BAREKENG: Journal of Mathematics and Its Applications
March 2023 Volume 17 Issue 1 Page 0535–0544
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss1pp0535-0544

IMPLEMENTATION OF THE STEP FUNCTION INTERVENTION AND EXTREME LEARNING MACHINE FOR FORECASTING THE PASSENGER'S AIRPORT IN SORONG

Nur Faizin [1], Achmad Fauzan[2*], Arum Handini Primandari[3]

[1,2,3]Statistics Study Program, Faculty of Mathematics and Natural Sciences, Universitas Islam Indonesia
JL. Kaliurang KM 14,5, Sleman, Yogyakarta, 55584, Indonesia

*Corresponding author's email: * **achmadfauzan@uii.ac.id***

ABSTRACT

Article History:

Received: 25th November 2022
Revised: 6th February 2023
Accepted: 17th February 2023

Keywords:

Intervention ARIMA;
Step Function;
Extreme Learning Machine

This study aims to forecast the number of passengers departing at the domestic departure terminal at Domine Eduard Osok Sorong Airport in 2022 using the Autoregressive Integrated Moving Average (ARIMA) method, ARIMA with Step Function Intervention, and Extreme Learning Machine (ELM). The knowledge of the number of passengers can help the airport prepare facilities. The residual ARIMA model (0,1,0) has no serial correlation (random walk) based on the Ljung-Box test. The MAPE value of the ARIMA model (0,1,0) is 65.47% which means poorly fitted. Because of it, the researchers propose an intervention in the ARIMA model. The RMSE and MAPE ARIMA Intervention (1,0,0) (0,1,0) [12] were 9,027.671 and 35.86%, respectively. Besides, this study also employed the ELM method, which has a MAPE error measurement value of 30.64%. The ELM method has the lowest error measurement results among the three methods. Therefore, the ELM method is suitable for forecasting the number of passengers with predicted values from June to September 2022 as follows: 47985, 37821, 31247, and 33578. On the other hand, intervention in ARIMA can reduce MAPE by 45%.

How to cite this article:

N. Faizin, A. Fauzan and A. H. Primandari., "IMPLEMENTATION OF THE STEP FUNCTION INTERVENTION AND EXTREME LEARNING MACHINE FOR FORECASTING THE PASSENGER'S AIRPORT IN SORONG", *BAREKENG: J. Math. & App.*, vol. 17, iss. 1, pp. 0535-0544, March 2023.

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

Research Article • Open Access

1. INTRODUCTION

Time series data consist of stochastic and deterministic (trends, seasonal fluctuations, irregular cycles, or jumps) and are sometimes disturbed by interventions at point changes that cause shifts in levels or trends [1]. Time series analysis generally obtains time series data patterns using past data. Data patterns from time series analysis can be used to predict future data [2], [3]. The Covid-19 pandemic has caused a drastic decline in data on the number of passengers at Domine Eduard Osok Sorong Airport, and it was quite significant in April 2020 after the enactment of Government Regulation Number 21 of 2020 concerning Large-Scale Social Restrictions (PSBB) and Presidential Decree Number 12 of 2020 regarding the Covid-19 pandemic as a national disaster [4]. Therefore, forecasting the number of passengers is crucial to predicting the passengers in 2022.

The researcher initially approached time series analysis using the Autoregressive Integrated Moving Average (ARIMA) method. However, due to drastic changes caused by the Covid-19 pandemic, the ARIMA method was carried out with intervention and compared with the Extreme Learning Machine (ELM) method. The results of this research are expected to help the management of Domine Eduard Osok Sorong Airport and the government, more specifically the Ministry of Civil Aviation, in adopting and evaluating policies during the Covid-19 pandemic and anticipating an increase in the number of passengers after the Covid-19 pandemic by preparing additional prayer facilities, check-in counters, airport lounges, parking lots, and other public facilities.

Several related studies include research by Alfiyatin *et al*. [5] using the ELM and backpropagation methods to predict Indonesia's inflation rate. Fransiska [6] forecasts monthly rainfall in Bengkulu City with the Seasonal Autoregressive Integrated Moving Average (SARIMA). Intervention modeling to analyze and forecast the number of airplane passengers at Soekarno-Hatta airport due to the Covid-19 pandemic was used by Rianda [4]. The results of this study found that the Covid-19 pandemic in Indonesia significantly impacted on reducing the number of airplane passengers at Soekarno-Hatta Airport by 43.48%. The analysis results produce a prediction of the Mean Absolute Percentage Error (MAPE) value of 7.79% on the training data and the MAPE value of 14.07% on the testing data.

The research conducted by Bayu Galih Prianda and Edy Widodo [7] used the Seasonal ARIMA and ELM Methods in forecasting the number of foreign tourists to Bali. The error value is 4.97%, and with the ELM method, an error value of 7.62% is obtained. The research by Syifania Putri forecast the Number of Departures of Domestic Sailing Passengers at the Port of Tanjung Perak Using the ARIMA and SARIMA Methods. The results of this study show that the analysis using the ARIMA method has a lower accuracy value than the SARIMA method, which is 16.15% [8].

2. RESEARCH METHODOLOGY

2.1 Autoregressive Integrated Moving Average (ARIMA)

The data in this study is the number of passengers departing at Domine Eduard Osok Sorong Airport, Indonesia. The data source comes from the airport from January 2017 to May 2022. The ARIMA model is used on stationary data or data experiencing differentiation to meet stationary. The ARIMA(p,d,q) model is a combination of the ARMA(p,q) model, and the differentiation equation is stated in the following equation [9][10].

$$X_t(1-B)(1-\phi_1 B)=\mu'+(1-\theta_1 B)e_t \tag{1}$$

where X_t: the value of the t-period response variable, $(1-\phi 1)$: the AR value, $(1-\theta_1 B)$: the MA value, and e_t: forecast error (error). The analysis steps using the ARIMA method (Nurjanah et al., 2018) are: (1) identifying the model and carry out the stationarity test, (2) forming possible ARIMA models, (3) estimating parameters and test the significance of parameters, (4) selecting the ARIMA model based on the smallest AIC value, (5) conducting a residual assumption test.

2.2 Seasonal Autoregressive Integrated Moving Average (SARIMA)

Time series data often have seasonal patterns that form specific interval patterns (months, weeks, and others). One way to represent data like this is to assume that the model has two components, is stated in the following equation.

$$X_t = S_t + N_t \tag{2}$$

where X_t is a component with a seasonal factor S_t, and $N_t$is a stochastic component which may be an ARMA model (Tantika et al., 2018), the general form of SARIMA (p,d,q) (P,Q,S) is stated in the following equation.

$$X_t(1-B)^d(1-B^S) = (1-B\Phi_1)(1-\theta_1 B^S)e_t \tag{3}$$

with $(1-B)^d$: non-seasonal differentiator, $(1-B^S)$: seasonal differentiator, $\theta_1 B$: MA *non seasonal*, B^S: MA *seasonal*, *et*: residual term. The steps in the SARIMA method are (Kafara et al., 2017): (1) model identification and stationarity test, (2) parameter estimation and significance test carried out on the parameters, (3) selection of several possibilities SARIMA models based on the lowest AIC value, (4) model diagnostic tests, and (5) determining the SARIMA model for forecasting.

2.3 Modeling Procedure

The first step in constructing an ARIMA model is plotting the time series data. The plot shows the pattern of time series data, which can be horizontal, trend, cyclical, or seasonal. Making time series data plots aims to investigate the stationarity of time series data. The stationarity of time series data is the first thing that must be considered because the AR and MA aspects of the ARIMA model are only related to time series data, which are stationary in variance and mean. The second step is Parameter estimation. Maximum likelihood estimation is used in parameter estimation. This method uses the principle of maximizing the likelihood function to estimate the parameters θ and ϕ in the ARIMA model. The ARIMA (p,q) model is stated in the following equation.

$$Z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots + \phi_p z_{t-p} + \alpha_t - \theta_1 \alpha_{t-1} - \theta_2 \alpha_{t-2} - \cdots - \theta_q \alpha_{t-q} \tag{4}$$

with Zt = Zt -μ and $\{\alpha_t\} \sim N\ (0,\sigma_\alpha^2)$, the probability density function of α = α1, α2,…, α_n is stated in the following equation.

$$P(\Phi,\mu,\sigma_{\varepsilon^2}) = (2\pi\sigma_{\varepsilon^2})^{\frac{n}{2}}\, exp\ exp\, [-\frac{1}{2\pi\sigma_{\varepsilon^2}}\sum_{t=1}^{n}\varepsilon_{t^2}] \tag{5}$$

Next is the function of the parameter $(\Phi,\sigma_{\varepsilon^2})$

$$nL(\Phi,\mu,\sigma_\alpha^2) = -\frac{n}{2}\ln \ln 2\pi\sigma_\alpha^2 - \frac{s(\Phi,\mu,\theta\ )}{2\sigma_\alpha^2} \tag{6}$$

with

$$s(\Phi,\mu,\theta) = \sum_{t=1}^{n}\alpha_t^2\,(\Phi,\mu,\theta|Z_*,\alpha_*,Z)$$

$s(\Phi,\mu,\theta)$ is the sum square function. The estimator value $\phi,\hat{\mu}$, and θ is obtained when maximizing the maximum likelihood estimator equation. After obtaining the estimator value, it can also be calculated from σ_{ε^2} from $\sigma_{\varepsilon^2} = \frac{s(\hat{\phi},\hat{\mu},\hat{\theta})}{df}$, with $df = (n-p)\ -\ (p+q+1) = n\ -\ (2p+q+1)$ (Tantika et al., 2018).

The third step is a diagnosis of the provisional ARIMA model to prove that the provisional model that has been set is sufficient. The next step is model selection, and the model selection criteria are based on the AIC value and the parsimony principle. The smallest value is used in selecting the best model. At the same time, the principle of parsimony is a criterion for selecting the best model by choosing a simpler *AR*(*p*) or *MA*(*q*) order value. Finally, forecasting using the best ARIMA model before intervention.

2.4 Intervention Analysis

Time series data can be affected by external events that can cause changes in time series data patterns. External events are called 'intervention,' such as natural disasters, government policies, promotions, wars, holidays, and others. A method is needed to model time series data and describe response patterns from existing interventions. The method that can be used is intervention analysis [11]. According to Box [12], the intervention event is assumed to occur at a known time point T and causes a change in the time series data pattern. Generally, intervention events impact two conditions: intervention events have a temporary impact and occur only within a specific time (pulse function), and intervention events have a long-term impact (step function). In general, there are two intervention models: The pulse function and the Step function.

The form of the intervention model or the intervention ARIMA model is presented by Box [12] as follows:

$$Y_t = \frac{\omega_s(B)B^b}{\delta_r(B)} I_t + \frac{\theta(B)}{\phi(B)(1-B)^d}\varepsilon_t \tag{7}$$

where Y_t: response variable at t, I_t: intervention variable, B: time delay or start time of intervention effect, $\omega_s(B)$: $\omega_0 - \omega_1 B - \cdots - \omega_s B^s$ (s indicates the length of time to stabilize), $\delta_r(B)$: $1 - \delta_1(B) - \cdots - \delta_r B^r$ (r pattern of intervention effects occurring since the intervention event at time T), $\frac{\theta(B)}{\phi(B)(1-B)^d}\varepsilon_t$: ARIMA model without intervention effect [13].

2.5 Extreme Learning Machine

Extreme learning machine (ELM) is an algorithm for Single Hidden Layer Feedforward Neural Network (SLFN) that converges much faster than traditional methods because it does not require iteration [14]. SLFN consists of three layers of neurons: input, hidden layer, and output. The name Single refers to one layer of the non-linear neuron model, which is a hidden layer. The input layer offers data features but does not perform any calculations, while the output layer is linear with no transformation function and no bias. The ELM method is the development of a feedforward artificial neural network. The algorithm of the ELM method does not train input weights or bias, but the ELM method is used to obtain output weights using the norm-least-squares solution and moore-penrose inverse in general linear systems. ELM has a fast-learning speed and produces good generalization performance by finding nodes that provide maximum output values. Parameters, such as input weight and bias, are chosen randomly [15].

Huang *et al.* [16] state three stages in the ELM method, including training data, activation function *g(x)*, and *m* hidden unit, then: (1) determining the input weight vector W_j and bias b_j the influence factor for the j-th hidden unit, b_j, $j = 1, \ldots, m$, (2) calculating the output matrix in the hidden layer H_{nxm}, and (3) calculating the output weight β. The neural network model of the ELM is presented in Figure 1.

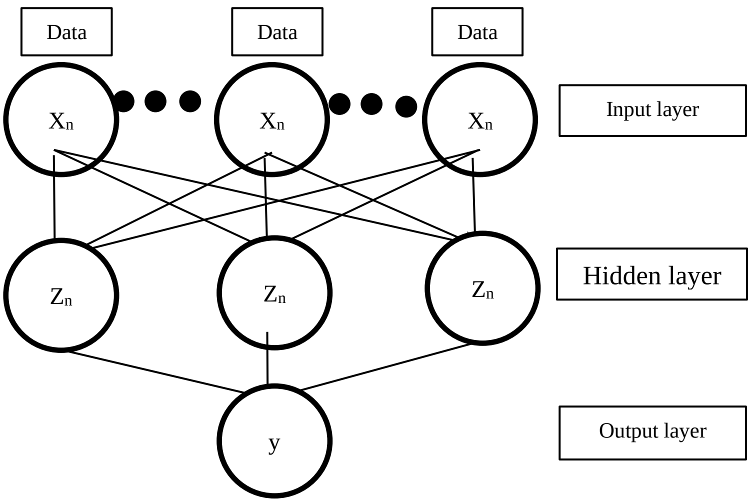

Figure 1. Neural Network Model of the Extreme Learning Machine

Based on **Figure 1** the ELM technique assigns the input layer weight *w*, and bias *b* randomly and never adjusts accordingly. Since the input weights (w) are fixed, the output weights (β) do not depend on them and have an immediate solution that does not require iteration. The training data is $S = \{(x_i, t_i) | x_i = (x_{i1}, x_{i2}, \ldots, x_{in})T \in R^n, t_i = (t_{i1}, ti2, \ldots, t_{im})T \in R^m\}$ where x_i is the input vector, and t_i is the target vector. Output o from ELM with $\widehat{N}$ hidden neurons is as follows [6], [13], [17]:

$$\sum_{i=1}^{\tilde{N}} \beta_i g(w_i x_j + b_i) = o_j,\ j = 1,\ 2,\ \ldots N \tag{8}$$

$g(\cdot)$ is the activation function on the hidden layer. The notation t_i indicates the target on the actual data, while the output (o) is the predicted NN data. The activation function in ELM is a non-linear function to provide a non-linear mapping of the system. Several activation functions are widely used, including the sigmoid, radial basis function, and hyperbolic tangent function. The training aims to minimize the error between the target and the ELM output. The most used object function is Mean Squared Error (MSE) [16].

$$MSE = \sum_{i=1}^{N} \left(t_{ij} - o_{ij}\right)^2,\ j = 1,\ 2,\ \ldots m \tag{9}$$

where N is the number of samples (training), $i,\ j$ are the indices for the sample data (training) and the output layer nodes. Sufficient (optimal) $w_i, b_i,\ \beta_i$ values should be found to predict the target value [16].

$$\sum_{i=1}^{\tilde{N}} \beta_i g(w_i x_j + b_i) = t_j,\ j = 1,\ 2,\ \ldots m \tag{10}$$

The formula is written in matrix form: $H\beta = T$. Therefore, training the SLFN is to find the best $w_i, b_i,\ \beta_i$. Determine β from $H\beta = T$ with inverse Moore Penrose so that $\beta = H^{+}T$, where H^{+} is a generalization matrix of Moore Penrose inverse H [14], [16].

2.6 Research Flow

Presents the flow of this research is presented in Figure 2.

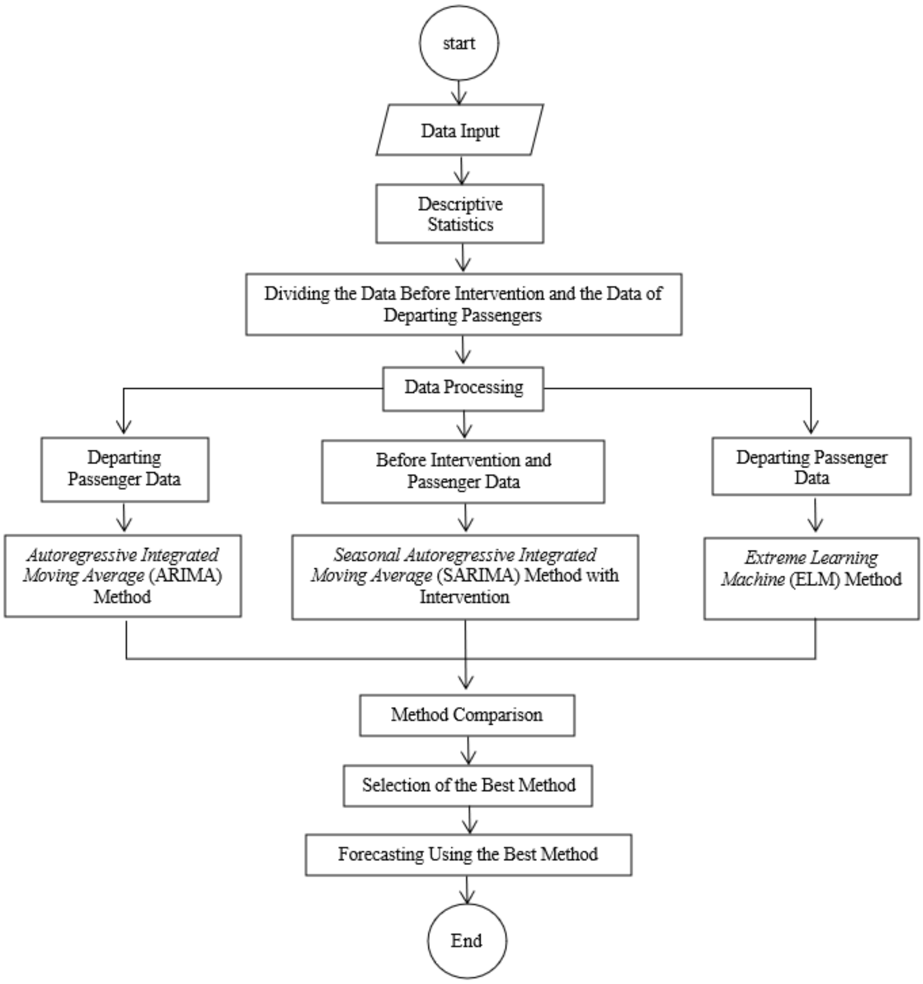

Figure 2. Research Stages Flowchart

Based on Figure 2, this research begins with descriptive statistics. Then, the researcher divides into the analysis based on the method, namely: (1) ARIMA, (1) SARIMA with Intervention, and (3) ELM. In the ARIMA Model, the process of parameter estimation and the best model of the ARIMA model is carried out. In the SARIMA method with intervention, the data is initially divided into 2, namely data before and after the intervention. In the data before the intervention, parameter estimation was carried out for the SARIMA model before the intervention. The results of SARIMA before the intervention was used to estimate the parameters of the SARIMA Intervention model.

Meanwhile, in the ELM method, the researcher carried out parameter iterations and took the smallest metric values, namely the Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) values. From the best ELM and the SARIMA method with intervention, a comparison was made to choose the best method. The results of the best method are used for future predictions.

3. RESULTS AND DISCUSSION

The time series plot on the number of passengers descrice in Figure 3.

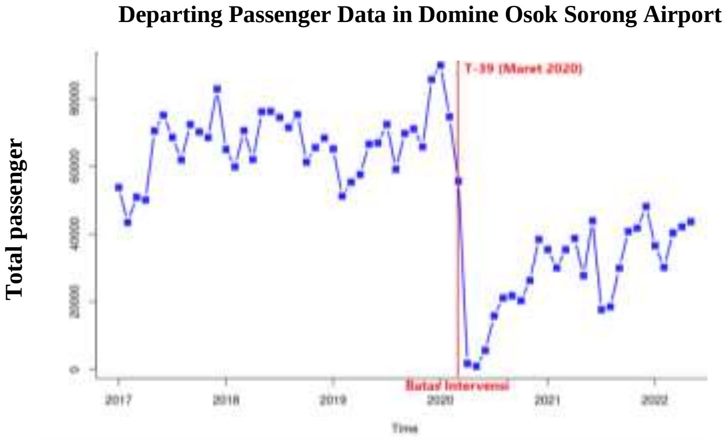

Figure 3. The plot of Passengers Departing January 2017 - May 2022

Figure 3 presents 65 data plots of passengers departing at the domestic departure terminal at Domine Eduard Osok Sorong Airport from January 2017 to May 2022. In 2020 the Covid-19 pandemic occurred, which resulted in a significant decrease in the number of passengers in April 2020 at Domine Eduard Osok Sorong Airport. Data at $t = 1, 2, \ldots, 39$ is data before the intervention, January 2017 to March 2020.

As a result of the Covid-19 pandemic, the government issued instruction to enforce public activity restrictions at villages/sub-districts (PPKM). Moreover, the government's policy that regulates travel by airplane during PPKM also caused a drastic decrease in the number of airplane passengers and flight activity in April 2020 at Domine Eduard Osok Airport in Sorong.

3.1 Autoregressive Integrated Moving Average (ARIMA)

3.1.1 Model identification and parameter estimation

The data used to build the ARIMA model is 65 data on the number of passengers departing at the domestic departure terminal at Domine Eduard Osok Sorong Airport from January 2017 to May 2022. The Augmented Dickey-Fuller test (unit root test) was used to test the stationarity of the data. The p-value of the Augmented Dickey-Fuller test with drift on passenger data is 0.01. Using $\alpha = 5\%$, the researchers reject H_0 (null hypothesis). Thus, the data does not contain unit roots (stationer). The researchers differentiated the data to develop another alternative model. Based on the smallest AIC value, the ARIMA (0,1,0) model is the best.

$$\begin{aligned} Y_t &= \hat{\phi}\, y_{t-1} + e_t \\ Y_t &= 0 + e_t \end{aligned} \tag{11}$$

3.1.2 ARIMA Model Diagnostic Test and Error Measure

The p-value of the Ljung-Box test on residual ARIMA (0,1,0) is 0.9851. By using $\alpha = 5\%$, the researchers fail to reject the null hypothesis. Thus, the residual ARIMA (0,1,0) has no serial correlation. The normality test for residuals using the Kolmogorov-Smirnov test has a p-value of 0.2893, more significant than $\alpha = 5\%$. The researchers failed to reject the null hypothesis on the normality test. Thus, the residuals are normally distributed. The ARIMA (0,1,0) model has an RMSE of 11,513.47 and a MAPE value of 65.47%.

3.2 Intervention Analysis

3.2.1 Seasonal ARIMA Modeling before Intervention

a. Model Identification

The data used to form the ARIMA model before the intervention was 39 data on the number of passengers departing at the domestic departure terminal at Domine Eduard Osok Sorong Airport from January 2017 to March 2020. A check will be carried out using the Augmented Dickey-Fuller test (unit root test) to find out precisely the stationarity of the data.

Augmented Dickey-Fuller test (unit root test) was used to test the stationarity of the 39 data. The p-value of the Augmented Dickey-Fuller test is 0.004201. Using $\alpha = 5\%$, the researchers reject H_0 (null hypothesis). Thus, data on the number of passengers before the pandemic did not contain unit roots or stationary data. Researchers form several alternative models by incorporating seasonal elements. Based on the smallest AIC value, the best model is ARIMA (1,0,0) (0,1,0) [12]. The model only differentiates data on seasonal lag, and the order for auto-regressive is p=1.

b. ARIMA Seasonal Model Diagnostic Test

Based on the calculation results of the test statistics for the Ljung-Box test is P-value = 0.7699 > α = 5% (0.05), which means no serial correlation, and the Kolmogorov-Smirnov test is P-value = 0.2027 > α = 5% (0.05)), which means that the data is normally distributed; it shows that at $\alpha = 0.05$ fails to reject H_0, which means that the residual value for the ARIMA (1,0,0) (0,1,0) [12] model fulfills the white noise and normality assumptions. By the fulfillment of the white noise assumptions used by the Ljung-Box Test and normality used by the Kolmogorov-Smirnov Test, the ARIMA (1,0,0) (0,1,0) [12] model is adequate and can be used for forecasting.

c. ARIMA Seasonal Modeling Before Intervention

After obtaining the best Seasonal ARIMA model and fulfilling the assumptions for the Ljung-Box test and the Kolmogorov-Smirnov test, seasonal ARIMA modeling is obtained before systematic intervention as follows:

$$\begin{aligned} Y_t &= \widehat{\phi}\, y_{t-1} + e_t \\ Y_t &= 0.6472 y_{t-1} + e_t \end{aligned} \tag{12}$$

Based on the ARIMA (1,0,0) (0,1,0) [12] model equation, the RMSE accuracy value is 7270,498, and the MAPE value is 7,946344.

3.2.2 Intervention ARIMA Seasonal Analysis

a. Intervention Parameter Estimation

The next step is to estimate the known intervention parameters. Using the software R, the value of the parameter ω_0 is -0.00052990, and the parameter δ_0 is 0.

b. Intervention ARIMA Seasonal Model Diagnostic Test

Based on the calculation of the test statistics for the Ljung-Box test, P-value = 0.9298 > α = 5% (0.05), which means no serial correlation, and the Kolmogorov-Smirnov test, P-value = 0.1728 > α = 5% (0.05)), which means that the data is typically distributed. It shows that at $\alpha = 0.05$ fails to reject H_0, which means that the residual value for the Intervention ARIMA (1,0,0) (0,1,0) [12] model fulfills the white noise and normality assumptions. The fulfillment of the white noise assumptions using the Ljung-Box Test and normality using the Kolmogorov-Smirnov Test concludes that the Intervention ARIMA (1,0,0) (0,1,0) [12] model is adequate and can be used for forecasting.

c. Intervention ARIMA Seasonal Modeling

After obtaining the best Seasonal ARIMA model and fulfilling the test assumptions for the Ljung-Box test and the Kolmogorov-Smirnov test, the Intervention Seasonal ARIMA model equation is stated in the following equation

$$Y_t = \frac{\omega_s(B)B^b}{\delta_r(B)} I_t + \hat{\phi} y_{t-1} + e_t$$
$$Y_t = -52990.044 P_{40} + 0.6472 y_{t-1} + e_t \qquad (13)$$

Based on forecasting using the Intervention ARIMA (1,0,0) (0,1,0) [12] model, the RMSE accuracy value is 9027,671, and the MAPE value is 35,86186.

3.3 Extreme Learning Machine

In order to optimize the ELM model, we involved the parameter such as hidden nodes, type of estimation, and replication (ensemble). The hidden nodes are the number of nodes located in hidden layer. Type of estimation is methods used to estimate output layer weights. The input of replication is the number of networks to train, the result is the ensemble forecast. The values of parameters are presented in Table 1.

Table 1. Presents The Trials By Combining Several Parameters to Obtain Optimal ELM Parameters.

No	Parameter	Value
1	Hidden nodes	5, 10
2	Types	Lasso, ridge, linear model
3	Replication	10, 20

In addition to using these parameters, some parameters are fixed, i.e., the median aggregating the several NN models. The input lags used are the latest six lags considering ACF and PACF and the period for significant declines during the pandemic. This lag is maintained (not trimmed) until the end of modeling. Because seasonality is challenging to identify, the ELM model does not accommodate seasonality. Table 2 shows the five combinations with the lowest MAPE values.

Table 2. Results of the Extreme Machine Learning Method Combination

No	*Hidden layer*	Type	Replication	MSE	RMSE	MAPE
1	20	lm	10	62607013.57	7912.46	30.64%
2	20	lm	20	61916445.11	7868.70	30.81%
3	20	lm	5	64356338.36	8022.24	31.18%
4	10	lm	10	94651843.57	9728.92	52.65%
5	10	lm	20	93054698.71	9646.49	53.39%

Based on Table 2, the smallest error measurement value is obtained by a combination of 20 hidden layers, weight optimization with a linear model, and 10 replications. The network architecture with this model in the first iteration is as follows:

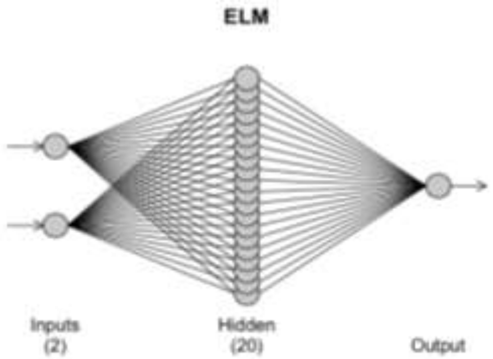

Figure 4. Plot Extreme Machine Learning

The model is constructed using the latest 6 lags on input. However, based on **Figure 4**, only 2 lags are kept in the final process (4 lags are trimmed). In this model, there are 20 nodes in the hidden layer; then, there is 1 output as forecasting data. A comparison of the error measurement values obtained from the ARIMA, Intervention Analysis, and ELM methods is presented in **Table 3**.

Table 3. Comparison of ARIMA Methods, Intervention Analysis, and Extreme Machine Learning

Analysis Method	RMSE Value	MAPE Value (%)
ARIMA	11513.47	65.47001
Intervention Analysis	9027.671	35.86186
Extreme Learning Machine	7912.46	30.64

Table 3 shows that the error measurement results using the ARIMA method are 65.47%, ARIMA Intervention is 35.86%, and ELM is 30.64%. Among these methods, the ELM method has smallest error measurement results. Therefore, the model is considered responsive to the significant change of pandemic. Thus, the ELM method was chosen as a suitable method for forecasting the number of airplane passengers at Domine Eduard Osok Sorong Airport. The forecast from ELM method is presented in **Figure 5**.

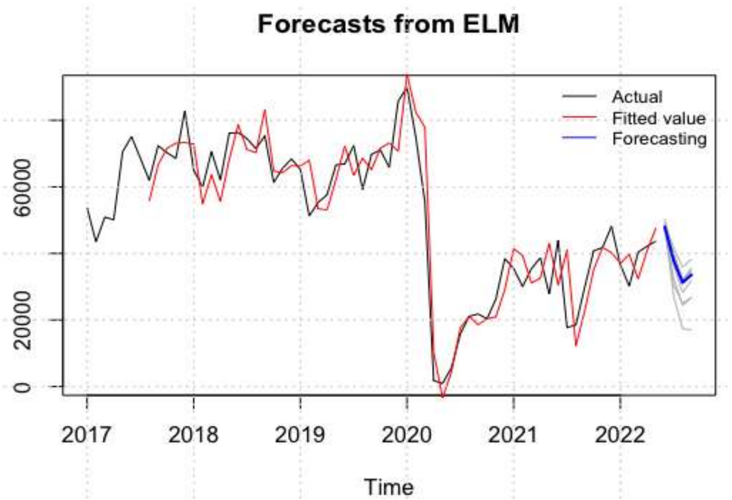

Figure 5. Plot for forecasting the number of airplane passengers using the ELM method

Figure 5 presents that the black line is the plot line for the number of airplane passengers at Domine Eduard Osok Sorong Airport from January 2017 to May 2022. The red line is the plot line for the fitted value overshadowing the actual data because there is a gap for the model to follow the actual data pattern. The blue lines are forecasting data plot lines with predicted values from June to September 2022 as follows: 47985, 37821, 31247, and 33578, respectively.

4. CONCLUSION

Based on the discussion that has been explained, the conclusions obtained in this study are as follows.

1. Based on the three methods used in forecasting the number of departing passengers at the domestic departure terminal at Domine Eduard Osok Sorong Airport, the ARIMA, ARIMA intervention, and ELM methods, the MAPE error measurement values were 65.47%, 35.86%, and 30.64%, respectively. The ELM method has the lowest error measurement results. Even though it does not divide data before and after the pandemic, the ELM model is quite responsive when significant trend changes occur. In this case, the ELM method was chosen to forecast the number of airplane passengers at Domine Eduard Osok Sorong Airport.
2. The MAPE value between the ARIMA intervention and ELM is not significant. Meanwhile, the value of MAPE between ARIMA and ARIMA intervention is quite significant. It proves that the intervention successfully reduced the level of forecasting errors appropriately. Future research can add a hidden layer to the neural network to address significant changes to the data. In addition, data pre-processing of transformation or normalization can be applied to reduce the error value.

AKNOWLEDGEMENT

The author thanks to Statistics Department, Faculty of Mathematics and Natural Science, Universitas Islam Indonesia which has assisted in the research process.

REFERENCES

[1] H. Bonakdari, J. P. Pelletier, and J. Martel-Pelletier, “A reliable time-series method for predicting arthritic disease outcomes: New step from regression toward a nonlinear artificial intelligence method,” *Comput. Methods Programs Biomed.*, vol. 189, p. 105315, 2020.

[2] H. Hartati, “Penggunaan metode ARIMA dalam meramal pergerakan Inflasi,” *J. Mat. Sains dan Teknol.*, vol. 18, no. 1, pp. 1–10, 2017.

[3] C. Sulistyowati, E. Farihah, and O. S. Hartadinata, *Anggaran perusahaan: teori dan praktika*. Scopindo Media Pustaka, 2020.

[4] H. N. Tantika, N. Supriadi, and D. Anggraini, “Metode seasonal ARIMA untuk meramalkan produksi kopi dengan indikator curah hujan menggunakan aplikasi R di kabupaten Lampung Barat,” *Matematika*, vol. 17, no. 2, pp. 49–58, 2018.

[5] A. N. Alfiyatin, W. F. Mahmudy, C. F. Ananda, and Y. P. Anggodo, “Penerapan Extreme Learning Machine (ELM) untuk peramalan laju inflasi di Indonesia,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 2, pp. 179–186, 2018.

[6] H. Fransiska, P. Novianti, and D. Agustina, “Permodelan curah hujan bulanan di kota Bengkulu dengan Seasonal Autoregressive Integrated Moving Average (SARIMA),” *Semin. Nas. Off. Stat.*, vol. 2019, no. 1, pp. 390–395, 2020.

[7] B. G. Prianda and E. Widodo, “Perbandingan metode seasonal ARIMA dan Extreme Learning Machine pada peramalan jumlah wisatawan mancanegara ke Bali,” *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 15, no. 4, pp. 639–650, 2021.

[8] S. Putri and A. Sofro, “Peramalan jumlah keberangkatan penumpang pelayaran dalam negeri di pelabuhan Tanjung Perak menggunakan metode ARIMA dan SARIMA,” *Stabilitas Sist. Din. Pertumbuhan Sel Kanker Dengan Ter. Radiasi*, vol. 10, no. 1, pp. 61–67, 2022.

[9] A. H. Primandari and M. D. Kartikasari, *Analisis Runtun Waktu dengan R*, 1st ed. Yogyakarta: Universitas Islam Indonesia, 2020.

[10] I. S. Nurjanah, D. Ruhiat, and D. Andiani, “Implementasi model Autoregressive Integrated Moving Average (ARIMA) untuk peramalan jumlah penumpang Kereta Api di pulau Sumatera,” *TEOREMA Teor. dan Ris. Mat.*, vol. 3, no. 2, p. 145, 2018.

[11] H. Panjaitan, A. Prahutama, and S. Sudarno, “Peramalan jumlah penumpang kereta api menggunakan metode ARIMA, intervensi dan ARFIMA (studi kasus : penumpang kereta api kelas lokal ekonomi DAOP IV Semarang),” *J. Gaussian*, vol. 7, no. 1, pp. 96–109, 2018.

[12] G. E. . Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*, 5th ed. New Jersey: Wiley Series in Probability and Statistics, 2015.

[13] A. Crystine, A. Hoyyi, and D. Safitri, “Analisis intervensi fungsi step (studi kasus pada jumlah pengiriman benda pos ke Semarang pada tahun 2006-2011),” *Concept Commun.*, vol. null, no. 23, pp. 301–316, 2019.

[14] J. M. Park and J. H. Kim, “Online recurrent extreme learning machine and its application to time-series prediction,” *Proc. Int. Jt. Conf. Neural Networks*, vol. 2017-May, pp. 1983–1990, 2017.

[15] Q. Humaini, “Jaringan syaraf tiruan Extreme Learning Machine (ELM) untuk memprediksi kondisi cuaca,” *Skripsi*, pp. 1–86, 2015.

[16] G. Huang, G. Bin Huang, S. Song, and K. You, “Trends in extreme learning machines: A review,” *Neural Networks*, vol. 61, pp. 32–48, 2015.

[17] Z. Kafara, F. Y. Rumlawang, and L. J. Sinay, “Peramalan curah hujan dengan pendekatan Seasonal Autoregressive Integrated Moving Average (SARIMA),” vol. 11, pp. 63–74, 2017.