division of Data by Type and Proportion Combination

Division Type	Training Data	Validation Data	Test Data
No Validation Data	60%	-	40%
	70%	-	30%
	80%	-	20%
With Validation Data	60%	20%	20%
	70%	15%	15%
	80%	10%	10%

b. Transforming data using Min-Max Scaling in the range [0,1] to be consistent and stable during training so that the model can handle data without scale bias. Min-Max Scaling is one of the preprocessing methods that transforms into a certain range by individually scaling each feature (Prasetyo et al., 2022). The data is transformed to ensure consistency of scale and increase stability during training. The Min-Max Scaling formula is as follows:

$$X_{sc} = \frac{(X - X_{min})}{(X_{max} - X_{min})} \quad (1)$$

c. Perform parameter tuning using grid search to test the optimal combination of hyperparameters, such as the number of neurons in the hidden layer, activation function, dropout rate, batch size, and number of epochs. Grid search is an exploratory method that tests various combinations of hyperparameter values on validation data to find the best configuration (Goodfellow et al., 2016). This method is effective when the number of hyperparameters tried is not too large because it is computationally inexpensive. In addition, to prevent overfitting, early stopping is applied which automatically stops training if the validation loss does not show a decrease after a few iterations.

Table 2. Tuning Parameters of Artificial Neural Network (ANN) with Grid Search

Parameters	Value
Number of Neurons in *Hidden Layer* (neurons)	5, 10, 20
Activation Function	ReLU, Tanh
Dropout Rate	0.0, 0.2
Batch size	16, 32
Number of Epochs	100

Using a model for forecasting one year ahead using multi-step forecasting techniques. The multi-step forecasting technique in long-term forecasting is done by making the previous month's prediction as input for the next month's prediction. The accuracy value of the model is used to evaluate how well the resulting forecasting model can accurately forecast (Goodfellow et al., 2016). In the initial stage, the model is evaluated for forecasting one year ahead to assess its accuracy in producing appropriate predictions. Model accuracy is measured using metrics such as Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE).

RMSE is a measure of the error rate in prediction results based on the difference between the actual value and the predicted value in the model with the following equation (Sari & Setiawan, 2024):

$$RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_{i=1}^{n}(D_i - Y_i)^2}{n}} \quad (2)$$

with:

D_i : actual data value in the i-th period

Y_i : forecast value in the i-th period

n : number of periods

Model performance assessment using RMSE does not have a minimum standard value. This is because RMSE depends on the unit of data used in the model. If the data used has a large unit, the resulting RMSE value will also be large, and vice versa. In addition, RMSE is known as a metric that has a common problem, which is quite sensitive to the presence of outliers (Zhang & Liu, 2023). Therefore, the RMSE value cannot be used as an absolute benchmark without considering the context and scale of the data used.

MAPE is the average value of the absolute difference that exists between the value of the prediction and the realized value expressed as a percent of the realized value (Nabillah & Ranggadara, 2020). MAPE is used if the size of the variable is an important factor in evaluating the accuracy of the forecast (Maricar, 2019). The MAPE value can be calculated using the following equation (Sari & Setiawan, 2024):

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|D_i - Y_i|}{D_i} x\ 100\% \quad (3)$$

with:
D_i = actual data value in period t
Y_i = forecast value in period t
n = number of periods

Table 3. MAPE Value Range

MAPE Range	Meaning
< 10%	Excellent Forecaster Model Capability
10 - 20 %	Good Forecaster Model Capability
20 - 50 %	Viable Forecaster Model Capability
> 50 %	Poor Forecaster Model Capability

RESULTS AND DISCUSSION

Data Exploration

The time series plot in Figure 1 shows the number of foreign tourist arrivals to Indonesia from January 1979 to August 2024 which illustrates fluctuations with an increasing trend from January 1979 to January 2020. The COVID-19 pandemic that occurred in the period January 2020 to February 2022 significantly affected the tourism sector. The Indonesian government implemented a policy of isolating some tourist attractions and limiting the number of tourists which caused a drastic decrease in the graph of the number of foreign tourist visits. This reflects the huge impact of the pandemic on the tourism sector in Indonesia. In addition, seasonal patterns are also seen in the data of the number of foreign tourist arrivals with the number of visits tending to increase at the beginning, middle and end of the year. The increase in the middle of the year can be attributed to the school vacation season in various countries, while the spike at the end of the year is often associated with Christmas and New Year holidays. This calendar effect shows that Indonesia's tourism sector is not only influenced by annual trends but also by global holiday seasons.

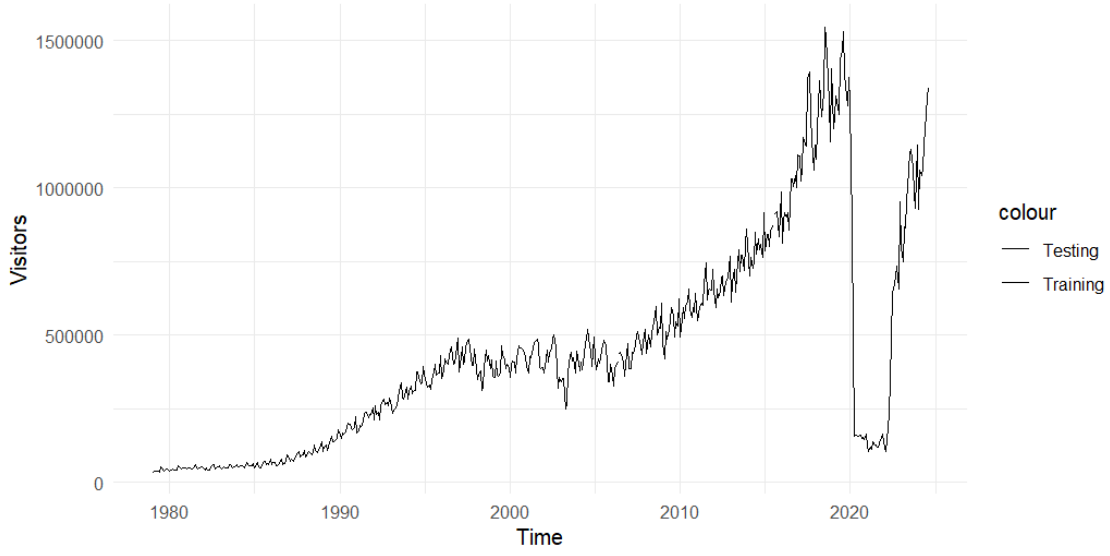


Figure 1. Data Exploration of the Number of Foreign Tourist Visits in Indonesia

Steep declines and sudden changes in patterns, especially during the COVID-19 pandemic, indicate high volatility in the data. This demands proper data sharing to build a stable and accurate model. Improper data sharing risks overfitting or underfitting leading to inaccurate predictions especially during periods of extreme fluctuations such as the pandemic. Therefore, proper data sharing is essential for the model to capture long-term patterns and adapt to sudden changes in trends or other external factors.

Data Sharing

Data division is an important step in the training process of the machine learning model. Training data is marked in blue, validation data in orange, and test data in green based on the division in Table 1. Conventional analysis methods usually involve dividing the data into training and test data only. The training data is used to build and train models to learn patterns in the data. The model tries to find the relationship between the input data and the desired output so that this process is influenced by the model parameters. Meanwhile, test data is data used to evaluate the best model that has been obtained through the training and tuning process with new data that is not included in the data for model formation. This aims to ensure that the performance of the model can be effectively used on real-world data. Certain studies add validation data sharing to ANN analysis to check or evaluate how well the model has performed by adjusting the hyperparameters and also avoiding overfitting.

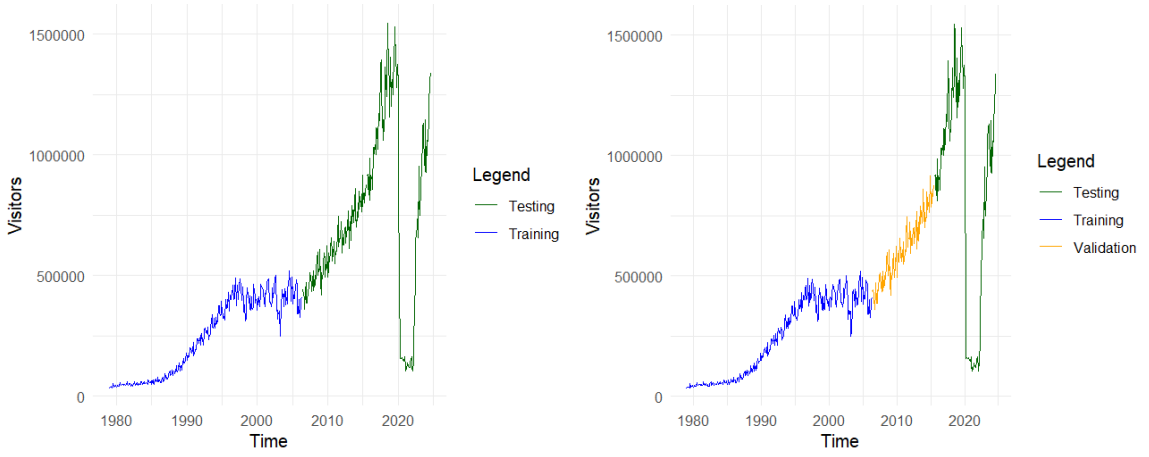


Figure 2. Example of the Distribution of Training Data, Validation Data, and Testing Data: (a) 60% Training Data and 40% Testing Data; (b) 60% Training Data, 20% Validation Data, and 20% Testing Data.

Tuning Parameters

Parameter tuning was performed in this analysis to find the parameters that minimize the error. Parameter tuning was performed for various proportions of training data, both with and without validation data, to evaluate the effect of data sharing on model performance. In addition, the test type was performed using a timestep of 6 to consider the seasonal pattern in the data. The following table

presents the parameter tuning results, including error values such as RMSE and MAPE, which were used as the main evaluation metrics.

Table 4. Best Parameter Tuning Results

No Validation Data							
Proportion of Training Data	Hidden Neuron	Activation Function	Dropout Rate	Batch Size	Epochs	RMSE	MAPE
60%	20	ReLU	0	16	100	28135.44	7.90
70%	20	ReLU	0	32	100	31019.87	7.82
80%	10	ReLU	0	16	100	34238.7	8.14
With Validation Data							
Proportion of Training Data	Hidden Neuron	Activation Function	Dropout Rate	Batch Size	Epochs	RMSE	MAPE
60%	20	ReLU	0	32	100	52476.27	6.77
70%	20	ReLU	0	16	100	64991.11	6.02
80%	20	ReLU	0	16	100	103583.04	6.78

Table 4 presents the results of parameter tuning for model optimization. The table shows several types of tests with parameter variations such as the number of hidden neurons, activation function, dropout rate, batch size, and number of epochs for each data division variation. Each type uses a timestep of 6 considering the seasonal pattern in each semester. These results show the initial performance of the model with a simple configuration.

In the first to third types, data sharing was performed without any validation data. Each type obtained a different parameter configuration and was tested using the RMSE and MAPE metrics. The RMSE value is seen to increase when the proportion of training data is increased, and on the other hand, the relative MAPE value decreases when the proportion of training data is increased, but in the third type there is an increase in the MAPE value from the second type, which is 7.82% to 8.14%. The lowest MAPE value is obtained from data sharing with a proportion of 70:30, which is 7.82%, using a configuration of 20 hidden neurons, ReLU activation function, dropout rate 0, batch size 32, and 100 epochs.

In the fourth through sixth types, data sharing was conducted involving validation data. As with the previous three types, parameter configuration was tested using the RMSE and MAPE metrics for each type. The RMSE value shows the same pattern as the type without validation data, indicating that the RMSE value gets larger when the proportion of training data is enlarged. The MAPE values show a volatile pattern, but it can be seen that the MAPE of the fourth and sixth types are almost the same at 6.77% and 6.88%. The lowest MAPE value is obtained from data sharing with a proportion of 70:15:15 which is 6.02%, using a configuration of 20 hidden neurons, ReLU activation function, dropout rate 0, batch size 16 and 100 epochs. In general, the RMSE value for the model configuration without validation data has a lower RMSE value and a higher MAPE compared to the model configuration with validation data. The second and fifth types using 70% of the data as training data appear to be quite

effective in building models that are effective in capturing data patterns, thereby increasing prediction accuracy.

Model Evaluation (RMSE and MAPE)

Performance analysis of parameter tuning shows that data sharing can affect the goodness of the model in the process of minimizing error in the training data. Furthermore, predictions were made for test data in the case of sharing without validation data, and predictions for validation data and test data in the case of sharing with validation data. The results of prediction and fitting of validation data and test data are shown in the following visualization.

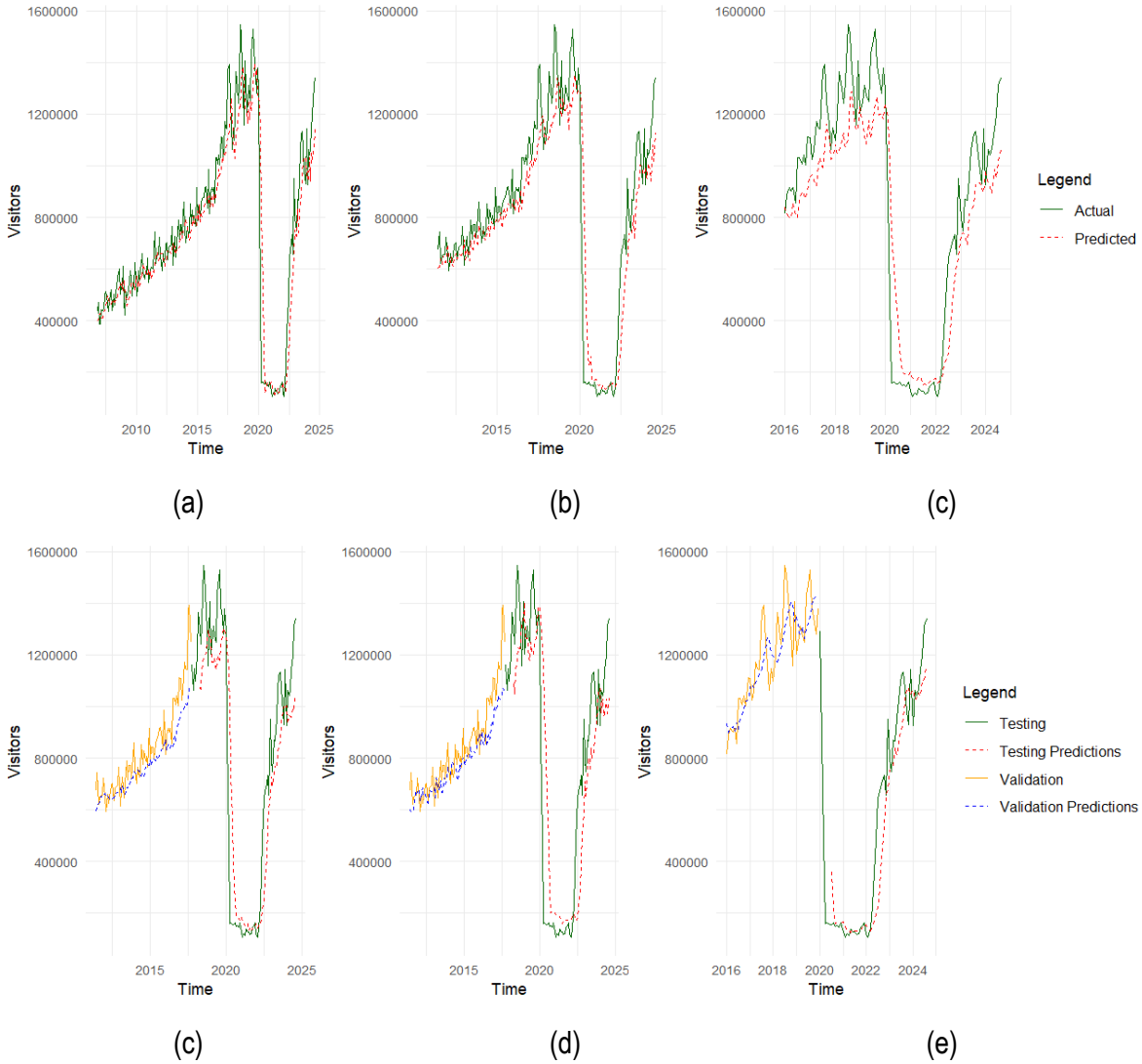


Figure 3. Prediction and Fitting Results of Validation Data and Test Data for All Proportion Variations. (a) 40% Test Data; (b) 30% Test Data; (c) 20% Test Data; (d) 20% Validation Data and 20% Test Data; (e) 15% Validation Data and 15% Test Data; (f) 10% Validation Data and 10% Test Data.

Figure 3 shows the prediction results for test data (top), as well as validation data and test data (bottom) for each variation of the proportion of training data (60%, 70%, and 80%) with a timestep of 6. The prediction lines for the case without validation data and with validation data tend to be similar for all variations of training data division. The accuracy value of the model for each variation of data division can be seen in the following table.

Table 5. MAPE and RMSE values

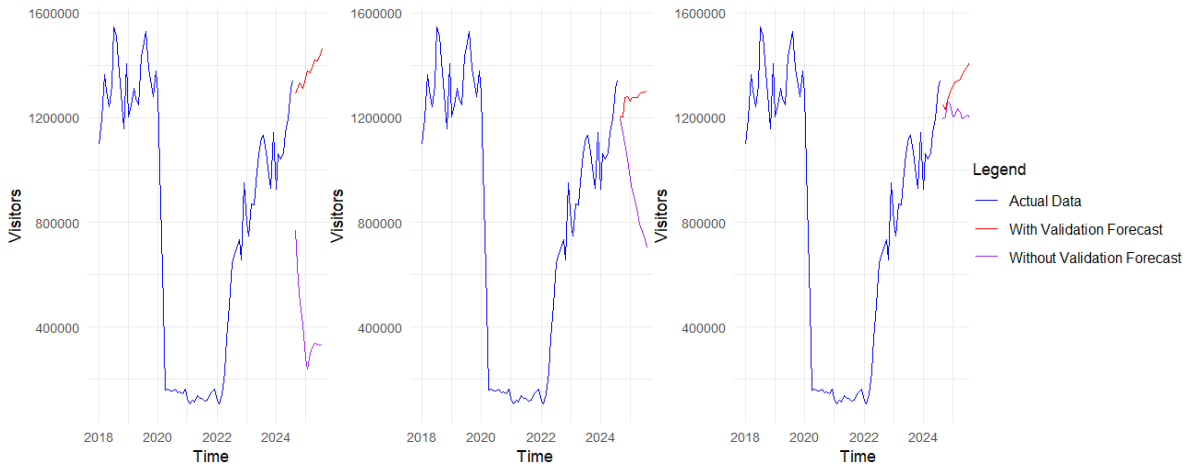
No Validation Data		
Training Proportion	Test RMSE	MAPE Test
60%	293683.65	24.64
70%	140841.46	16.24
80%	167129.62	23.05
With Validation Data		
Training Proportion	Test RMSE	MAPE Test
60%	201970.55	31.75
70%	215758.19	39.62
80%	137488.96	24.45

The models formed from the two types of data sharing provide different results and interpretations. In general, the model without validation data provides a smaller MAPE value than the model using validation data. On the contrary, when viewed from the RMSE, the model with validation data actually produces an average RMSE that is smaller than the model without validation data for all variations in the proportion of data sharing, except for the proportion of 70% training data sharing. At 70% training data proportion, the model without validation data obtained a smaller RMSE value than the model with validation data.

The different patterns of goodness-of-fit testing results of the RMSE and MAPE metrics on various data sharing variations may occur because they measure error in different ways. RMSE is known as a metric that has a common problem, which is quite sensitive to the presence of outliers (Chai & Draxler, 2014). This statement can be a reason to explain the inversely proportional pattern of RMSE and MAPE as the proportion of training data division increases or decreases, because the data used in this analysis does have a drastic decrease which is indicated as an outlier in the COVID-19 period.

The results of RMSE and MAPE on predictions without validation data and using validation data can be a reference for determining the right proportion of division to get the best value of one of the metrics or even both. However, it is still necessary to further review the performance of the model to find out whether the model really learns from actual data, one of which is by looking at the forecasting pattern. The pattern formed from the forecasting results can be matched with the actual data before forecasting for each model formed.

Visualization of Results



(a) (b) (c)

Figure 4. Forecasting with Variations in Training Data Share: (a) 60%; (b) 70%; (c) 80%

Figure 4(a) shows the forecasting results from the time period September 2024 to one year ahead for each type of data split (without validation data and using validation data) with three variations in the proportion of training data. The first graph shows the forecasting results of the 60% training data share. It can be seen that the forecasting results without validation data have a negative trend and the seasonal pattern tends to be different from the actual data. Forecasting results with validation data show a seasonal pattern with a positive trend in accordance with the expected pattern, and for the seasonal pattern shows a pattern that tends to be similar to the actual data in each semester.

Figure 4(b) shows the forecasting results from a 70% share of the training data. The forecasting results without validation data show a linear pattern with a negative trend, eliminating the seasonal pattern in the actual data. The forecasting results with validation data move in a positive trend and show a seasonal pattern that tends to be similar to the actual data.

Figure 4(c) shows the forecasting results of 80% training data share. Forecasting results without validation data show results that are quite similar to the seasonal pattern of actual data with the direction of movement tending to be straight (without trend). The forecasting results with validation data show a similar pattern to the forecasting pattern of the 60% training data share, but this time the seasonal pattern becomes almost invisible and forms a straight line.

Forecasting using validation data tends to maintain the seasonal pattern of the actual data and continue the trend pattern of the previous data. Forecasting patterns with seasonal patterns that are almost identical to the actual data are obtained at 60% and 70% of the training data. Meanwhile, forecasting without validation data tends to eliminate the seasonal pattern of actual data with a downward trend, except for the 80% training data division which maintains the seasonal pattern with a constant direction of movement without a trend.

CONCLUSION

The proportion of data division has a significant influence on the performance of predicting tourist visits to Indonesia using ANN. In general, dividing the data into two parts, namely training and testing, produces a smaller MAPE value compared to the model that divides the data into three parts (training, testing, and validation). However, models without validation data are not able to capture the seasonal patterns contained in the data, so the prediction results for the upcoming period are less accurate. In contrast, models using validation data can overcome this problem. The best model is obtained with the proportion of training, validation, and testing data of 80%, 10%, and 10%, respectively, which results in a MAPE of 24.45%. By using validation data, the problem of forecasting patterns that arise in the type without validation data can be overcome, as seen from the model's ability to maintain seasonal patterns in the prediction results for the next year.

REFERENCE

Amir, F., Utami, E., & Hanafi, H. (2024). Literature Study on the Development of Neural Networks For Weather Forecasting. *Jurnal Teknologi*, *17*(1), 49–57. https://doi.org/10.34151/jurtek.v17i1.4637

Birba, D. E. (2020). *A Comparative study of data splitting algorithms for machine learning model selection*.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (3rd Edition). O'Reilly Media, Inc.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. The MIT Press.

Hassoun, M. (2003). *Fundamentals of Artificial Neural Networks*. The MIT Press.

Kamilia, M., & Yeni, F. (2023). Validasi Data Pelanggan Menggunakan Customer Data Management dan Geographic Information System Melalui Website MyCX dan Starclick. *Journal of Network and Computer Applications*, *2*(1), 37–43.

Maricar, M. A. (2019). Analisa Perbandingan Nilai Akurasi Moving Average dan Exponential Smoothing untuk Sistem Peramalan Pendapatan pada Perusahaan XYZ. *Jurnal Sistem Dan Informatika*, *13*(2), 36–45.

Muraina, I. O. (2022). Ideal Dataset Splitting Ratios in Machine Learning Algorithms: General Concerns for Data Scientist and Data Analysts. *7th International Mardin Artuklu Scientific Researches Conference*, 496–504.

Nabillah, I., & Ranggadara, I. (2020). Mean Absolute Percentage Error untuk Evaluasi Hasil Prediksi Komoditas Laut. *JOINS (Journal of Information System)*, *5*(2), 250–255. https://doi.org/10.33633/joins.v5i2.3900

Niazkar, H. R., & Niazkar, M. (2020). Application of artificial neural networks to predict the COVID-19 outbreak. *Global Health Research and Policy*, *5*(1), 50. https://doi.org/10.1186/s41256-020-00175-y

Prasetyo, V. R., Mercifia, M., Averina, A., Sunyoto, L., & Budiarjo, B. (2022). Prediksi Rating Film Pada Website Imdb Menggunakan Metode Neural Network. *NERO (Networking Engineering Research Operation)*, *7*(1), 1–8.

Russell, S., & Norvig, P. (2003). *Artificial Intelligence: A Modern Approach*. Pearson Education, Inc.

Saikia, P., Baruah, R. D., Singh, S. K., & Chaudhuri, P. K. (2020). Artificial Neural Networks in the

domain of reservoir characterization: A review from shallow to deep models. *Computers & Geosciences*, *135*, 104357. https://doi.org/10.1016/j.cageo.2019.104357

Sari, A. S. N., & Setiawan, E. P. (2024). Comparison of Fuzzy Time Series Lee, Chen, and Singh on Forecasting Foreign Tourist Arrivals to Indonesia in 2023. *Jurnal Matematika, Statistika Dan Komputasi*, *21*(1), 10–32. https://doi.org/10.20956/j.v21i1.34914

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, *15*(56), 1929–1958.

Zhang, X., & Liu, C.-A. (2023). Model averaging prediction by K-fold cross-validation. *Journal of Econometrics*, *235*(1), 280–301. https://doi.org/10.1016/j.jeconom.2022.04.007