Hybrid Forecasting Model To Predict Air Passenger and Cargo in Indonesia

Ratna Sulistyowati, *Suhartono, Heri Kuswanto, Setiawan
Department of Statistics,
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
ratnasuli@bps.go.id, suhartono@statistika.its.ac.id,
heri_k@statistika.its.ac.id, setiawan@statistika.its.ac.id
*Corresponding Author

Erni Tri Astuti
Department of Statistics,
Sekolah Tinggi Ilmu Statistik,
Jakarta, Indonesia
erni@stis.ac.id

***Abstract*— Forecasting of air passenger and cargo have a major influence on the master plan of the airport infrastructure development and investment by the civil airline. This research aims to obtain the most accurate predictive value of the air passenger and cargo at three international airports Indonesia, namely, Soekarno Hatta, I Gusti Ngurah Rai, and Juanda Airport. Those international airports are the three largest contributors to the number of air passengers and cargo volumes in Indonesia. This research uses a hybrid forecasting method that combines linear and nonlinear models. The combination of two linear and nonlinear models is able to obtain accurate predictions. The first phase is linear modeling with time series regression model (TSR) and Autoregressive Integrated Moving Average with Exogenous Factor (ARIMAX). In the second phase, the error of the linear model is analyzed by using machine learning methods such as Neural Network (NN) and Support Vector Regression (SVR) to capture nonlinear patterns. There are four hybrid models that be applied and compared, i.e. TSR-NN, TSR-SVR, ARIMAX-NN, and ARIMAX-SVR based on the Mean Absolute Percentage Error (MAPE). The results show that hybrid ARIMAX-NN and TSR-NN give more accurate prediction than hybrid TSR-SVR and ARIMAX-SVR.**

Keywords: Air passenger, Cargo, Time Series Regression, ARIMAX, Neural Networks, Support Vector Regression, Hybrid

I. INTRODUCTION

In recent years, the demand for global air transport services has continued to rise in almost all countries as reflected in the increasing frequency of flights, the number of air passengers and the volume of cargo. Biederman [1] suggests that long-term forecasts for air passenger and cargo are the key variable in airports related infrastructure projects, as well as investments by civilian airlines. Research on air traffic has been carried out in almost all countries. Several studies have developed an appropriate model to predict the air traffic flow [2-3]. Forecasting using Autoregressive Integrated Moving Average (ARIMA) model for air passenger and cargo data has been widely used as in [4-5].

Air passenger and cargo movements often implicitly have a nonlinear pattern, which appears from irregular seasonal and trend. The nonlinear model is more representative of the existing data condition. The most implemented in nonlinear time series forecasting is the machine learning approach, such as Artificial Neural Network (ANN) and Support Vector Machine (SVM). Previous studies have reported that the ANN model can predict the air passengers accurately as in [6–8]. Moreover, the SVM model also provides a promising alternative to air traffic forecasting as in [9-10].

Based on the results of M3 competition [11], the combination of forecasting methods will improve the accuracy of forecasting. Zhang [12] has been proposed a hybrid methodology using ARIMA for linear components and ANN for nonlinear components. In addition, ARIMA-ANN model can improve forecasting performance compared to individual models. Some studies that applied hybrid methodology can be seen in [13–15].

The need for air transportation services in Indonesia is also expected to grow in line with the economic and population growth. According to the Ministry of Transportation Republic of Indonesia [16], Indonesia currently has 289 airports with 24 international airports. Three international airports with the largest contribution in Indonesia, namely, Soekarno Hatta, I Gusti Ngurah Rai, and Juanda Airport. That three international airport contributes 55.52 percent of the number of air passengers departing from Indonesia and 68.50 percent of the total volume of cargo loading from Indonesia.

This paper develops four hybrid models of TSR-NN, TSR-SVR, ARIMAX-NN, and ARIMAX-SVR to obtain the most accurate forecasts for air passenger and cargo at the three largest international airports in Indonesia. The results of this studies can be used as input for government policies related to the planning and development of airport infrastructure.

II. METHODS

A. Time Series Regression (TSR) Model

In general, TSR model is basically the same as regression. The TSR model was used in this study is a model with the component of trend, seasonal and calendar variation effects. The trend is defined as the long-term direction that is continuously up or down [17]. Seasonality is a repeating pattern with the same period, for example, 12 months per year. Meanwhile, calendar variations are seasonal patterns with varying periods. The composition of the day varies from month to month and year to year such as Eid Fitri. The TSR model with trend, seasonality, and calendar variation can be described as:

$$Y_t = \delta t + \sum_{m=1}^{M} \beta_m S_{m,t} + \sum_{g=1}^{G} \gamma_g V_{g,t} + \sum_{g=1}^{G} \varphi_g V_{g,t-1} + \sum_{g=1}^{G} \vartheta_g V_{g,t+1} + N_t \quad (1)$$

where δ is a linear trend parameter, β is the seasonal parameter, and $S_{m,t}$ is a seasonal dummy variable. If the data are monthly, then $M = 12$. If the data are quarterly data, then $M = 4$, and so on. $V_{g,t}$ is a dummy variable of the effects of calendar variations, $V_{g,t-1}$ is a dummy variable of one month before the occurrence of calendar variation effects, and $V_{g,t+1}$ is a dummy variable of one month after the occurrence of calendar variation effects. If the effects of calendar variations made in weekly, then $G=4$. If the effects of calendar variations made in daily, then $G=30$, and so on. Total effects of calendar variations can be identified based on a plot time series. N_t is a white noise error. If the error is not white noise, lag is used as an additional predictor variable. Selection of lag can be determined based on a plot of ACF and PACF [18].

B. Autoregressive Integrated Moving Average with Exogeneous (ARIMAX) Model

ARIMAX model is the development of ARIMA model to include significant exogenous factors that are considered in the model. ARIMA model is a linear combination forecasting model from the autoregressive model (AR) and the moving average model (MA). ARIMAX model used in this paper is a model with a trend, seasonal and calendar variations which generally can be written:

$$Y_t = \delta t + \sum_{m=1}^{M} \beta_m S_{m,t} + \sum_{g=1}^{G} \gamma_g V_{g,t} + \sum_{g=1}^{G} \varphi_g V_{g,t-1} + \sum_{g=1}^{G} \vartheta_g V_{g,t+1} + \frac{\theta_q(B)\Theta_Q(B^S)}{\phi_p(B)\Phi_P(B^S)} N_t \quad (2)$$

The first stages of ARIMAX model building are creating a time series regression models with the trend, seasonal and calendar variations, in order to obtain the error [18]. The error is modeled with ARIMA (Box-Jenkins procedure). Order of the ARIMA model used for the original data and the input variables simultaneously in equation (2). Then the next stages are checking the residual white noise and normal distribution, and test the significance parameters.

C. Neural Network (NN) Model

Artificial Neural Network (ANN) or known as Neural Network (NN), originally developed to mimic the human brain works, are composed of a number of interconnected simple processing elements called neurons or nodes [19]. Feed Forward Neural Network (FFNN) is the most common form of NN architecture used in time series modeling. FFNN consists of one input layer, one or more hidden layers, and one output layer. The NN model specification in selecting the number of hidden neurons used by cross-validation method [20]. The FFNN architecture with p neuron in the input layer, q neurons in the hidden layer and one neuron in the output layer is shown in Figure 1.

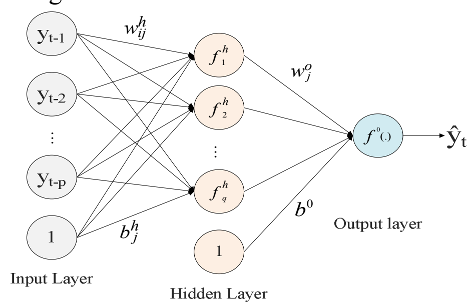

Figure 1 Architecture FFNN

The architecture describes the output value $(\hat{y}_t)$ and the inputs $(y_{t-1}, y_{t-2}, \ldots, y_{t-p})$ with mathematical formulated as:

$$\hat{y}_t = f^0 \left[\sum_{j=1}^{q} \left\{ w_j^0 f_j^h \left[\sum_{i=1}^{p} w_{ij}^h y_{t-i} + b_j^h \right] + b^0 \right\} \right], \quad (3)$$

where w_{ij}^h is the weight of the i th neuron input layer to the j th neuron hidden layer, b_j^h is the bias of the j th neuron in the hidden layer $(j = 1, 2, \ldots, q)$, w_j^0 is the weight of the j th neuron from hidden layer to the neuron in the output layer, and b^0 is the bias of the neuron in the output layer. f_j^h is the activation function in the hidden layer using a sigmoid function, i.e. $f(x) = (1+\exp(-x))^{-1}$. f^0 is the activation function in the output layer with the linear function $f(x) = x$.

D. Support Vector Regression (SVR) Model

SVM was originally developed by Cortes and Vapnik [21] to solve problems in the development of SVM classification but also overcomes the problem of regression called the SVR. SVR tries to obtain the best hyperplane using principle of structural risk minimization (SRM), by dividing data and minimizing the distance between hyperplane and data. The regression function of the SVR methods is expressed as:

$$f(x) = w^T \phi(x) + b, \quad (4)$$

where w is the weighting vector, $\phi(x)$ is a function that maps nonlinearly x from the input space into a high dimension features space and b is the bias. The coefficients of w and b are estimated by minimizing the risk function [21] described in the following equation:

$$R(f(x)) = C \sum_{i=1}^{T} L_\varepsilon(y_i, f(x_i)) + \frac{1}{2}\|w\|^2, \quad (5)$$

where $L\varepsilon(y_i, f(x_i)) = \begin{cases} 0, & \text{if} \left|y_i - f(x_i)\right| \le \varepsilon \\ \left|y_i - f(x_i)\right| - \varepsilon, & \text{Otherwise} \end{cases}$

with L_ε is ε-insensitive loss function, C and ε are the parameters that have been determined to obtain optimum global results. The concept of the loss function is to minimize the value as follows:

$$R(w, \xi, \xi^*) = \min \frac{1}{2}\|w\|^2 + \mathrm{C} \sum_{i=1}^{n} \left(\xi_i + \xi_i^*\right), \quad (6)$$

with limits: $y_i - w^T\phi(x_i) - b < \varepsilon + \xi_i$,

$w^T\phi(x_i) - b < \varepsilon + \xi_i$,

$w^T\phi(x_i) + b - y_i < \varepsilon + \xi_i^*$, and

$\xi_i^*, \xi_i > 0$, $i = 1, 2, \ldots, n$.

Optimizing on such limitations can be solved using *Lagrangian* in the following form:

$$\begin{aligned} L(w,b,\xi,\xi^*,\alpha_i,\alpha_i^*,\beta_i,\beta_i^*) &= \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{n}\left(\xi_i+\xi_i^*\right)\right) \\ &- \sum_{i=1}^{n} \alpha_i\left[w\varphi(x_i)+b-y_i+\varepsilon+\xi_i^*\right] \\ &- \sum_{i=1}^{n} \alpha_i^*\left[y_i - w\varphi(x_i) - b + \varepsilon + \xi_i^*\right] - \sum_{i=1}^{n}\left(\beta_i\xi_i + \beta_i^*\xi_i^*\right), \end{aligned} \quad (7)$$

with the approach Karush-Kuhn-Tuck will be obtained:

$$\begin{aligned} \partial(\alpha_i, \alpha_i^*) &= \sum_{i=1}^{n} y_i(\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^{n} (\alpha_i + \alpha_i^*) \\ &- \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j). \end{aligned} \quad (8)$$

The kernel function $K(x_i, x_j)$ can be expressed as inner product $\phi(x_i)^T \phi(x)$. One of the kernel function is most often used is the *Gaussian radial basis function* (RBF) [22] as follows:

$$K(x_i, x) = \exp\left(-\frac{\|x_i - x\|^2}{2\sigma^2}\right), \tag{9}$$

where σ^2 is a kernel parameter. The SVR function express as:

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) K(x_i, x) + b. \tag{10}$$

In forecasting time series data, the inputs used are the lag of the observational data $x = [y_{t-1}, y_{t-2}, \ldots, y_{t-p}]$. SVR architecture implemented with FFNN model approach is shown in Figure 2.

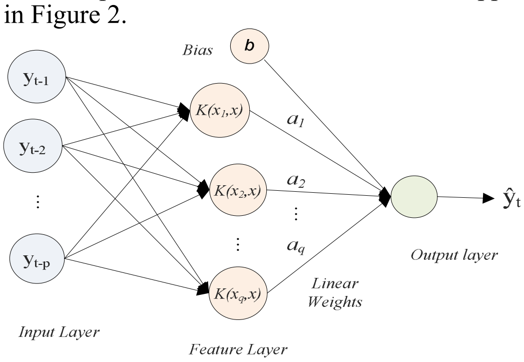

Figure 2 Architecture SVR with FFNN Approach

E. Hybrid Forecasting Model

Zhang [16] represented a hybrid forecasting model consisting of linear and nonlinear models formulated as:

$$Y_t = L_t + N_t + e_t \tag{11}$$

where L_t is the linear component and N_t is the nonlinear component of the hybrid model. This paper compares four hybrid forecasting models i.e. TSR-NN, TSR-SVR, ARIMAX-NN, and ARIMAX-SVR. The flowchart of the four hybrid models is present in Figure 3.

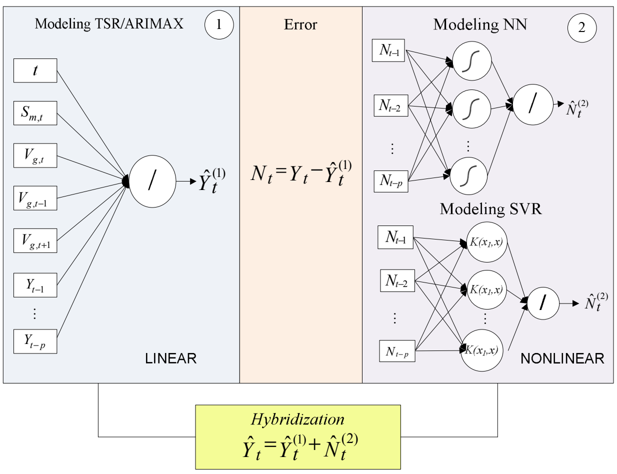

Figure 3 Flowchart Modeling Hybrid TSR-NN, TSR-SVR, ARIMAX-NN and ARIMAX-SVR

F. Best Model Selection Criteria

The best model is chosen by evaluating the forecast accuracy. Criteria for selecting the best model is Mean Absolute Percentage Error (MAPE) in testing data. MAPE has the advantage of being scale-independent, and so are frequently used to compare forecast performance between different data sets [17]. MAPE is formulated by:

$$MAPE = \left(\frac{1}{L}\sum_{i=1}^{L}\left[\left|\frac{Y_{n+l} - \hat{Y}_n(l)}{Y_{n+l}}\right|\right]\right)100\%, \tag{12}$$

where n is the number of training data, L is the number of testing data, Y_{n+l} is the actual data and $\hat{Y}_n(l)$ is the forecast of Y_{n+l}. Measures based on MAPE has the disadvantage when there is any observation being infinite or undefined, or having an extreme value close to zero [17]. MAPE also has the disadvantage that put a heavier penalty on negative errors than on positive errors. In this study, there were no observations with extreme values near zero and had symmetric errors.

III. DATA

The research data in this paper exists monthly data on the number of air passenger and cargo volume at the three international airports in Indonesia namely Soekarno Hatta (CGK), I Gusti Ngurah Rai (DPS), and Juanda Airport (SUB). These data are periodically published by the Statistics Indonesia (BPS), that is collected from PT Angkasa Pura 1 and PT Angkasa Pura 2. The time period of the data is in the range from January 2001 to August 2017. The research variables are described in Table 1.

TABLE 1 THE RESEARCH VARIABLES

Passenger (in persons)		**Cargo (in tons)**	
Variable	**Description**	**Variable**	**Description**
Domestic Passenger		**Domestic Cargo**	
$Y_{1,1,t}$	Departed from CGK	$Z_{1,1,t}$	Loaded from CGK
$Y_{1,2,t}$	Departed from DPS	$Z_{1,2,t}$	Loaded from DPS
$Y_{1,3,t}$	Departed from SUB	$Z_{1,3,t}$	Loaded from SUB
$Y_{2,1,t}$	Arrived at CGK	$Z_{2,1,t}$	Unloaded at CGK
$Y_{2,2,t}$	Arrived at DPS	$Z_{2,2,t}$	Unloaded at DPS
$Y_{2,3,t}$	Arrived at SUB	$Z_{2,3,t}$	Unloaded at SUB
International Passenger		**International Cargo**	
$Y_{3,1,t}$	Departed from CGK	$Z_{3,1,t}$	Loaded from CGK
$Y_{3,2,t}$	Departed from DPS	$Z_{3,2,t}$	Loaded from DPS
$Y_{3,3,t}$	Departed from SUB	$Z_{3,3,t}$	Loaded from SUB
$Y_{4,1,t}$	Arrived at CGK	$Z_{4,1,t}$	Unloaded at CGK
$Y_{4,2,t}$	Arrived at DPS	$Z_{4,2,t}$	Unloaded at DPS
$Y_{4,3,t}$	Arrived at SUB	$Z_{4,3,t}$	Unloaded at SUB

The predictors used are trends, seasonal and calendar variations dummy variable as shown in Table 2.

TABLE 2 THE PREDICTORS

Dummy	**Description**
Trends	$t = 1, 2, \ldots, n$
Seasonal	$S_{1,t} = \begin{cases} 1, & \text{for January} \\ 0, & \text{otherwise} \end{cases} \quad \ldots \quad S_{12,t} = \begin{cases} 1, & \text{for December} \\ 0, & \text{otherwise} \end{cases}$
Calendar Variation	$V_{i,t} = \begin{cases} 1, & \text{for Eid Fitri at week } i\text{-th month } t\text{-th, with } i=1,2,3,4 \\ 0, & \text{Otherwise} \end{cases}$
	$V_{i,t-1} = \begin{cases} 1, & \text{for month } t\text{-th before Eid Fitri at week } i\text{-th, with } i=1,2,3,4 \\ 0, & \text{Otherwise} \end{cases}$
	$V_{i,t+1} = \begin{cases} 1, & \text{for month } t\text{-th after Eid Fitri at week } i\text{-th, with } i=1,2,3,4 \\ 0, & \text{Otherwise} \end{cases}$

IV. RESULTS AND DISCUSSION

A. Characteristic of Air Passenger and Cargo Data

The growth of the total number of air passengers and cargo volume in the three largest international airports in Indonesia, namely Soekarno Hatta, I Gusti Ngurah Rai, and Juanda Airport is shown in Figure 4.

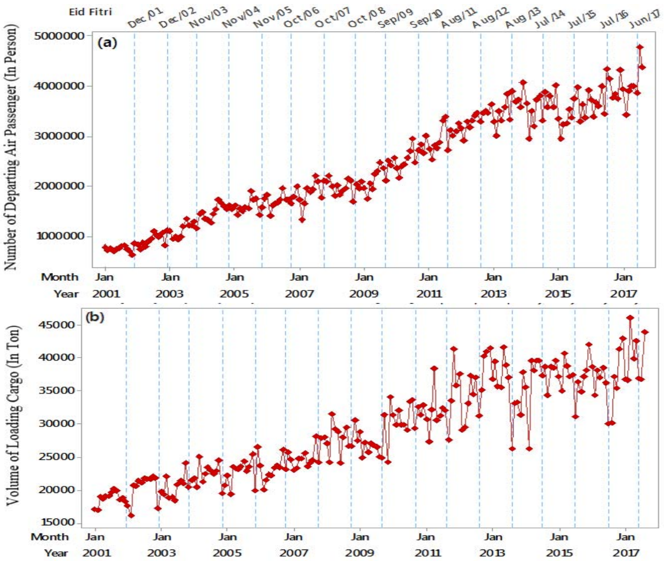

Figure 4 Time Series Plot of Total Number of Departing Air Passenger (a) and Loading Cargo Volume (b) at Three International Airports

Figure 4 indicates the increasing pattern of the total number of passengers and the total volume of cargo at the three airports from 2001 to 2017. Both graphs show the seasonal pattern with an increasing number of air passengers and cargo volume in July and December. July is the month in which students begin the new academic year, while December is the month for a celebration of Christmas and New Year.

The plot illustrates that Eid Fitri has an effect on increasing the number of air passengers and cargo. Figure 5 visually determines whether there is an influence of the week of Eid Fitri on air passenger in Eid Fitri month, one month before Eid Fitri, and one month after Eid Fitri.

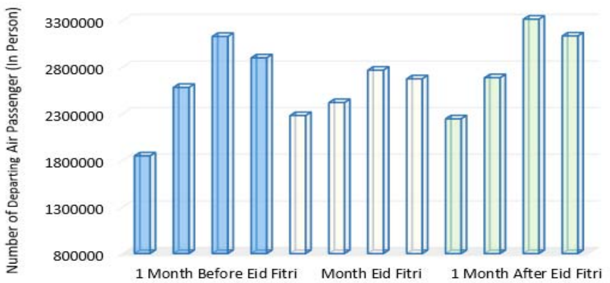

Figure 5 Week's Influence On Average Departing Passenger at Three Airport

Figure 5 explains that the effect of the Eid Fitri occurring in different weeks will affect different air passenger and cargo increases. For the data of the number of departing passengers, when Eid Fitri occurred in the first week then the average passenger is high during the month of Eid Fitri. However, if the Eid Fitri occurs in the second, third and fourth week, then the average passenger during Eid Fitri is not higher than one month before and one month after Eid Fitri. The effect of the week of Eid Fitri in three airports for cargo data is described in Figure 6.

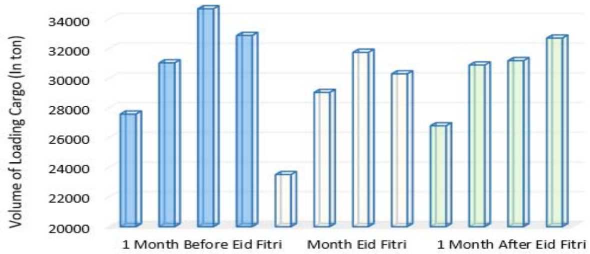

Figure 6 Week's Influence On Average Loading Cargo at Three Airport

For the cargo data in Figure 6, it is apparent that if Eid Fitri occurred in the first, second, third and fourth week, then the average cargo is high at one month before Eid Fitri and one month after Eid Fitri. It implies that the movement of cargo data is not easy to predict than air passenger data. Therefore, for the modeling of air passenger and cargo data in the three airports required the influence of the week one month before of Eid Fitri, the week of Eid Fitri and the week after one month of Eid Fitri.

B. Modeling of Air Passenger and Cargo First Phase

The first phase is modeling air passenger and cargo at each airport i.e. Soekarno Hatta, I Gusti Ngurah Rai and Juanda Airport, using the trend, seasonal and calendar variations with TSR and ARIMAX models. In the TSR model, the first step is checking residual assumptions include white noise and normal distribution. If the residual white noise assumption is not fulfilled, significant lags will be included in the model. After all the assumption is satisfied, then estimate and test the significance of the parameter.

Meanwhile, on ARIMAX modeling with the component trend, seasonal and calendar variation, the data is firstly modeled by the time series regression model, then the residual is modeled by ARIMA based on ACF and PACF pattern. After that, checking the assumption of residual is done to know whether the residual follow satisfy white noise and normal distribution. If all the assumption is fulfilled, then parameter estimation and significance testing parameter are performed.

C. Modeling of Air Passenger and Cargo Second Phase

The second phase is modeling by the hybrid method, i.e. TSR-NN, TSR-SVR, ARIMAX-NN, and ARIMAX-SVR. In TSR-NN or ARIMAX-NN modeling, the error of TSR or ARIMAX model was modeled using FFNN three layer. The input used in second phase modeling is a significant lag data in first phase modeling. It aims to capture the neglected nonlinear component of TSR or ARIMAX modeling through advanced modeling using the error. The selection of the number of neurons in the hidden layer was chosen by design experiment from the number of neurons that generate the minimum MAPE testing data.

Determination of parameters in TSR-SVR and ARIMAX-SVR hybrid modeling was using grid search method. This method intends to obtain optimum global results. The inputs used in the TSR-SVR and ARIMAX-SVR models were similar to the inputs used in the TSR-NN and ARIMAX-NN models. The SVR model has three parameters, i.e. σ^2, epsilon, and C. To achieve the optimum parameters, it is necessary to combine the values of the three parameters in order to obtain the smallest MAPE value.

The combined value of σ^2 is used between 0.01 and 1 [10]. The value of epsilon parameters used from range 10 to

100. The value of C used from range 1000 to 100000. The combination of the three parameters offers thousands of combinations which are then selected with the minimum MAPE values testing data. The accuracy performance of TSR-NN, ARIMAX-NN, TSR-SVR, and ARIMAX-SVR models by MAPE values for each data of air passenger and cargo at three international airports are presented in Table 3.

TABLE 3. COMPARISON MAPE VALUE OF TSR-NN, ARIMAX-NN, TSR-SVR AND ARIMAX-SVR HYBRID MODELS

Variable	MAPE Training Data				MAPE Testing Data			
	TSR-NN	ARIMAX-NN	TSR-SVR	ARIMAX-SVR	TSR-NN	ARIMAX-NN	TSR-SVR	ARIMAX-SVR
Domestic Passenger								
Departed from CGK	2.46	3.56	3.89	5.11	5.77	**4.11**	5.26	5.19
Departed from DPS	5.46	5.99	5.30	0.16	**5.32**	5.60	6.95	9.22
Departed from SUB	5.11	5.73	0.05	0.96	14.02	20.87	9.49	**7.16**
Arrived at CGK	3.61	5.26	5.18	6.17	5.51	**4.97**	6.41	5.47
Arrived at DPS	4.08	3.77	3.54	3.19	6.26	**5.41**	7.19	9.35
Arrived at SUB	4.50	4.18	0.05	1.09	23.07	6.30	10.57	**5.51**
International Passenger								
Departed from CGK	5.42	4.52	1.64	2.60	**4.41**	4.86	5.19	4.43
Departed from DPS	3.59	6.41	5.78	0.13	10.92	**7.31**	12.62	7.48
Departed from SUB	6.51	6.85	2.28	6.50	7.49	**5.88**	9.29	9.17
Arrived at CGK	6.17	4.93	6.19	0.43	6.86	7.31	6.86	**6.70**
Arrived at DPS	5.34	6.43	4.88	0.10	**5.70**	6.65	7.73	8.60
Arrived at SUB	5.84	9.36	2.80	8.71	11.32	**7.13**	8.80	10.23
Domestic Cargo								
Loaded from CGK	5.79	5.77	2.06	5.26	10.52	**10.44**	12.86	11.81
Loaded from DPS	10.11	13.18	21.82	14.78	49.19	56.47	50.79	**42.69**
Loaded from SUB	6.32	9.12	2.69	11.04	15.78	21.99	10.28	**7.97**
Unloaded at CGK	6.95	4.87	0.67	2.00	28.41	24.13	14.41	**13.96**
Unloaded at DPS	10.63	9.23	6.98	9.38	**13.57**	23.90	19.56	13.84
Unloaded at SUB	11.42	8.93	4.02	8.54	**6.59**	12.44	10.09	7.73
International Cargo								
Loaded from CGK	9.55	9.48	8.12	7.59	**10.49**	11.33	17.67	15.74
Loaded from DPS	9.06	8.17	5.51	5.17	35.06	**35.00**	35.80	46.00
Loaded from SUB	11.38	13.08	1.64	17.09	**12.41**	13.54	15.69	17.42
Unloaded at CGK	11.46	14.21	5.31	10.82	**16.00**	19.60	18.44	20.02
Unloaded at DPS	15.69	14.53	11.22	12.73	26.18	32.59	31.56	**22.90**
Unloaded at SUB	16.97	19.25	3.25	15.00	23.63	**16.83**	20.21	16.93

From Table 3, it can be seen that the use of hybrid models provides more accurate forecasting results, especially in training data. On average, the TSR-NN hybrid model for air passenger and cargo in training data was able to reduce the MAPE value of the TSR model by 36.52 percent on air passenger data and by 20.84 percent in cargo data. The TSR-SVR model on average can reduce MAPE value of TSR model in training data up to 45.14 percent for air passenger data and 50.96 percent for cargo data. However, for testing data, the TSR-NN model can only reduce the value of the MAPE TSR model by 11.94 percent for air passenger data and by 5.43 percent for cargo data. Therefore, the TSR-SVR forecast for testing data does not always yield better results than the TSR model.

On average, ARIMAX-NN hybrid model on training data was able to reduce the MAPE value of ARIMAX models by 35.73 percent on air passenger data and 19.14 percent on cargo data. The ARIMAX-SVR model on average can reduce MAPE value of ARIMAX model in training data up to 62.64 percent for passenger data and 19.98 percent for cargo data. For testing data, ARIMAX-NN model was only able to reduce the MAPE value of ARIMAX model by 10.74 percent for passenger data and 6.45 percent for cargo data. While in data testing, ARIMAX-SVR forecast results do not always yield better forecasting accuracy than the ARIMAX model.

Based on the MAPE value criteria in testing data, shows that the best forecast results for each air passenger and cargo data on the three airports resulted in different forecasting models. The ARIMAX-NN hybrid method is the best method of forecasting 9 data of air passenger and cargo or as much as 38 percent. The best TSR-NN hybrid method for forecasting 8 data of air passenger and cargo. Moreover, the best ARIMAX-SVR hybrid method for forecasting 7 data of air passenger and cargo. These results indicate that the use of NN in hybrid models give better results than the SVR especially for data of air passenger and cargo at the three international airports in Indonesia.

D. Prediction of Air Passenger and Cargo Until 2018

After obtaining the best forecasting model, the further forecast was made for the September 2017 to December 2018. The predicted numbers of total air passengers for the three airports is presented in Figure 7.

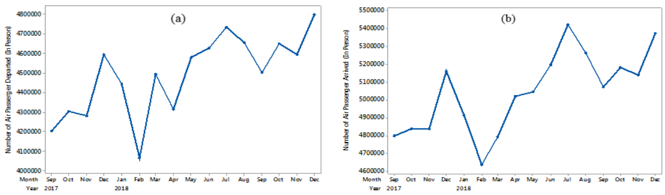

Figure 7 Total Predicted Number of Air Passengers Departing (a) and Arriving (b), In Three International Airports

Figure 7 shows that an estimated the number of air passengers increase highly in December of 2018 both on passenger departure and arrival airport due to Christmas and New Year holidays. In addition, the increasing demand is also expected to be high due to official travels of government and private employees at the end of the year. Otherwise, in February 2018, the passenger movement is expected to decline, because if there is no big event, the number of days in this month is less than the other month. The Eid Fitri in 2018 is estimated in the second week of June. Therefore, it is expected that the increasing number of passengers will occur from May, June, and July.

The total predicted value in three airports of cargo volume is shown in Figure 8. Moreover, Figure 8 explains the predicted numbers of air cargo tend to follow fluctuation pattern depends on global economic conditions.

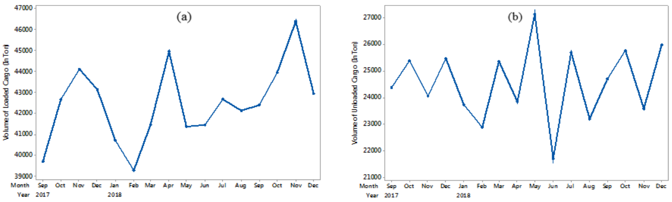

Figure 8 The Total Predicted Value of Loading Cargo (a) and Unloading Cargo (b) In Three International Airports

V. CONCLUSION

The growth of air passenger and cargo in the three airports, namely Soekarno Hatta, I Gusti Ngurah Rai, and Juanda Airport revealed that the data have increase pattern from year to year. In addition, the movement of air passenger and cargo at the three airports has a periodic pattern that is during academic year holiday in July as well as December Christmas and New Year holiday. The air passenger and cargo patterns also increase in Eid Fitri holiday none as follow the effect of calendar variation.

The four hybrid methods, i.e. TSR-NN, TSR-SVR, ARIMAX-NN, and ARIMAX-SVR, was able to reduce the error of the individuals TSR or ARIMAX model in training data, although the testing data does not always yield better forecasting accuracy. Based on the comparison of forecast performance, the results show that ARIMAX-NN and TSR-NN methods yielded better forecasting performance than TSR-SVR and ARIMAX-SVR. Thus the use of NN in hybrid models has better performance compared with the SVR for predictive air passenger and cargo data at the three largest international airports in Indonesia.

ACKNOWLEDGMENT

The author wishes to thank Statistic Indonesia (BPS) and Institut Teknologi Sepuluh Nopember (ITS) which give the author opportunity to study in Master Program of Statistics Department. This research also was supported by DRPM-DIKTI under the scheme of "Penelitian Berbasis Kompetensi", project No. 532/PKS/ITS/2017.

REFERENCES

[1] P. S. Biederman, "The role of forecasting at trans world airlines," *The Journal of Business Forecasting*, vol. 12, no. 3, p. 3, 1993.

[2] H. Matsumoto, "International urban systems and air passenger and cargo flows: some calculations," *Journal of Air Transport Management*, vol. 10, no. 4, pp. 239–247, 2004.

[3] S. S. Blunk, D. E. Clark, and J. M. McGibany, "Evaluating the long-run impacts of the 9/11 terrorist attacks on US domestic airline travel," *Applied Economics*, vol. 38, no. 4, pp. 363–370, 2006.

[4] C. Lim and M. McAleer, "A seasonal analysis of Malaysian tourist arrivals to Australia," *Mathematics and Computers in Simulation*, vol. 48, no. 4, pp. 573–583, 1999.

[5] W. H. K. Tsui, H. O. Balli, A. Gilbey, and H. Gow, "Forecasting of Hong Kong airport's passenger throughput," *Tourism Management*, vol. 42, pp. 62–76, 2014.

[6] J. Faraway and C. Chatfield, "Time series forecasting with neural networks: a comparative study using the airline data," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 47, no. 2, pp. 231–250, 1998.

[7] L. R. Weatherford, T. W. Gentry, and B. Wilamowski, "Neural network forecasting for airlines: A comparative analysis," *Journal of Revenue and Pricing Management*, vol. 1, no. 4, pp. 319–331, 2003.

[8] T. O. Blinova, "Analysis of possibility of using neural network to forecast passenger traffic flows in Russia," *Aviation*, vol. 11, no. 1, pp. 28–34, 2007.

[9] H. J. Heng, B. Z. Zheng, and Y. J. Li, "Study of SVM-based air-cargo demand forecast model," in *Computational Intelligence and Security, 2009. CIS'09. International Conference on*, 2009, vol. 2, pp. 53–55.

[10] A. Shabri, "Least Square Support Vector Machines as an Alternative Method in Seasonal Time Series Forecasting," *Applied Mathematical Sciences. Sci.*, vol. 9, no. 124, pp. 6207–6216, 2015.

[11] S. Makridakis and M. Hibon, "The M3-Competition: results, conclusions and implications," *International Journal of Forecasting*, vol. 16, no. 4, pp. 451–476, 2000.

[12] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[13] D. Ö. Faruk, "A hybrid neural network and ARIMA model for water quality time series prediction," *Engineering Applications of Artificial Intelligence*, vol. 23, no. 4, pp. 586–594, 2010.

[14] Suhartono, I. Puspitasari, M. S. Akbar, and M. H. Lee, "Two-level seasonal model based on hybrid ARIMA-ANFIS for forecasting short-term electricity load in Indonesia," in *Statistics in Science, Business, and Engineering (ICSSBE), 2012 International Conference on*, 2012, pp. 1–5.

[15] P. F. Pai and C. S. Lin, "A hybrid ARIMA and support vector machines model in stock price forecasting," *Omega*, vol. 33, no. 6, pp. 497–505, 2005.

[16] Ministry of Transportation Republic of Indonesia, *Transportation Statistic Volume 1*. Jakarta: Ministry of Transportation Republic of Indonesia, 2016.

[17] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2014.

[18] Suhartono, M. H. Lee, and D. D. Prastyo, "Two levels ARIMAX and regression models for forecasting time series data with calendar variation effects," in *AIP Conference Proceedings*, 2015, vol. 1691, p. 050026.

[19] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: The state of the art," *International Journal of Forecasting*, vol. 14, no. 1, pp. 35–62, 1998.

[20] U. Anders and O. Korn, "Model selection in neural networks. Neural Networks," *Neural Networks*, vol. 12, pp. 309–323, Mar. 1999.

[21] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[22] W. K. Härdle, D. D. Prastyo, and C. M. Hafner, "Support vector machines with evolutionary model selection for default prediction," in *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, Oxford University Press Inc. New York, NY, 2014.