ANFIS Performance Evaluation for Predicting Time Series with Calendar Effects

Putriaji Hendikawati, Subanar, Abdurakhman and Tarno

Abstract—**In reality, most time series observations take the form of multivariate data that are influenced by many factors. In real-world modeling problems, too many inputs can increase calculation complexity due to the many parameters that must be estimated and resulting in reduced accuracy. This study uses the Adaptive Neuro-Fuzzy Inference System (ANFIS) method to apply various data preprocessing techniques, such as regression, Autoregressive Integrated Moving Average (ARIMA), and Autoregressive Integrated Moving Average with Exogenous Variable (ARIMAX), for the determination of potential input variables for time-series data subject to the calendar effect. The hotel room occupancy rate in the Special Region of Yogyakarta (DIY), which is influenced by the calendar effect, is predicted with this method. Preprocessing and correct sampling from, input data can have an impact on the prediction results. In general, data preprocessing improves efficiency. The empirical study shows that ANFIS preprocessing with the ARIMAX model provides the best results. This model obtained the smallest root mean square error (RMSE) for training and testing under the ANFIS model, i.e., 26,025.779 and 67,468,167, respectively. This empirical study shows that the preprocessing data that has been corrected according to calendar variations will positively impact the prediction performance. For ANFIS architecture, it can be considered to use triangular and gaussian membership functions with a minimal number of clusters and the grid-partitioning clustering method.**

Index Terms—**ARIMAX, ANFIS, data preprocessing, predicting, time series, calendar effect.**

I. INTRODUCTION

A method that is a combination of Artificial Neural Network (ANN) and Fuzzy Inference System (FIS), which is currently developing with many advantages, is Adaptive Neuro-Fuzzy Inference System (ANFIS). In recent years, ANFIS has been successfully used in modeling time series data in various fields, including forecasting [1]. Forecasting is a technique for estimating future conditions based on historical time series data and can help in planning and decision making. Forecasting time series data involves univariate analyses, whereas, in reality, most observations take the form of multivariate data that is influenced by many factors. However, forecasting, due to the fact that it considers other factors that influence historical data, requires other analyzes.

ANFIS modeling is based on fuzzy sets, membership functions, and inference systems. Generally, the selection of the ANFIS framework, such as which input variables to be use, the number of membership functions to be employed, and the number of fuzzy rules to be adopted, is done by trial and error. It is not uncommon in real-world modeling problems to build many potential frameworks for the model. A large number of inputs disturbs the transparency of the underlying model and increases the complexity of the calculations needed to build the model. When applying the ANFIS method, too many inputs can result in many training parameters, complicating the system and perhaps reducing the effect of ANFIS itself [2]. Therefore, an input selection method that prioritizes each input candidate and can be used according to the ANFIS framework is required. The selection of inputs includes removing noise or irrelevant inputs, removing inputs that depend on other inputs, making the underlying model more concise and transparent, and reducing the time taken to construct the model [2]. Also, according to [3], this preprocessing method reduces non-random noise from data, standardizes the data, and reduces the effect of scaling the data in the estimation process. The choice of indicators as input can help eliminate excessive inputs [4]. Preprocessing and correct sampling from input data can have an impact on the predictive results.

One time series model, an extension of the ARIMA time series, is called ARIMAX and consists of the ARIMA model with exogenous variables. In this model, the factors that influence the dependent variable Y at time t consist of the previous Y data over time and other independent variables measured at time t. Previous research on ARIMAX has discovered that the exogenous variables also influence the forecasting results. Calendar variations constitute one of the exogenous variables that can affect the prediction results of time series data. In Indonesia, a country with a muslim majority, calendar variations appear during religious holidays such as Eid al-Fitr. This Eid holiday exhibits repeating patterns that vary in length because events occur on different dates each year.

The Special Region of Yogyakarta (DIY) is one of the famous tourist destinations in Indonesia. Therefore, it is not surprising that every year the number of tourists, both foreign and domestic, who visit the DIY continues to increase. In this study, forecasting for the hotel occupancy rate in this region will be done by considering several factors affecting this rate, including the number of foreign and domestic tourists visiting the DIY. Data from Central Bureau of Statistics show that the number of tourists using hotel facilities in this region has increased every year.

In ANFIS, there are no fixed rules for determining the

Manuscript received October 09, 2020; revised April 05, 2021.

This research was carried out under the financial support from Lembaga Pengelola Dana Pendidikan (LPDP) Indonesia.

*P. Hendikawati is a PhD candidate of Mathematics Department, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia. She is also a lecturer from Mathematics Department, Universitas Negeri Semarang, Semarang 50229, Indonesia, (e-mail: putriaji.mat@mail.unnes.ac.id).

Subanar is a professor of the Mathematics Department, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia (e-mail: subanarseno7@gmail.com).

Abdurakhman is an associate professor of the Mathematics Department, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia (e-mail: rachmanstat@ugm.ac.id).

Tarno is an associate professor of the Statistics Department, Universitas Diponegoro, Semarang, Indonesia (e-mail: tarno.stat@gmail.com).

inputs used in the model. This study aims to compare the performances of various data preprocessing techniques, including regression, ARIMA, and ARIMAX, and determine which inputs are most suitable that influence the model. Calendar variations due to the Eid al-Fitr holiday will also be examined to discover how they impact the hotel room occupancy rate in the DIY. This research is expected to provide information and recommendations that will allow hotel managers to improve services and government agencies to develop policies related to the tourism sector. The rest of this paper is organized as follows. In Section 2, we present related works, and in Section 3, we describe the materials and methods used in this paper. In Section 4, we present the research framework. Section 5 contains an empirical study using data with calendar effects, compares the performances of the preprocessing methods, and examines the forecasting accuracy. In Section 6, we draw our conclusions.

II. RELATED WORKS

This section discusses previous studies related to preprocessing and prediction with ANFIS and calendar variations in time-series data. In particular, it will examine studies that focus on using ANFIS for prediction and finding the preprocessing techniques providing the best ANFIS architectures. It will then concentrate on the prediction of time series affected by calendar variations.

A. Preprocessing and Prediction with ANFIS

Empirical research on the use of ANFIS for time series data modeling has been carried out in recent years by several researchers, including [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], and [21]. Their research shows that the ANFIS approach is quite reliable and accurate when used in forecasting.

Preprocessing, and correct sampling from, input data can have an impact on the predictive results. In general, data preprocessing improves efficiency and generalization of data analysis. Some researchers have investigated preprocessing methods in ANFIS. [3] have preprocessed the data with autocorrelation function (ACF) for the ANFIS input. Meanwhile, [22] proposed a new hybrid forecasting method using the combined Multiple Output Dependent Data Scaling (MODDS)-ANFIS method, where MODDS was used to preprocess the data.

The data is preprocessed to scale each attribute in the dataset at intervals according to the proposed scaling method and improve the prediction algorithm's performance. [23] conducted a survey on preprocessing techniques used in data mining analyses. [24] used the ANFIS and Neural Fuzzy System method to predict inflation rate. [25] examined the development of an optimal ANFIS architecture formation procedure based on the lagrange multiplier test procedure; specifically, the input selection procedure, the determination of the number of membership functions, and the rules were examined. [26] used the imputation method in the data preprocessing stage to overcome the impacts of missing data on the observed values.

B. Prediction of a Time-Series Subject to a Calendar Effect

Calendar variations consists of two types, namely, trading and holiday variations. Trading day variations are caused by the number of trading days in each month. Holiday variations are due to lunar calendar system variations. Several authors have examined the calendar variation models. [27], [28], [29], [30], [31], [32], and [33] examined the effects of trading days. Holiday effects due to calendar variations have been studied by [29], [31], [34], [35], [36], and [37]. Major celebrations such as Eid, Easter, and the Chinese New Year can influence business activities and consumer behavior patterns.

A study by [38] showed that ARIMAX provides better forecasting results for data subject to calendar effects than feed forward neural network. [39] also developed a calendar variation model with ARIMAX that showed better forecasting results than the ARIMA seasonal method, decomposition, and even a neural network. [40] examined the effect of other variables on forecasting using the ARIMAX and VAR models. The results of their study indicate that exogenous variables also influence forecasting results.

III. THEORY AND METHODS

A. Regression with Categorical Variables

The relationship between a response variable and k predictor variables for subject i is determined by the multiple regression equation model formulated in [41].

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + \varepsilon_i. \quad (1)$$

where Y is the dependent or response variable, the $X_{i1}, X_{i2}, ..., X_{ik}$ represent the independent or predictor variables, β_0 indicates the intercept, the $\beta_1, \beta_2, ..., \beta_k$ indicate the regression coefficients, and ε_i is the error term for the model.

In a regression analysis, it is often not only quantitative predictor variables that affect the response variables but also qualitative attributes, such as calendar variations. To accommodate the existence of qualitative variables in the regression model, we use binary dummy variables with values of 0 or 1, depending on whether the observations are from a population with certain characteristics or not. In this study, the predictor variables are a combination of quantitative and dummy variables. The regression equation is expressed in the form

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 D_{i1} + \varepsilon_i. \quad (2)$$

where D_{i1} is the dummy variable. The regression model parameters are estimated using the least-squares method. Appropriate regression models can be used for predictions of the response variable based on the predictor variables [42].

B. Autoregressive Integrated Moving Average (ARIMA)

In equation (1), $X_{i1}, X_{i2}, ..., X_{ik}$ can be various variables affecting the response variable. If these variables are defined as $X_{i1} = Y_{t-1}, X_{i2} = Y_{t-2}, ..., X_{ik} = Y_{t-k}$, then equation (1) becomes

$$Y_t = a + b_1 Y_{t-1} + b_2 Y_{t-2} + ... + b_k Y_{t-k} + e_t \quad (3)$$

Equation (3) is still a regression equation, but it differs from (1) in that the variables in the right-hand section of equation

(3) represent previous values of the dependent variable Y_t. These time-lagged values trigger autoregression (AR).

The ARIMA model is widely used in time series predictions. ARIMA (p, d, q) is a combination of the AR model (p) and the moving average (MA) model (q), as seen in the equation below (see [43], [44], and [45]).

$$Z_t - \phi_1 Z_{t-1} - \phi_2 Z_{t-2} - ... - \phi_p Z_{t-p} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}. \quad (4)$$

In equation (4), the variable Z takes the place of the variable Y in equation (3). Using a backshift operator, Equation (4) can be written as

$$(1 - \phi_1 B - \cdots - \phi_p B^p) Z_t = (1 - \theta_1 B - \cdots - \theta_q B^q) a_t, \quad (5)$$

$$\phi_p(B) Z_t = \theta_q(B) a_t, \quad (6)$$

where $\phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\theta_q(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ each of which is a stationary AR and MA process a_t is a white noise process. a_t is a white noise process if $a_1, a_2, a_3, , ..., a_n$ iid with $E(a_t) = 0$, $Var(a_t) = \sigma_a^2$ constant and $a_t \sim WN(0, \sigma_a^2)$.

The ARIMA model can be used by including only certain significant lag parameters. This model is called a subset or additive model. The ARIMA subset model is part of the generalized ARIMA model, so it cannot be expressed in general terms. For example, the subset model for ARIMA is ARIMA ([1 3],0,[1,12]) which is written as

$$\left(1 - \phi_1 B - \phi_3 B^3\right) Z_t = \left(1 - \theta_1 B - \theta_{12} B^{12}\right) a_t.$$

This subset ARIMA model includes lag 1 and 3 in the autoregressive section, while in the moving average section includes lag 1 and 12, with several other parameters being zero.

C. Autoregressive Integrated Moving Average with Exogenous Variable (ARIMAX)

Time series modeling can be performed not only by using existing historical data but also by adding other variables that are considered to have a significant influence on the data to increase forecasting accuracy. The ARIMAX model is a modification of the ARIMA model with the addition of a predictor variable. In general, the form of the ARIMAX(p, d, q) model is given by

$$(1-B)^d \phi_p(B) Y_t = \mu + \theta_q(B)\epsilon_t + \alpha_1 X_{1t} + ... + \alpha_k X_{kt} \quad (7)$$

A calendar variation is one of the predictors that can be used in ARIMAX modeling.

D. Adaptive Neuro-Fuzzy Inference System (ANFIS)

The ANFIS architecture is an adaptive network that uses supervised learning and has a function similar to that of the Takagi-Sugeno fuzzy inference system. Assume that there is a fuzzy inference system that has two inputs x and y and one output f. The rule base then contains two fuzzy if-then rules of type Takagi-Sugeno as follows.

Rule 1: if x is A_1 and y is B_1, then $f_1 = p_1 x + q_1 y + r_1$

Rule 2: if x is A_2 and y is B_2, then $f_2 = p_2 x + q_2 y + r_2$

The ANFIS network in this paper consists of the five layers [46] described below:

Layer 1: Every node i in this layer is adaptive and has node function $O_{1,i} = \mu_{A_i}(x)$, for $i = 1, 2$ or $O_{1,i} = \mu_{B_{i-2}}(y)$, for $i = 3, 4$ where x, y are the inputs to node i and A_i, B_i, are the linguistic labels associated with the activation parameters for this layer. The output $O_{1,i}$ is the membership grade of the fuzzy set $A(A_1, A_2, B_1, B_2)$ given by the input membership function. The membership function for A can be any appropriate parameterized membership function. For example, consider the generalized bell membership function i.e. $\mu_{A_i}(x) = \frac{1}{(1+|\frac{-c_i}{a_i}|^{2b})}$ where μ_{A_i} is the degree of the membership function for the fuzzy set A_i and a, b, c are the parameter that can change the shape of the membership function and referred as premise parameters.

Layer 2: Every node in this layer is a fixed neuron and represents the firing strength of a rule. Each node multiplies all entry signals and sends them to the next node. Typically, the $T-norm$ operators, such as the AND, are used to represent the i-rule and obtain the output $O_{2,i} = w_i = \mu_{A_i}(x).\mu_{B_i}(y), i = 1, 2$.

Layer 3: Every node in this layer is labeled by N and called the normalized firing strength. Each node calculates the ratio of the first firing strength (w_i) to the sum of the overall firing strength in the previous layer, i.e., $O_{3,i} = \bar{w_i} = \frac{w_i}{w_1+w_2}$.

Layer 4: Nodes in this layer adapt to the output, defined as $O_{4,i} = \bar{w_i} f_i = \bar{w_i}(p_i x + q_i y + r_i), i = 1, 2$ where w_i is the normalized firing strength in the third layer, and p_i, q_i, and r_i being the consequent parameters.

Layer 5: A single neuron is the sum of all the outputs from the fourth layer. It is labeled as Σ, i.e. $overall\ output = O_{5,i} = \Sigma i \bar{w_i} f_i = \frac{\Sigma_i w_i f_i}{\Sigma_i w_i}$.

The hybrid algorithm learning method proposed by [2] is used to update these two parameters, which can train premise and consequent parameters to adapt to their environment. The hybrid algorithm is a combination of the back-propagation and least-squares methods. In a hybrid algorithm, the parameters for the premise and the consequences will pass through the network. A hybrid algorithm is used because the back-propagation algorithm used to train parameters in adaptive networks has been found to have convergence problems and tend to become trapped in local minimums. When the premise parameter is obtained, the final output will be a linear combination of the consequent parameters [46], namely,

$$\begin{aligned} f &= (\frac{w_1}{w_1+w_2})f_1 + (\frac{w_2}{w_1+w_2})f_1 \\ &= \bar{w_1}(p_1 x + q_1 y + r_1) + \bar{w_2}(p_2 x + q_2 y + r_2) \\ &= (\bar{w_1}x)p_1 + (\bar{w_1}y)q_1 + (\bar{w_1})r_1 + (\bar{w_2}x)p_2 + \\ &\quad (\bar{w_2}y)q_2 + (\bar{w_2})r_2. \end{aligned}$$

The hybrid learning algorithm consists of two parts, namely, the forward and backward paths. On the forward path, the premise parameters on the first layer must be in stable condition. The least-squares estimator (LSE) method is applied to correct the consequent parameters in the fourth layer. The LSE method can be applied to accelerate the convergence rate in the hybrid learning process due to the linear consequence parameters. Furthermore, after the consequent parameters are obtained, the input data is passed back to the adaptive network input, and the resulting output will be compared with the actual output.

In the backward path, the consequence parameter must be in a steady state. When an error occurs during the comparison between the output produced and the actual output, it propagates back to the first layer. Simultaneously, the premise parameters in the first layer are updated using the gradient descent or back-propagation learning method. The combination of LSE and gradient descent in a hybrid learning algorithm can ensure a faster convergence rate because it can reduce the dimensional search space in the original back-propagation method [47]. The procedure for the hybrid learning ANFIS method used in this study is based on that of [46].

IV. RESEARCH FRAMEWORK

The flowchart of the proposed model is given in Fig 1. The proposed method consists of two stages, namely, preprocessing the data and ANFIS analysis.

A. Preprocessing the Data

There are three main steps to ANFIS: preprocessing the data, determining the fuzzy rules, and evaluating model performance. Data preprocessing begins with data collection. The data used in this study are secondary data obtained from *visitingjogja.com*, which is the official portal of the DIY Provincial Tourism Office. The data are the hotel room occupancy rates for both starred and non-starred hotels in the DIY in monthly periods from January 2008 to December 2017. The number of foreign and domestic tourist visits to the DIY during the same period are also used. Here, the DIY hotel room occupancy rate (Y_t) is the dependent variable, and the numbers of foreign tourist visits ($X_{1,t}$) and domestic tourist visits ($X_{2,t}$) are the independent variables. Three dummy variables are used for the calendar variations, i.e., the calendar effects of the month in which Eid occurs (X_3), one month before the Eid (X_4), and one month after the Eid (X_5). For dummy variables, the value is 1 for time at calendar variation dan 0 for others.

A data preprocessing process was developed to obtain the most appropriate inputs for ANFIS. The preprocess methodology in this study includes selecting and determining variables that significantly influence the time series data. The significance of each coefficient variable in relation to the dependent variable is tested. In general, data preprocessing facilitates efficiency and improves model generalization capabilities. In this study, a stochastic forecasting analysis method, namely, regression, is combined with ARIMAX to capture better information for forecasting. To choose the best model for the preprocessing stage, we measure the model's accuracy using the sum of squared residuals (SSR), Akaike information criterion (AIC), and coefficient of determination (R^2). The model with the smallest SSR and AIC and with the largest R^2 is said to be a better fit for the data. In

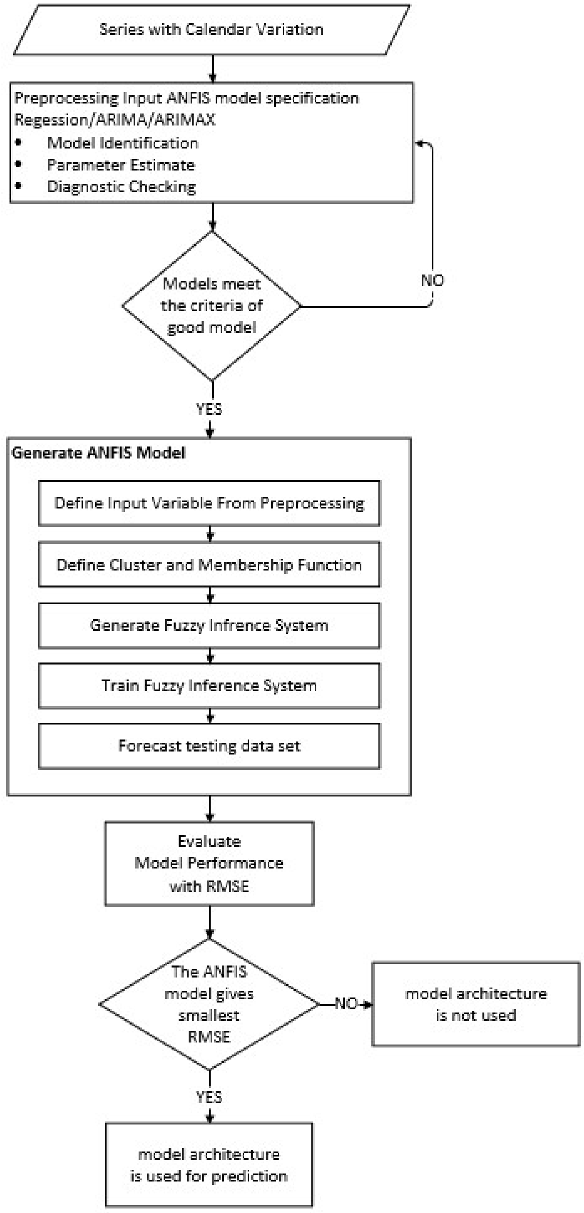

Fig. 1: Procedure of the Proposed Model

addition, the proposed model must also undergo diagnostic checking to discover whether the residuals meet the white noise assumptions and are normally distributed. Furthermore, adequate preprocessing techniques based on statistical tests will produce more valid and reliable results than determining the inputs by trial and error or guesswork.

B. ANFIS Modeling for Forecasting

Sugeno's ANFIS model with an architecture consisting of the following four stages is used.

a. Determine the input.
In the analysis using ANFIS, we divided the dataset into 80% training data (96 data points) and 20% testing data (24 data points). Data from January 2008 to December 2015 were used as the training data, while the testing data were the data from January 2016 to December 2017. Based on the data preprocessing results, the input variables used were the significant variables in the best model selected. Meanwhile, the

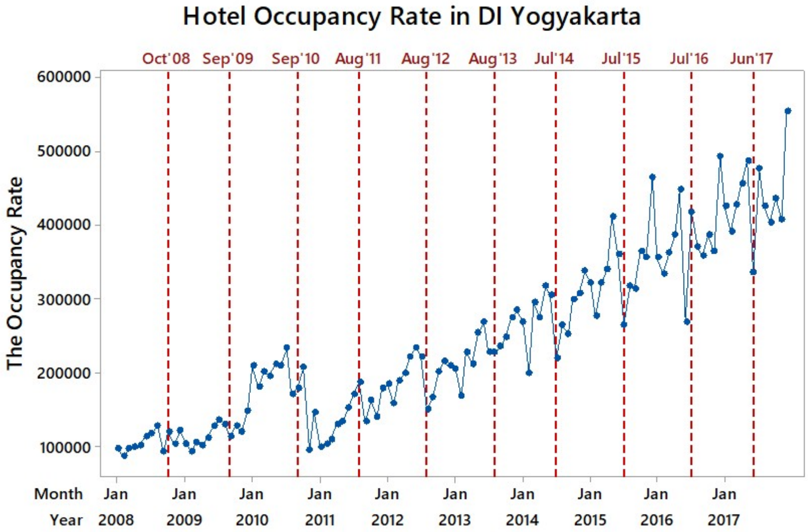

Fig. 2: Plot Hotel Occupacy Rate in DIY

target value was the hotel occupancy rate in the one subsequent period. At this stage, the data clustering method were determined, namely, grid partitioning and sub clustering.

b. Determine the membership function and fuzzy rules. Four membership functions were used: the triangular, trapezium, generalized bell shape, and Gaussian functions, while the output was modeled with a constant and linear function. The number of rules used corresponds to the number of membership functions (clusters) used.

c. Determine the learning algorithm. There are two types of ANFIS learning algorithms, i.e., the back-propagation and hybrid algorithms. This study applied the hybrid algorithm. According to [2], the hybrid learning method is more efficient. In the forward step, the least-squares method is used to identify the consequent parameters when the input is passed to layer 4. Next, in the backward step, the gradient descent determines the parameters for the premise.

d. Evaluate model performance. After a significant model is obtained, the forecast value of the training and testing process is then calculated using the $RMSE$ criteria. The $RMSE$ has the formula $\sqrt{\frac{1}{n}\sum_{t=1}^{n}(Z_t-\hat{Z}_t)^2}$, where $\hat{Z}_t$ is as the predicted value, Z_t is the actual value, and n is the predicted amount of data. The testing error (RMSE testing) use as the measure of the model performance [2]. The best ANFIS architecture achieved occurs when the testing error is minimal. The smaller the RMSE of the testing data, the better the architecture for prediction.

V. EXPERIMENTAL RESULTS

A. Preprocessing the ANFIS Data Input

Fig 2 shows the pattern in the data during the Eid al-Fitr period. The calendar variations are repeated and increased almost every year, except in 2015 and 2017 which show a decline. The month of the Eid al-Fitr holiday is shown as a vertical dotted line, with the month shifting forward every 3 years.

Data preprocessing in ANFIS begins with modeling and estimating data with regression and ARIMA. The regression model is used to investigate the relationship between the exogenous variables and the dependent variable, while ARIMA is used to examine the effect of past data on the hotel occupancy rate. All models to be used must pass diagnostic checking. The best ARIMA model is then remodeled by entering two exogenous variables that are considered to affect the hotel occupancy rate in the DIY and incorporating calendar effect variables. As explained in section IV-A, the Eid al-Fitr holiday is represented by a dummy variable. We choose the preprocessing model with a number of significant parameters at the 5% level of significance, a large R^2 value, small residuals, and a small AIC value as the best model for explaining the hotel room occupancy rate data. Preprocessing determines the input data used in ANFIS in several steps, as shown in Table I, II, III, and IV.

Table I shows some significant regression model alternatives, with and without calendar effects. Based on empirical studies of the regression model, it can be seen that the hotel room occupancy rate is influenced by another variable, namely, the number of foreign (X_1) dan domestic tourists (X_2) either partially or simultaniously. The number of foreign and domestic tourists both have a positive impact on the hotel room occupancy rate (the coefficients for the two variables are both positive). The regression models without a calendar effect indicate that the model with one input, namely, the number of foreign tourists, is the best model meeting the requirements of the regression test. Meanwhile, the best regression model with calendar dummy variables shows that only the number of foreign tourists (X_1) significantly affects the hotel room occupancy rate when three calendar dummy variables (X_3, X_4, X_5) are used. The three calendar effects have a significant negative impact on the hotel room occupancy rate. From these results it can be seen that the number of domestic tourist visits has no significant effect on the hotel room occupancy rate, so this variable is excluded from the model.

TABLE I: Prepocessing Data: Determination of Input for ANFIS using Regression Model

Preprocessing Model	Coefficient (Sig)	R^2	SSR	AIC	Dignostic Checking	Conclusion
Without Calendar Effect						
Liniear regression X_1 *	X_1: 12.202 (0.00)	0.708	4.63E+11	24.928	fulfilled	the best model without calendar effect
Linear regression X_2	X_2: 1.098 (0.00)	0.865	2.15E+11	24.160	not fulfilled	model cannot be used
Multiple regression X_1, X_2	X_1: 3.944 (0.00) X_2: 0.762 (0.00)	0.906	1.49E+11	23.809	not fulfilled	model cannot be used
With Calendar Effect						
Regression with dummy X_1, X_3	X_1: 12.427 (0.00) X_3: -56664.03 (0.006)	0.726	4.34E+11	24.879	fulfilled	alternative input model
Regression with dummy X_2, X_3	X_2: 1.099 (0.000) X_3: -7540.367 (0.588)	0.865	2.14E+11	24.174	not fulfilled	coef X_3 not sig model cannot be used
Regression with dummy X_1, X_2, X_3	X_1: 4.254 (0.000) X_2: 0.744 (0.000) X_3: -28453.47 (0.016)	0.911	1.41E+11	23.776	not fulfilled	model cannot be used
Regression with dummy X_1, X_3, X_4	X_1: 12.611 (0.000) X_3: -60644.83 (0.003) X_4: -40740.49 (0.044)	0.736	4.19E+11	24.861	fulfilled	alternative input model
Regression with dummy X_1, X_3, X_5	X_1: 12.774 (0.000) X_3: -64169.58 (0.001) X_5: -71049.66 (0.00)	0.755	3.89E+11	24.788	fulfilled	alternative input model
Regression with dummy X_1, X_3, X_4, X_5 **	X_1: 13.032 (0.000) X_3: -69742.31 (0.000) X_4: -50234.92 (0.01) X_5: -77337.37 (0.00)	0.769	3.67E+11	24.745	fulfilled	the best model with calendar effect

[*] The best regression model without calendar variation.
[**] The best regression model with calendar variation.

TABLE II: Prepocessing Data: Identification of ARIMA Model

Model	Coefficient (Sig)	R^2	SSR	AIC	Dignostic Checking	Conclusion
ARIMA(1,1,0)	AR(1): -0.459 (0.00)	0.205	3.833	-0.561	WN	alternative input model
ARIMA(1,1,1)	AR(1): -0.083 (0.633) MA(1): -0.490 (0.00)	0.244	3.648	-0.594	WN	coef AR(1) not sig model cannot be used
ARIMA([12],1,0)	AR(12): 0.508 (0.00)	0.267	3.535	-0.614	not WN	model cannot be used
ARIMA([1,12],1,0) *	AR(1): -0.384 (0.00) AR(12): 0.450 (0.00)	0.417	2.812	-0.826	WN	the best ARIMA model to construct ARIMAX

[*] The best ARIMA model

TABLE III: Prepocessing Data: Determination of Input for ANFIS using ARIMA([1,12],1,0) without Calendar Variation

Preprocessing Model	Coefficient (Sig)	R^2	SSR	AIC	Dignostic Checking	Conclusion
ARIMA([1,12],1,0) with X_1	AR(1): -0.387 (0.000) AR(12): 0.448 (0.000) X_1: 7.50E-07 (0.389)	0.423	2.782	-0.819	WN	coef X_1 not sig model cannot be used
ARIMA([1,12],1,0) with X_2	AR(1): -0.387 (0.000) AR(12): 0.447 (0.000) X_2: 8.39E-08 (0.287)	0.427	2.764	-0.827	WN	coef X_2 not sig model cannot be used
ARIMA([1,12],1,0) with X_1, X_2	AR(1): -0.387 (0.000) AR(12): 0.443 (0.000) X_1: 2.57E-07 (0.252) X_2: -2.05E-06 (0.424)	0.430	2.747	-0.816	WN	coef X_1, X_2 not sig model cannot be used

[**] There is no viable model at this stage

TABLE IV: Prepocessing Data: Determination of Input for ANFIS using ARIMA([1,12],1,0) with Calendar Variation

Preprocessing Model	Coefficient (Sig)	R^2	SSR	AIC	Dignostic Checking	Conclusion
ARIMA([1,12],1,0) with X_1, X_2 X_3, X_4, X_5	AR(1): -0.439 (0.00) AR(12): 0.368 (0.00) X_1: 1.05E-06 (0.74) X_2: 7.40E-08 (0.77) X_3: -0.145 (0.00) X_4: -0.108 (0.09) X_5: -0.035 (0.66)	0.471	2.554	-0.848	WN	coef X_1, X_2, X_4, X_5 not sig model cannot be used
ARIMA([1,12],1,0) with X_1, X_2, X_3	AR(1): -0.426 (0.00) AR(12): 0.398 (0.00) X_1: -4.60E-07 (0.87) X_2: 1.65E-07 (0.52) X_3: -0.156 (0.00)	0.458	2.615	-0.854	WN	coef X_1, X_2 not sig model cannot be used
ARIMA([1,12],1,0) with X_1, X_3	AR(1): -0.433 (0.00) AR(12): 0.394 (0.00) X_1: 1.38E-06 (0.09) X_3: -0.166 (0.00)	0.454	2.632	-0.865	WN	coef X_1 not sig model cannot be used
ARIMA([1,12],1,0) with X_2, X_3	AR(1):-0.428 (0.00) AR(12): 0.397 (0.00) X_2: 1.27E-07 (0.08) X_3: -0.159 (0.00)	0.458	2.616	-0.870	WN WN	coef X_2 not sig model cannot be used
ARIMA([1,12],1,0) with X_3 *	AR(1): -0.406 (0.00) AR(12): -0.422 (0.00) X_3: -0.117 (0.02)	0.435	2.725	-0.844	WN	the best ARIMAX model
ARIMA([1,12],1,0) with X_3, X_4, X_5	AR(1): -0.405 (0.00) AR(12): 0.412 (0.00) X_3: -0.108 (0.02) X_4: -0.055 (0.33) X_5: 0.006 (0.95)	0.437	2.713	-0.817	WN	coef X_4, X_5 not sig model cannot be used
ARIMA([1,12],1,0) with X_1, X_3, X_4	AR(1): -0.440 (0.00) AR(12): 0.347 (0.00) X_1: 1.76E-06 (0.05) X_3: -0.156 (0.00) X_4: -0.109 (0.06)	0.467	2.573	-0.876	WN	coef X_4 not sig model cannot be used
ARIMA([1,12],1,0) with X_2, X_3, X_4	AR(1): -0.433 (0.00) AR(12): 0.355 (0.00) X_2: 1.52E-07 (0.05) X_3: -0.146 (0.00) X_4: -0.100 (0.07)	0.468	2.565	-0.878	WN	coef X_4 not sig model cannot be used
ARIMA([1,12],1,0) with X_3, X_4	AR(1): -0.405 (0.00) AR(12): 0.405 (0.00) X_3: -0.107 (0.02) X_4: -0.054 (0.33)	0.438	2.712	-0.834	WN	coef X_4 not sig model cannot be used

[*] The best ARIMAX model

Table II shows the results for the ARIMA model. In addition to exogenous variables that affect the hotel room occupancy rate, it is believed that previous data also have a significant influence on rate predictions. The results indicate that the data from one and twelve previous periods affect the current forecast. The ARIMA([1,12], 1,0) model delivers the best result.

Because the hotel room occupancy rate data is time series data, past data can forecast future data. Based on the data plot, there are no clear seasonal patterns so seasonal ARIMA is not used here. However, the Eid holiday period shows an increasing pattern at these times. Data from one and twelve months previously have a significant effect on the hotel occupancy rate; in particular, a model utilizing this data has the largest R^2 and the smallest SSR and AIC values for the ARIMA([1,12],1,0) model. This best ARIMA model is then used to build the ARIMAX model.

Tables III and IV show the ARIMA modeling results obtained by adding exogenous variables, namely, the numbers of foreign and domestic tourist visits and the Eid calendar effect variable as independent variables. The ARIMA([1,12],1,0) model with the calendar effect for the month in which Eid occurred (X_3) is the best model. The model has independent variables that significantly influence the occupancy rate which is indicated by a significant variable coefficients, fulfill all criteria, and have the smallest

AIC value.

Examining Table IV, The exogenous variables in the best ARIMA model do not make a positive contribution when the calendar effect is considered, i.e., the numbers of foreign and domestic tourists were found to be insignificant variables when combined with the previous one-month (Y_{t-1}) and the previous twelve-months (Y_{t-12}) hotel room occupancy rate data. The variables from one previous period have negative coefficients, while the variables from twelve periods ago have positive coefficients. The data for the previous month (Y_{t-1}) can be assumed to be the hotel occupancy rate for the month before the Eid holiday. This period is the month of ramadhan, during which muslims fast for one full month and spend a lot of time worshipping at home with their families. This is because they do not travel outside the city, so not many use hotel facilities. This month corresponds to the pre-Eid calendar month dummy variable, which does not significantly affect the hotel room occupancy rate. However, things change if we include the dummy calendar effect variable for the month of the Eid holiday (X_3). It can be seen that the dummy variable for the month of the Eid holiday has a significant influence on the ARIMA([1,12],1,0) model. Of the five exogenous variables used to predict hotel occupancy rates, it was found that the variable X_2 did not significantly affect hotel occupancy rates. Only variables X_1, X_3, X_4, and X_5 had significant influence. Apart from these exogenous variables, past data have a big influence on forecasting.

From Table I, II, III, and IV, based on the criteria for selecting the best preprocessing model, there are eight models that meet the requirements with the best four models. The four best models are: 1) regression with the number of foreign tourist visits variable (X_1) without calendar effects; 2) the regression with the calendar effect, i.e. utilizing the number of foreign tourist visits (X_1) and the calendar effects during, a month before and a month after Eid (X_3, X_4, and X_5), 3) the ARIMA ([1,12],1,0) without calendar effect, and 4) the ARIMA ([1,12],1,0) model with calendar effect for the month of Eid (X_3). These models were chosen because all their independent variables have a significant effect on the hotel room occupancy rate and pass diagnostic checking. The models have the largest R^2 and the smallest AIC values in their class. The best model along with other models that meet the requirements, in the next stage are used as an alternative model to determine the ANFIS input variable.

B. Forecasting Accuracy of the Proposed ANFIS Method

Forecasting with the ANFIS method was performed according to the procedure stage, as described in subsection IV-B. The first step involves determining the inputs and the number and type of membership functions. Forecasting with the ANFIS method is done by using all the best preprocessing models with significant coefficients. The data preprocessing obtained eight alternative models with significant variables consisting of three models without calendar variation, and five models take in to account the calendar variation.

Table V and VI shows the results of the ANFIS analysis without and with calendar variations with several architecture modifications. The input variable is determined from the previous data preprocessing. Four membership functions are used. In this case, the number of rules and membership functions are limited to two and three because when there are too many membership functions, there will be more parameters to be estimated than the amount of data, guaranteeing that overfitting will occur. Two clustering methods were chosen, namely, grid partitioning and sub-clustering, with output functions of a constant and a linear function, respectively. In the training process, the error tolerance is set to 0, and the maximum number of epochs is set to 10.

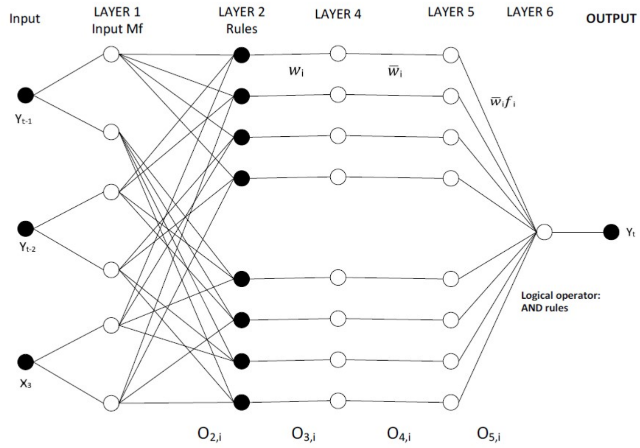

Fig. 3: Architecture of the best ANFIS model with 3 input and 2 rules

As seen in Table V and VI, the value marked with bold typeface indicates the smallest RMSE. The ARIMA ([1,12],1,0) and ARIMA ([1,12],1,0) with the calendar variations the month in which the Eid holiday occurs (X_3) has the smallest RMSE. The ANFIS prediction in Table V and VI show that the lowest RMSE training are 27159.608 and 26025.779 obtained by using a Gauss membership function with two rules and cluster. Meanwhile, the smallest RMSE testing are 69988.490 and 67468.167 obtained by using a triangular membership function with two rules and cluster. Both achieved when using the grid partition clustering method. Fig 3 shows architecture of the best ANFIS model with 3 input variabeles and 2 rules. The best ANFIS architecture obtained when the input variables are taken into account the calendar variation. This result shows that the ARIMAX model used to determine significant input variables at the data preprocessing stage provides more accurate results than the regression or ARIMA methods.

Fig 4 shows the plot of the forecast results for the original data in the training and testing process using the best ANFIS model. Fig 4(a) shows the training process where the circle shape shows the original data, and the star point is the prediction result. Meanwhile, Fig 4(b) shows the testing process where the dot shows the original data and the star point is the prediction result. It is seen that the prediction results can follow and approach the pattern of the original data, although there are still some data that have relatively large errors.

From these result, some initial remarks can be drawn. Firstly, based on these empirical studies, the hotel occupancy rate is influenced by other variables beyond the past data. Time lags of historical data contain information for future predictions [48]. Various studies has shown that, with the ability to study the pattern data from previous data, artificial

TABLE V: ANFIS Result with Significan Input from Prepocessing Data without Calendar Variation

Preprocessing Model	Input Variable	Clustering Type	Num Type	MFs MFs	Output	RMSE Training	RMSE Testing
Regression	X_1	Grid Partitioning	Trimf	[2]	Constan	57672.932	103181.531
			Trapmf			59635.052	118098.932
			Gbellmf			57685.989	108255.347
			Gaussmf			57354.940	104318.216
			Trimf	[3]	Constan	57201.555	112047.224
			Trapmf			62739.147	126373.159
			Gbellmf			59176.730	113724.584
			Gaussmf			57958.426	112048.735
		Sub Clustering	Gaussmf	[2]	Linear	56611.188	117910.658
ARIMA(1,1,0)	Y_{t-1}	Grid Partitioning	Trimf	[2]	Constan	37975.184	81227.508
			Trapmf			46943.388	77486.333
			Gbellmf			41366.088	74431.883
			Gaussmf			38773.095	75148.042
			Trimf	[3]	Constan	37519.553	83854.202
			Trapmf			46623.828	92390.913
			Gbellmf			40993.206	84309.036
			Gaussmf			38648.669	82544.181
		Sub Clustering	Gaussmf	[2]	Linear	46559.075	91504.404
ARIMA([1,12],1,0)	Y_{t-1}, Y_{t-12}	Grid Partitioning	Trimf	[2 2]	Constan	30047.194	**69988.490**
			Trapmf			33011.992	98549.858
			Gbellmf			30219.245	95725.934
			Gaussmf			29557.367	84195.215
			Trimf	[3 3]	Constan	28212.409	71091.577
			Trapmf			30435.774	95629.417
			Gbellmf			27575.040	92262.742
			Gaussmf			**27159.608**	70922.423
		Sub Clustering	Gaussmf	[3 3]	Linear	28833.444	69391.064

[*] The smallest RMSE of the ANFIS architecture with input without calendar variations.

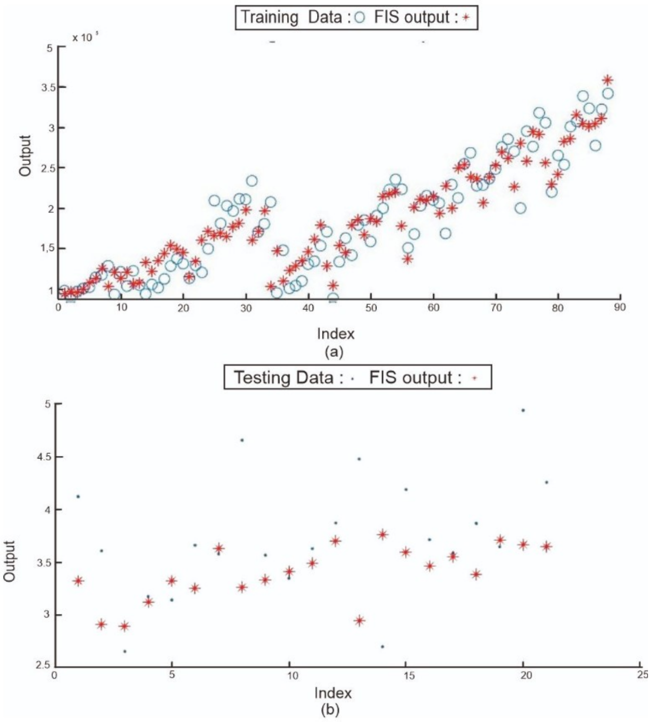

Fig. 4: Plot of the original data vs forecast data from the best ANFIS model in (a) the training and (b) testing process

network method is more efficient in forecasting results with less errors [21]. Also, we can see that the calendar variations affect the hotel room occupancy rate. The ARIMAX model and the influence of the calendar variations used in data preprocessing to determine the input variables for ANFIS turned out to provide better forecasting results than the regression and ARIMA method. This result indicates that it is necessary to pay attention to the effects of calendar variations, especially around major holidays. The results obtained indicate the importance of conducting further research to develop forecasting models that not only consider past time-series data but also the effects of other variables. We can do this by also taking into account the effects of calendar variations related to events and the timing of events in certain areas. Besides, data preprocessing is important in determining the ANFIS input needed to obtain more accurate prediction results.

In the ANFIS performance, the triangular and gaussian membership functions show smaller error values (RMSE) compared to the trapezoidal and generalized bell functions both in the training and testing processes. Also, using a small number of membership functions (clusters) will tend to reduce the error. The RMSE value in the testing process tends to be obtained when using a small number of membership functions. The empirical results show that 2 clusters (MFs number) tend to decrease the error value and give the best results. In the use of two clustering methods, the sub-clustering method shows that the RMSE training

TABLE VI: ANFIS Result with Significan Input from Prepocessing Data with Calendar Variation

Preprocessing Model	Input Variable	Clustering Type	Num Type	MFs	Output	RMSE Training	RMSE Testing
Regression	X_1, X_3	Grid partitioning	Trimf	[2 2]	Constan	56925.931	102073.835
			Trapmf			59026.047	117470.694
			Gbellmf			57012.195	107556.528
			Gaussmf			56627.357	103575.296
			Trimf	[3 3]	Constan	56360.431	108917.522
			Trapmf			62262.960	124591.637
			Gbellmf			58485.141	111160.573
			Gaussmf			57139.162	109193.991
		Sub clustering	Gaussmf	[2 2]	Linear	55679.394	117053.973
	X_1, X_3, X_4	Grid partitioning	Trimf	[2 2 2]	Constan	50728.948	99153.183
			Trapmf			52133.088	108908.004
			Gbellmf			49981.057	100987.412
			Gaussmf			49684.910	98653.201
			Trimf	[3 3 3]	Constan	50046.129	99556.211
			Trapmf			57943.283	114896.956
			Gbellmf			53053.691	101002.475
			Gaussmf			51119.801	99178.103
		Sub clustering	Gaussmf	[2 2 2]	Linear	48673.577	108833.142
	X_1, X_3, X_5	Grid partitioning	Trimf	[2 2 2]	Constan	54406.879	95892.886
			Trapmf			57792.791	115125.088
			Gbellmf			55386.260	103648.296
			Gaussmf			54675.557	98696.589
			Trimf	[3 3 3]	Constan	54245.889	97422.267
			Trapmf			60106.453	112637.733
			Gbellmf			56084.596	97884.616
			Gaussmf			54922.493	96914.921
		Sub clustering	Gaussmf	[2 2 2]	Linear	54064.224	105007.211
	X_1, X_3, X_4, X_5	Grid partitioning	Trimf	[2 2 2 2]	Constan	45363.803	94059.172
			Trapmf			49279.495	105423.459
			Gbellmf			46175.551	96153.865
			Gaussmf			45200.885	93562.182
			Trimf	[3 3 3 3]	Constan	45062.803	91820.891
			Trapmf			53065.963	116720.669
			Gbellmf			47170.872	109608.855
			Gaussmf			45735.132	97710.262
		Sub clustering	Gaussmf	[2 2 2 2]	Linear	44900.822	94811.387
ARIMA([1,12],1,0)	Y_{t-1}, Y_{t-12}, X_3	Grid partitioning	Trimf	[2 2 2]	Constan	28743.002	**67468.167**
			Trapmf			31802.361	96732.891
			Gbellmf			28977.405	92197.854
			Gaussmf			28295.001	81286.960
			Trimf	[3 3 3]	Constan	26896.384	102180.111
			Trapmf			29643.142	133699.447
			Gbellmf			26600.782	133480.064
			Gaussmf			**26025.779**	163272.577
		Sub clustering	Gaussmf	[4 4 4]	Linear	27660.067	78326.828

[*] The smallest RMSE of the ANFIS architecture with an input containing a calendar variation.

value is smaller than the grid partition method, but the performance in the testing process is better when using the grid partitioning method. In the training process, the use of the sub-clustering method shows a smaller error than grid partitioning. However, when using grid partitioning, the testing process's predictions always show the smallest error. Therefore, the use of grid partition for clustering is better.

VI. CONCLUSION

Based on the empirical results, it is important to pay attention to the influence of calendar variables in the data. It has been proven that calendar variations in this case, Eid al-Fitr events significantly influence data forecasting and allow ANFIS to produce more accurate forecasting results. Determining ANFIS input variables with time-series data containing calendar variations can then be done by going through the data preprocessing using the ARIMAX model for more accurate and statistically reliable results than relying on the trial-error method. The ANFIS model has the smallest RMSE value when the data is preprocessed using an ARIMAX with calendar variations. This result shows that ARIMAX can give better results than other preprocessing models, namely regression and ARIMA, that not take the calendar effect into account. This study shows that appropriate data preprocessing has a substantial impact on forecasting performance. Further research will be necessary to develop another approach or method that can be used to determine input variables for the ANFIS method when exogenous variables and calendar effects are considered. For ANFIS architecture, it can be considered to use triangular and gaussian membership functions with minimal rules and clusters and using the grid-partitioning method. For further developments, the proposed ARIMAX ANFIS method can also be applied to other cases influenced by calendar variations in Indonesia such as Eid al-Adha, Easter, Hindu religious holidays, and various other events related to calendar movements.

REFERENCES

[1] P. Hendikawati, Subanar, Abdurakhman, and Tarno, "A Survey of Time series Forecasting from Stochastic Method to Soft Computing", *Journal of Physics: Conference Series*, vol. 1613, 2020.

[2] J. S. Jang, "Input Selection for ANFIS Learning", in *Proceedings of IEEE 5th International Fuzzy Systems*, Vol. 2, pp. 1493-1499, 1996.

[3] A.Azadeh, M. Saberi, and S.M. Asadzadeh, "An Adaptive Network based Fuzzy Inference System Auto Regression Analysis of Variance Algorithm for Improvement of Oil Consumption Estimation and Policy Making: The Cases of Canada, United Kingdom, and South Korea", *Applied Mathematical Modelling*, vol. 35, no. 2, pp. 581-593, 2011.

[4] Z. M. Yunos, S. M. Shamsuddin, and R. Sallehuddin, "Data Modeling for Kuala Lumpur Composite Index with ANFIS", In *2008 Second Asia International Conference on Modelling & Simulation (AMS)*, pp. 609–614, 2008.

[5] M. Alizadeh, R. Rada, A. K. G. Balagh, and M. M. S. Esfahani, "Forecasting Exchange Rates: A Neuro-Fuzzy Approach", In *IFSA-EUSFLAT Conf*, pp.1745-1750, 2009.

[6] S. M. Fahimifard, M. Homayounifar, M. Sabouhi, and A. R. Moghaddamnia, "Comparison of ANFIS, ANN, GARCH and ARIMA Techniques to Exchange Rate Forecasting", *Journal of Applied Sciences*, vol. 9, no. 20, pp.3641-3651, 2009.

[7] H. Xu, and H. Xue, "Improved ANFIS to Forecast Atmospheric Pollution", *Management Science and Engineering*, vol. 2, no. 2, pp. 54-61, 2010.

[8] C. H. Cheng, and L.Y. Wei, "One Step Ahead ANFIS Time Series Model for Forecasting Electricity Loads", *Optimization and Engineering*, vol. 11, pp303-317, 2010.

[9] G. S. Atsalakis, E. M. Dimitrakakis, and C. D. Zopounidis, "Elliott Wave Theory and Neuro-Fuzzy Systems, in Stock Market Prediction: The WASP System", *Expert Systems with Applications*, vol. 38, no. 8, pp. 9196-9206, 2011.

[10] L. Wei, T. Chen, and T. Ho, "A Hybrid Model Based on Adaptive-Network-Based Fuzzy Inference System to Forecast Taiwan Stock Market", *Expert System with Applications*, vol. 38, pp. 13625-13631, 2011.

[11] M. Mordjaoui, and B. Boudjema, "Forecasting and Modelling Electricity Demand using ANFIS Predictor", *Journal of Mathematics and Statistics*, vol. 7, no. 4, pp. 275-281, 2011.

[12] F. K. Wang, K.K. Chang, and C.W. Tzeng, "Using Adaptive Network Based Fuzzy Inference System to Forecast Automobile Sales", *Expert System with Applications*, vol. 38, pp. 10587-10593, 2011.

[13] M. Ashish, and B. Rashmi, "Prediction of Daily Pollution using Wavelet Decomposition and Adaptive Network Based Fuzzy Inference System", *International Journal of Environmental Science*, vol. 2, no. 1, pp. 185-196, 2011.

[14] K. S. Lei, and F. Wan, "Applying Ensemble Learning Techniques to ANFIS for Air Pollution Index Prediction in Macau", In *International Symposium on Neural Networks*, pp. 509-516, Springer-Verlag, Berlin, 2012.

[15] Tarno, Subanar, D. Rosadi, and Suhartono, "Analysis of financial time series data using Adaptive Neuro Fuzzy Inference System (ANFIS)", *International Journal of Computer Sciences Issues*, vol. 10 no. 2, pp. 491-496, 2013.

[16] M. Savic, et.all, "Adaptive Network Based Fuzzy Inference System (ANFIS) Model Based Prediction of The Surface Ozone Concentration", *Journal of Serbian Chemical Society,* vol. 79, no. 10, pp. 1323-1334, 2014.

[17] K. Prasad, A.K. Gorai, and P. Goyal, "Development of ANFIS Model for Air Quality Forecasting and Input Optimization for Reducing the Computational Cost and Time", *Atmospheric Environment*, vol. 128, pp. 246-262, 2016.

[18] E. Çakıt, and W. Karwowski, "Predicting the Occurrence of Adverse Events using an Adaptive Neuro-Fuzzy Inference System (ANFIS) Approach with The Help of ANFIS Input Selection", *Artificial Intelligence Review*, vol. 48, no. 2, pp. 139-155, 2017.

[19] D. Karaboga, and E. Kaya, "Adaptive Network Based Fuzzy Inference System (ANFIS) Training Approaches: A Comprehensive Survey", *Artificial Intelligence Review*, vol. 52, no. 4, pp. 2263-2293, 2019.

[20] Hassan, Saima, Jafreezal Jaafar, Brahim B. Samir, and Tahseen A. Jilani, "A Hybrid Fuzzy Time Series Model for Forecasting", Engineering Letters, vol. 20, no. 1, pp88-93, 2012.

[21] Wongsinlatam, Wullapa, and Suntaree Buchitchon, "Criminal Cases Forecasting Model using A New Intelligent Hybrid Artificial Neural Network with Cuckoo Search Algorithm", IAENG International Journal of Computer Science, vol. 47, no. 3, pp481-490, 2020.

[22] K. Polat, "A Novel Data Preprocessing Method to Estimate the Air Pollution (SO2): Neighbor-Based Feature Scaling (NBFS)", *Neural Computing and Applications*, vol. 21, pp. 1987-1994, 2012.

[23] K. Gibert, M. Sànchez–Marrè, and J. Izquierdo, "A survey on Pre-processing Techniques: Relevant Issues in The Context of Environmental Data Mining", *AI Communications*, vol. 29, no. 6, pp. 627-663, 2016.

[24] N. R. Sari, A. P. Wibawa, and W. F. Mahmudy, "Comparison of ANFIS and NFS on Inflation Rate Forecasting", In *5th International Conference on Electrical, Electronics and Information Engineering (ICEEIE)*, pp. 123-130, 2017

[25] Tarno, Y. Wilandari, Suparti, and D. Ispriyanti, "Volatility Modeling of Financial Time Series Data Using ANFIS", *Advance Science Letters*, vol. 23, no. 7, pp. 6562-6566, 2017.

[26] T. W. Septiarini, and S. Musikasuwan, "Investigating the Performance of ANFIS Model to Predict the Hourly temperature in Pattani, Thailand", *Journal of Physics: Conference Series*, vol. 1097, no. 1, pp. 012085, 2018.
[27] L. M. Liu, "Note—Analysis of Time Series with Calendar Effects", *Management Science*, vol. 26, no. 1, pp. 106-112, 1980.
[28] S. C. Hillmer, "Forecasting Time Series with Trading Day Variation", *Journal of Forecasting*, vol. 1, no. 4, pp. 385-395, 1982.
[29] W. S. Cleveland, and S. J. Devlin, "Calendar Effects in Monthly Time Series: Modeling and Adjustment", *Journal of the American Statistical Association*, vol. 77, no. 379, pp. 520-528, 1982.
[30] W. R. Bell, and S. C. Hillmer, "Modeling Time Series with Calendar Variation", *Journal of the American statistical Association*, vol. 78, no. 383, pp. 526-534, 1983.
[31] R. Sullivan, A. Timmermann, and H. White, "Dangers of Data Mining: The Case of Calendar Effects in Stock Returns", *Journal of Econometrics*, vol. 105, no. 1, pp. 249-286, 2001.
[32] G. Kling, and L. Gao, "Calendar effects in Chinese stock market", *Annals of Economics and Finance*, vol. 6, no. 1, pp. 75-88, 2005.
[33] K. Evans, and A. Speight, "International Macroeconomic Announcements and Intraday Euro Exchange Rate Volatility", *Journal of the Japanese and International Economies*, vol. 24, no. 4, pp. 552-568, 2010.
[34] P. Brockman, and D. Michayluk, "The Persistent Holiday Effect: Additional Evidence", *Applied Economics Letters*, vol. 5, no. 4, pp. 205-209, 1998.
[35] R. C. Vergin, and J. McGinnis, "Revisiting the Holiday Effect: Is It On Holiday?", *Applied Financial Economics*, vol. 9, no. 5, pp. 477-482, 1999.
[36] F. J. Seyyed, A. Abraham, and M. Al-Hajji, "Seasonality in Stock Returns and Volatility: The Ramadan Effect", *Research in International Business and Finance*, vol. 19, no. 3, pp. 374-383, 2005.
[37] P. Alagidede, "Day of The Week Seasonality in African Stock Markets", *Applied Financial Economics Letters*, vol. 4, no. 2, pp. 115-120, 2008.
[38] Suhartono, "Neural Networks, ARIMA and ARIMAX Models for Forecasting Indonesian Inflation", *Widya Journal of Management and Accounting*, vol. 5, no. 3, 2005.
[39] M. H. Lee, and N. A. Hamzah, "Calendar Variation Model Based on ARIMAX for Forecasting Sales Data with Ramadhan Effect", in *Proceeding Regional Conference on Statistical Science*, pp. 349-361, 2010.
[40] W. Anggraeni, K. B. Andri, and F. Mahananto, "The Performance of ARIMAX Model and Vector Autoregressive (VAR) Model in Forecasting Strategic Commodity Price in Indonesia", *Procedia Computer Science*, vol. 124, pp. 189-196, 2017.
[41] A. C. Rencher, and G. B. Schaalje, "Linear Models in Statistics", *John Wiley & Sons*, 2018.
[42] S. Chatterjee, and A. S. Hadi, "Regression Analysis by Example", *John Wiley & Sons*, 2015.
[43] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, "Time Series Analysis, Forecasting and Control", *Prentice Hall*, 1994.
[44] S. Makridakis, S. C. Wheelwright, and R. J. Hyndman, "Forecasting Methods and Applications", *John Wiley & Sons*, 2008.
[45] W. S. Wei, "Time Series Analysis: Univariate and Multivariate Methods", *Pearson Addison Wesley*, 2006.
[46] J. S. R. Jang, C. T. Sun, and E. Mizutani, "Neuro-fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence", *Prentice Hall*, 1997.
[47] P. C. Nayak, K. P. Sudheer, D. M. Rangan, and K. S. Ramasastri, "A Neuro-Fuzzy Computing Technique for Modeling Hydrological Time Series", *Journal of Hydrology*, vol. 291, no. 1-2, pp. 52-66, 2004.
[48] Yudistira, Novanto, "COVID-19 Growth Prediction using Multivariate Long Short Term Memory", IAENG International Journal of Computer Science, vol. 47, no. 4, pp829-837, 2020.

Putriaji Hendikawati born in Jakarta on August 18, 1982, is a lecturer and researcher at the Mathematics Department, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Negeri Semarang. As a lecturer in the Mathematics Department, the author teaches several courses with interest in Statistics. She earned a bachelor's degree in mathematics (S.Si.) from Universitas Negeri Semarang (2004) and magister in mathematics, majoring in Statistics (M.Sc.) from Universitas Gadjah Mada (2010) and currently carrying out a doctoral program Universitas Gadjah Mada. Research interest in applied statistics and time-series data analysis.

Subanar born in Trenggalek on August 31, 1951, is a lecturer and researcher at the Mathematics Department, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Gadjah Mada (UGM), Yogyakarta, Indonesia. This mathematics professor in statistics earned a bachelor's degree in mathematics (Drs.) from Universitas Gadjah Mada (1976) and a doctorate (Ph.D.) from the University of Wisconsin-Madison (1987). He has written several books and is active in contributing ideas in journals and various international scientific conferences.

Abdurakhman is a lecturer and researcher at the Mathematics Department, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada (UGM), Yogyakarta, Indonesia. As a lecturer in the Mathematics Department, the author teaches several courses with interest in Statistics. He earned a bachelor's (S.Si.), magister (M.Si.), and doctoral (Dr.) degree in mathematics majoring in Statistics from Universitas Gadjah Mada. Research interest in applied statistics, especially in financial and actuarial mathematic statistics.

Tarno born in Boyolali on July 6, 1963, is a lecturer and researcher at the Statistics Department, Faculty of Sains and Mathematics, Universitas Diponegoro. He obtains a doctoral degree in mathematics (Dr.), majoring in Statistics from Universitas Gadjah Mada (2017). Research interest in applied statistics and time-series data analysis.