FOREIGN TOURIST ARRIVAL FORECASTING TO BALI USING CASCADE FORWARD BACKPROPAGATION

[a]Ayu Nikki Asvikarani, [b]I Made Widiartha, [c]Made Agung Raharja

[a,b,c] Computer Science Department, Faculty of Math and Science, Udayana University,
Bukit Jimbaran, Bali, Indonesia-803611, Telp: 081237787141
Email: nikkiasvikarani@gmail.com, madewidiartha@unud.ac.id, made.agung@unud.ac.id

Abstract

Bali has a recognized tourism potential in the world arena. In order to improve the quality and development of the tourism sector in the midst of global competition, it is necessary to formulate appropriate strategies by decision makers such as private parties and government. In support of more accurate decision making, the authors make a system of forecasting the number of foreign tourist visits to Bali Province using Cascade Forward Backpropagation (CFB) method with coverage of Australia, Japan, and United Kingdom which are the top 3 countries with the highest foreign tourist arrival to Bali in that years. Factors used as input in forecasting include the number of visits of foreign tourists the previous year, the population of countries of origin of foreign tourists, Gross Domestic Product at current prices of countries of origin of foreign tourists, and Relative Consumer Price Index Origin of foreign tourists. In this study, optimization of activation function parameters, hidden neurons, and learning rate to obtain forecasting results with the lowest error rate. Forecasting results using the CFB method produces a fairly good accuracy with MAPE range of 6 - 30% where the activation function tanh work better than sigmoid activation function.

Key words: artificial neural network, cascade forward backpropagation, forecasting, tourism.

INTRODUCTION

Based on data from the United Nations World Tourism Organization (UNWTO), tourism has experienced continued growth and deepened its diversification to become one of the fastest growing economic sectors in the world [1]. In Indonesia, based on data from the visit of foreign tourists belonging to the Central Statistics Agency, Bali's Ngurah Rai Airport is the entrance of foreign tourists with the highest number of visits. The Destin Asia's Readers Choice Award also places Bali as the main destination for holiday for 12 years in a row [2]. Bali received 4,927,937 foreign tourist arrivals in the January-December 2016 period where this figure rose 23.14% from the January-December 2015 period which reached 4,001,835 [3]. Although Bali's tourism trends have continued to increase over the past few years, the right development strategy must continue to be developed to be able to improve the quality and quantity of the tourism sector, especially in the face of the ASEAN Economic Community (AEC) which is an economic integration to build ASEAN as a single market and production base with the aim of making ASEAN more dynamic and competitive. AEC has been enacted since December 31, 2015 by carrying out 4 (four) pillars including a single market and a unified production base, competitive economic zones, equitable economic growth, and increased ability to integrate with the global economy. AEC will produce freer flow of goods, services, investment and labor in all ASEAN member countries [4]. This will cause economic competition to become tighter, including in tourism which is focused on services. Responding to this, a special strategy is needed in increasing the number of foreign tourists visiting Indonesia, especially to the Province of Bali. All forms of efforts to improve quality and quantity in the tourism sector cannot be separated from the budget that must be prepared by both the government and the private sector. Direct budget planning without strategy can provide potential losses to budget providers. One of the basics in developing strategies and making decisions is by forecasting [5].

Forecasting is an activity of estimating what will happen in the future by considering past events and the influence of current conditions. Forecasting an accurate number of foreign tourist visits will certainly provide many benefits to managers and investors in making decisions related to operations, planning and marketing, as well as investment strategies and assisting the government in making proper budget planning [6]. Based on this background, in this study the application of the Artificial Neural Network Cascade Forward Backpropagation (ANN-CFB) method was applied in predicting the number of foreign tourists visiting Bali.

RESEARCH METHOD

This research uses data of foreign tourist arrivals to Bali and their factors. The factors of foreign tourist arrivals are population of origin country, Gross Domestic Product (GDP) real of origin country and Consumer Price Index (CPI) in Indonesia relative to CPI of origin country [7]. Those data are taken from 1990 to 2016 and has time series characteristic. All data will be divided into 80% training data and 20% testing data [8].

Before entering into the ANN CFB process several stages will be carried out in analyzing the feasibility of the data including multiple correlation analysis (R test), determination analysis (R^2 test), and outlier detection of the data using IBM SPSS Statistics 20 software. Next will be the normalization process for the data that passes the preprocessing stage uses the equation below. In this study there is a comparison between the performance of two activation functions, namely binary sigmoid and tanh, so that there are two ranges of normalization values, 0 - 0.9 for binary sigmoid activation function and -0.9 - 0.9 for tanh activation function [9].

$$\hat{x} = \frac{(x-a)*(j-i)}{(b-a)} + i \tag{1}$$

with:

$\hat{x}$ = normalization data x
a = smallest value of a data set
b = largest value of a data set
i = bottom range data values
j = upper range data values

Artificial Neural Network Cascade Forward Backpropagation (ANN-CFB) has a similar way to Feed Forward Neural Network in using backward propagation algorithm in renewing its weight, but at CFB each layer is connected to all previous neuron layers [10].
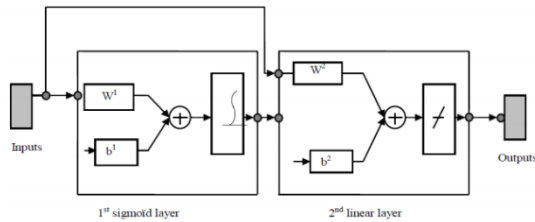
Fig 1. Cascade Forward Backpropagation Model.

The following is an abstract regarding the Cascade Forward Backpropagation algorithm [11].

1. Initialize weights with small random numbers.
2. For each combination (*pq, dq*) in the training sample:
   Forward propagation of the neural network layer:

$$a^0 = p_q; \ a^k = f^k(W^k a^{k-1} - b^k), \\ k = 1, \dots M \quad (2)$$

   Sensitivity backwards propagation at the neural network layer:

$$\partial^M = -2F'^M(n^M)(d_q - a^M); \\ \partial^k = F'^k(n^k)(W^{k+1})^T \partial^{k+1}, k \\ = M - 1, \dots, 1 \quad (3)$$

   Changes in weight and bias:

$$\nabla W^k = -\eta \delta^k (a^{k-1})^T, k = 1, \dots, M, \\ \nabla b^k = \eta \delta^k, k = 1, \dots, M, \quad (4)$$

3. Test the termination conditions.

RESULT AND ANALYSIS

Foreign Tourist Arrival Forecasting to Bali Using Cascade Forward Backpropagation uses 26 pairs of data from 1990-2016. There are 3 (three) countries tested which are the top 3 countries with the highest foreign tourist arrival to Bali in that year. They are Australia, Japan, and United Kingdom. The data used has 4 entries, namely the visit of foreign tourists in the previous year, the population of countries of origin of foreign tourists, GDP of countries of origin of foreign tourists, and the consumer price index of countries of origin of foreign tourists towards Indonesia. The output from the system is in the form of numbers of foreign tourists visiting Bali Province the following year. Furthermore, the feasibility of the data will be analyzed including multiple correlation analysis (R test), determination analysis (R2 test), and outlier detection of the data using IBM SPSS Statistics 20 software. Multiple Correlation Analysis (R) is used to determine the relationship between two or more independent variables (x1, x2, ..., xn) to the dependent variable (y) simultaneously. R values range from 0 to 1 where the closer to 1 means the stronger the relationship and the closer to 0, the weaker the relationship. While the Determination Analysis is used to determine the percentage contribution of the influence of independent variables (x1, x2, ..., xn) simultaneously to the dependent variable (y) [12]. Following is the result of R and R2 test of each country as shown in table 1.

Table 1. The result of R and R^2 test of each country.

No	Country	R	R Square	Ajusted R Square	Std. Error of the Estimate	Change Statistics				
						R Square Change	F Change	df1	df2	Sig. F. Change
1	Australia	,980[a]	,960	,953	68157,492	,960	126,995	4	21	,000
2	United Kingdom	,932[a]	,868	,843	13694,909	,868	34,561	4	21	,000
3	Japan	,804[a]	,647	,580	49105,936	,647	9,615	4	21	,000

a. Predictors: (Constant), consumer price index, number of arrivals, gross domestic product, population.

Based on the results of the R and R2 test of the three countries' datasets it can be concluded that all independent variables simultaneously are very influential on the dependent variable. All these data are divided into 80% training and 20% test data after normalizing process.

Desktop based application written in Python is made to implement Cascade Forward Backpropagation method. This

application will train and test the network and showed the comparison results between those 2 of two activation functions, namely binary sigmoid and tanh in chart form.

The network consists of 4 (four) input states where each of them represents input for foreign tourist arrivals, population of origin country, GDP of origin country and CPI of Indonesia relative to CPI of the origin country. There is 1 (one) hidden layer with specified number of hidden states. There is also 1 (one) output state which is numbers of foreign tourists visiting Bali Province the following year.

There are 120 configuration of network test for each time period. The configuration variables are the activation function, amount of hidden state, and learning rate. The maximum epoch in this test is 100000. The activation function is binary sigmoid and tanh. The number of hidden states is 2, 4, 8, 16, 32, and 64. The learning rate are 0.005, 0.006, 0.007, 0.008, 0.009, 0.01, 0.02, 0.03, 0.04, and 0.05.

Following is the summary of training results as shown in table 2.

Table 2. Architectural Summary Table of the Best Training Results

No	Countries	Activation Function	Hidden Neuron	Learning Rate	MSE Training
1	Australia	Sigmoid	16	0.05	0.0024643418184
2	Australia	Tanh	32	0.001	0.0093471661369
3	Japan	Sigmoid	4	0.006	0.0178291759086
4	Japan	Tanh	16	0.004	0.0704538826734
5	United Kingdom	Sigmoid	2	0.04	0.0050225487951
6	United Kingdom	Tanh	64	0.004	0.0135677712337

In the picture 2a – 4b will be displayed test results that represent forecast test comparisons of foreign tourist arrival to Bali by Cascade Forward Backpropagation and real data based on the best training results from each country. The blue line represents real data and the orange line represents ANN-CFB forecast.
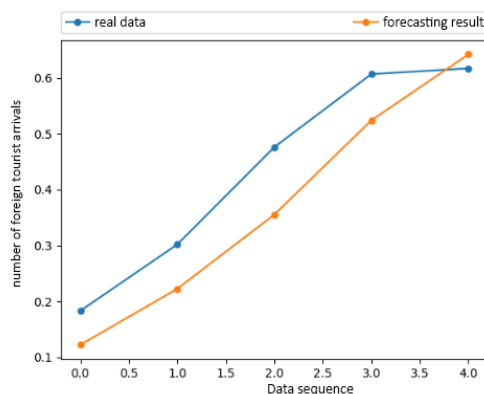
Fig 2a. Forecast comparison chart between ANN-CFB and real data of Australia with Sigmoid activation function.
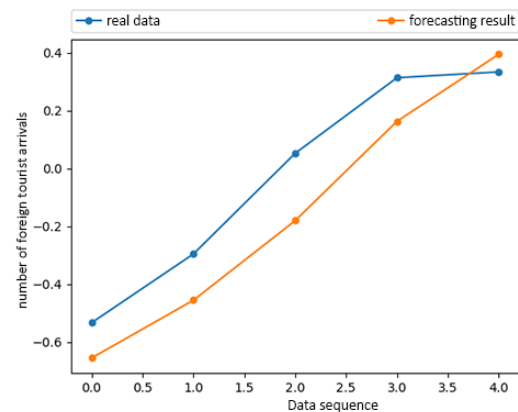
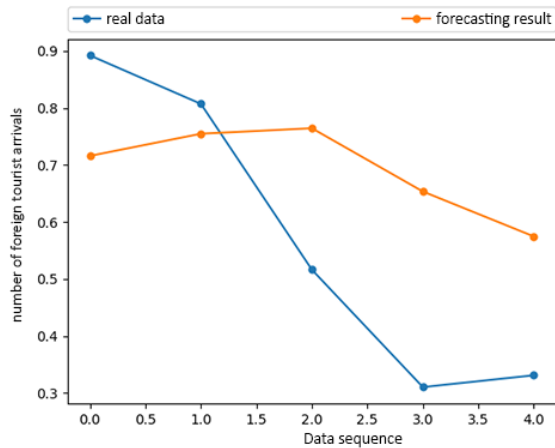Fig 2b. Forecast comparison chart between ANN-CFB and real data of Australia with Tanh activation function.

Fig 3a. Forecast comparison chat between ANN-CFB and real data of Japan with Sigmoid activation function.
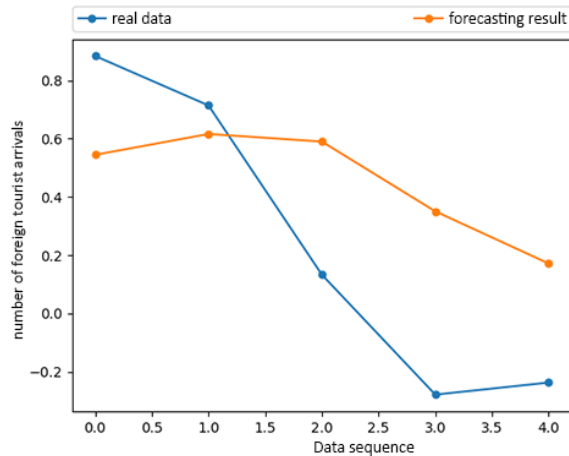

Fig 3b. Forecast comparison chart between ANN-CFB and real data of Japan with Tanh activation function.
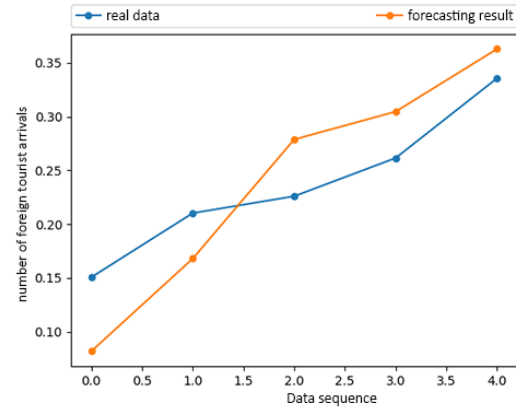

Fig 4a. Forecast comparison between ANN-CFB and real data of United Kingdom with Sigmoid activation function.
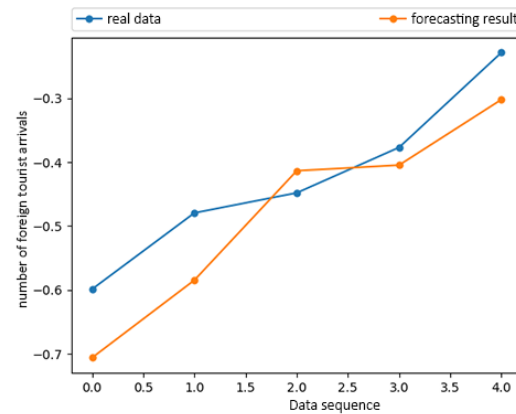

Fig 4b. Forecast comparison between ANN-CFP and real data of United Kingdom with Tanh activation function.

Following is the summary of testing results as shown in table 3.

Table 3. Architectural Summary Table of the Testing Results

No	Countries	Activation Function	Hidden Neuron	Learning Rate	MSE Testing	MAPE
1	Australia	Sigmoid	16	0.05	0.006405286	0.156330972
2	Australia	Tanh	32	0.001	0.024192653	0.154600411
3	Japan	Sigmoid	4	0.006	0.054417087	0.294608829
4	Japan	Tanh	16	0.004	0.179442555	0.266053573
5	United Kingdom	Sigmoid	2	0.04	0.002382481	0.090114425
6	United Kingdom	Tanh	64	0.004	0.005982572	0.066711597

Table 3 shows the network architecture with the best MSE results in each country. Based on the results obtained, the tanh activation function shows better performance in all countries, with a MAPE difference of about 1-3% better than sigmoid. The tanh activation function requires a smaller learning

rate compared to sigmoid but requires more hidden neurons than sigmoid.

Following is an example of the effect of changes in learning rate and hidden neurons in one of the 3 (three) countries tested namely Australia. In seeing the effect of the number of neurons on MSE, an experiment with a combination of training parameters was carried out which produced the smallest MSE in each activation function. The best architecture of the binary sigmoid function has 16 hidden neurons and a learning rate of 0.05,

while the best architecture of the tanh activation function has 32 hidden neurons and a learning rate of 0.001. In this experiment changes were made to hidden neurons ranging from 2, 4, 8, 16, 32, and 64 with a learning rate of 0.05 for the activation function of the sigmoid binary and learning rate of 0.001 for the tanh activation function. Figure 10 is an MSE table of the best training parameter combinations in terms of the number of neurons in the hidden layer.
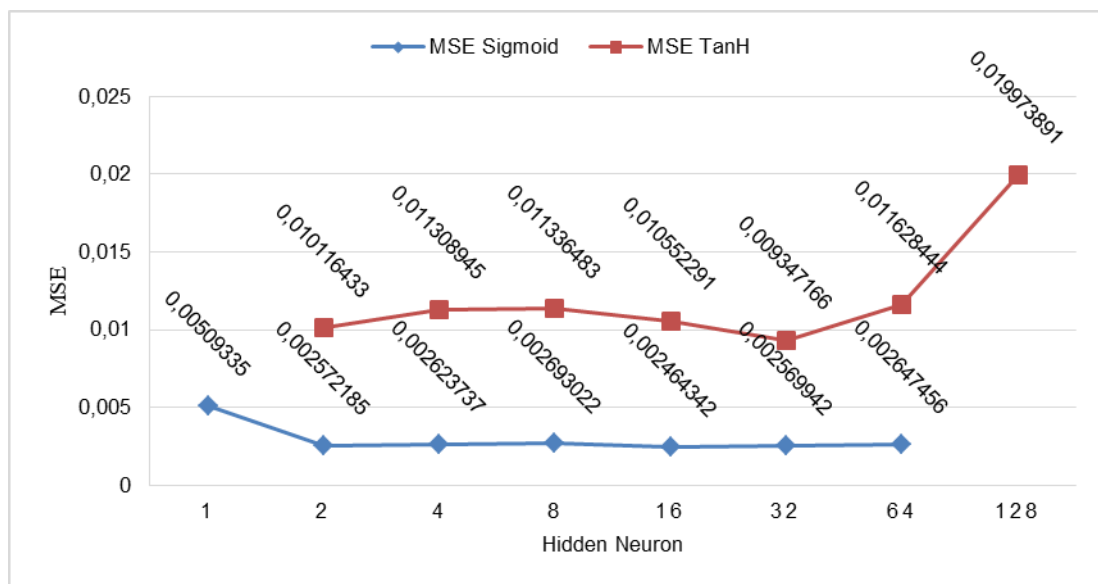

Fig 5. Graphic Effect of Changing the Number of Hidden Neurons on MSE Australia Dataset

Next is shown the effect of learning rate changes on Australia dataset network architecture. In this experiment changes were made to the learning rate values ranging from 0.005 - 0.009 and 0.01 - 0.06 with 16 hidden neurons for the function of binary sigmoid

activation function. While the tanh activation function changes in learning rate from 0.0005, and 0.001 - 0.1 with 32 hidden neurons. Figure 11 is an MSE table of the best training parameter combinations in terms of learning rate
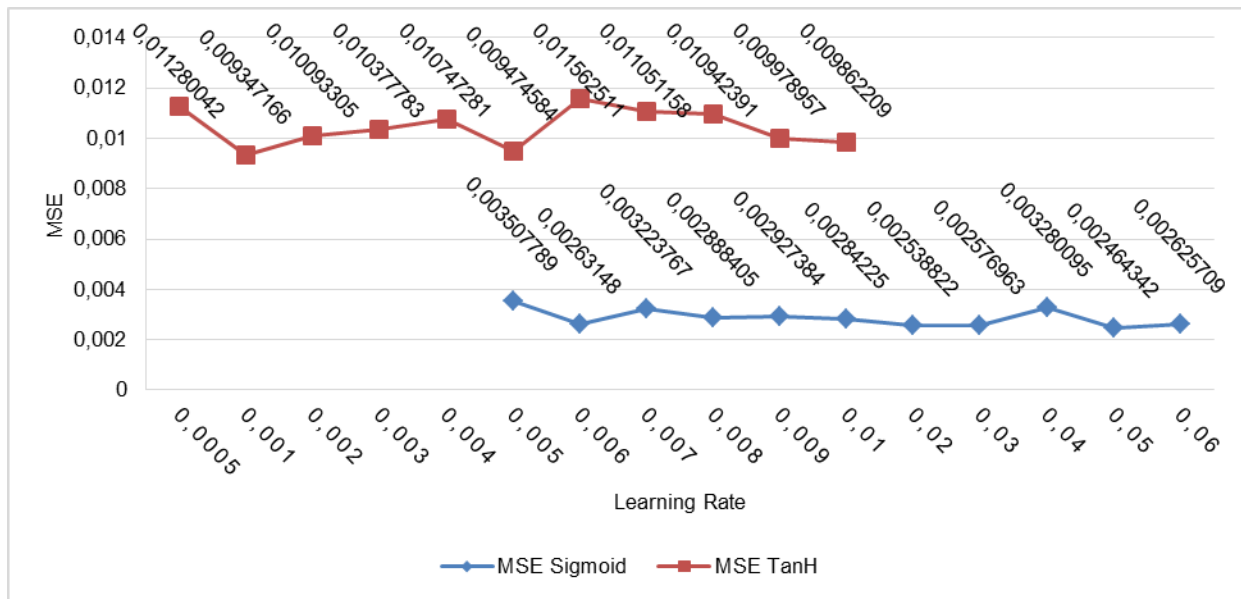
Fig 6. Graphic Effect of Changing the Learning Rate on MSE Australia Dataset

CONCLUSION

The conclusions that can be drawn from this study include the following.

1.   In this study ANN-CFB algorithm has been successfully implemented in forecasting the number of foreign tourists visiting Bali Province. Three parameters were changed, namely the activation function, the number of hidden neurons, and the learning rate. Based on the results obtained, the tanh activation function shows better performance in 3 countries, namely Australia, Japan, and the United Kingdom, with a MAPE difference of about 1-3% better than sigmoid. For each change in parameters, the number of neurons and the learning rate causes the MSE to fluctuate with a downward trend. The tanh activation function requires a smaller learning

.

rate compared to sigmoid but requires more hidden neurons than sigmoid.

In this study the results of each forecast are obtained by using 2 (two) different activation functions, namely the activation function of sigmoid and tanh. For Australia, the best architecture is obtained by architecture with the activation function tanh, 32 hidden neurons and a learning rate of 0.001 which results in MSE of 0.024 and MAPE of 15%. The best network architecture for the Japanese state dataset is architecture with the activation function of tanh, hidden neuron 16 and learning rate 0.004 which results in MSE 0.17 and MAPE 26%. The best network architecture for the UK country dataset is architecture with the Tanh activation function, 64 hidden neurons and a learning rate of 0.004, which results in an MSE of 0.002 and a MAPE of 6%

REFERENCES

[1] United Nations World Tourism Organization, "2017 Edition UNWTO," *UNWTO Tour. Highlights*, pp. 1–16, 2017, [Online]. Available: www.unwto.org.

[2] Contributor, "Readers' Choice Awards 2017," *DestinAsian Media Group*, Singapore, 2017.

[3] Disparda.baliprov.go.id, "Number of Foreign Tourists Arrival to the Province of Bali based on Nationality," *Disparda.baliprov.go.id*, 2017. .

[4] Asean.org, "Asean Economy Community," *Asean.org*, 2017. http://asean.org/asean-economic-community/ (accessed Nov. 20, 2017).

[5] G. P. Zhang, *Neural Networks in Business Forecasting*. 2004.

[6] L. C. David, *Industrial and Business Forecasting Method*. London: Butterworths, 1982.

[7] M. Hertinmalyana, A. H. Rahayu, and R. R. Wati, "Analysis of Demand and Consumption of International Visitors to Indonesia (from selected countries)," *13th Glob. Forum Tour. Stat.*, p. 400, 2014.

[8] Data.worldbank.org, "World Bank Open Data," *data.worldbank.org*. .

[9] W. O. Vihikan, I. K. G. D. Putra, and I. P. A. Dharmaadi, "Foreign tourist arrivals forecasting using Recurrent Neural Network Backpropagation Through Time," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 15, no. 3, pp. 1257–1264, 2017, doi: 10.12928/TELKOMNIKA.v15i3.5993.

[10] D. S. Badde, A. Gupta, and V. K. Patki, "Cascade and Feed Forward Back propagation Artificial Neural Network Models for Prediction of Compressive Strength of Ready Mix Concrete," *IOSR J. Mech. Civ. Eng.*, no. 2278–1684, pp. 1–6, 2009.

[11] S. Tengeleng and N. Armand, "Performance of using cascade forward back propagation neural networks for estimating rain parameters with rain drop size distribution," *Atmosphere (Basel).*, vol. 5, no. 2, pp. 454–472, 2014, doi: 10.3390/atmos5020454.

[12] B. P. Santosa and Ashari, *Statistical Analysis with Microsoft Excel and SPSS*. Yogyakarta: Andi Offset, 2005.

[13] Badan Pusat Statistik, "Number of Foreign Tourist Arrivals based on Entrance and Nationality," *bps.go.id*, 2017.

[14] Badan Pusat Statistik Provinsi Bali, "Statistics of Foreign Tourists in Bali Province," *bps.go.id*, 2016.