

## Forecasting Currency in East Java: Classical Time Series vs. Machine Learning\*

J A Putri<sup>1</sup>, Suhartono<sup>2‡</sup>, H Prabowo<sup>3</sup>, N A Salehah<sup>4</sup>, D D Prastyo<sup>5</sup>,  
and Setiawan<sup>6</sup>

<sup>1,2,3,4,5,6</sup>Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia

<sup>‡</sup>corresponding author: [suhartono@statistika.its.ac.id](mailto:suhartono@statistika.its.ac.id)

Copyright © 2021 J A Putri, Suhartono, H Prabowo, N A Salehah, D D Prastyo, and Setiawan. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

Most research about the inflow and outflow currency in Indonesia showed that these data contained both linear and nonlinear patterns with calendar variation effect. The goal of this research is to propose a hybrid model by combining ARIMAX and Deep Neural Network (DNN), known as hybrid ARIMAX-DNN, for improving the forecast accuracy in the currency prediction in East Java, Indonesia. ARIMAX is class of classical time series models that could accurately handle linear pattern and calendar variation effect. Whereas, DNN is known as a machine learning method that powerful to tackle a nonlinear pattern. Data about 32 denominations of inflow and outflow currency in East Java are used as case studies. The best model was selected based on the smallest value of RMSE and sMAPE at the testing dataset. The results showed that the hybrid ARIMAX-DNN model improved the forecast accuracy and outperformed the individual models, both ARIMAX and DNN, at 26 denominations of inflow and outflow currency. Hence, it can be concluded that hybrid classical time series and machine learning methods tend to yield more accurate forecasts than individual models, both classical time series and machine learning methods.

**Keywords:** arimax, arimax-dnn, deep neural network, currency, forecasting.

---

\* Received: Jan 2021; Reviewed: Jan 2021; Published: Jun 2021

## 1. Introduction

East Java's economic growth in the second quarter of 2018 reached 5.6% (YoY), where this value is higher than the national one, namely 5.3% (YoY). East Java's economic performance is better than DKI Jakarta, West Java, and Banten. The increase in economic growth will result in more currency circulating in society. This is reinforced by the condition that East Java experiences a net outflow condition. This condition occurs when the money withdrawal transaction is greater than the deposit of money. The total outflow of Rp. 36.56 trillion increased significantly by 99.39% from the first quarter of 2018 in line with the moment of Eid al-Fitr that fell in June 2018 ([BI], 2018).

The need for currency varies significantly from time to time. This is caused by several factors. One of the influencing factors is the variation of the calendar of holidays or holidays such as the Eid al-Fitr. Suhartono, Lee, and Hamzah introduced a time series regression model for modeling data containing calendar variation patterns, i.e. ARIMAX (Lee et al., 2010). The effect of variations in the Eid al-Fitr calendar has a different effect on data inflow and outflow. This is because people tend to save money back after the holiday, so the inflow value will also be influenced by the month after Eid al-Fitr. The outflow value will increase a month before Eid al-Fitr because people will withdraw money before Eid al-Fitr. Therefore, this study will use predictor variables, namely the dummy variable for the month of Eid al-Fitr and one month after Eid al-Fitr for data inflow, while for the outflow data will use the dummy variables for the month of Eid al-Fitr and one month before Eid al-Fitr.

Time series data, especially financial data, generally form a non-linear data pattern. Non-linear modeling for forecasting often uses the artificial neural network method or so-called neural network (NN). NN is one of the machine learning methods. One of the developments in the neural network is a deep neural network (DNN) that has more layers. This model is expected to be able to recognize more complex processes (Schmidhuber, 2015). The research of The M4 Competition: Results, findings, conclusion, and way forward states that the combined or hybrid method can capture both trend and seasonal data patterns, and on average it can increase the accuracy value compared to individual methods (Makridakis et al., 2018b). The results of the study by Zhang provide the same statement, namely combining linear and non-linear models is an effective way to increase the accuracy value (Zhang, 2003).

Based on several studies conducted, it shows that there is no one method of forecasting that is always the best for predicting various types of data. Each method has its advantages and disadvantages, so it is necessary to compare several forecasting methods (Makridakis et al., 2018a). In general, there are three approaches used to forecast inflow and outflow in East Java, namely the linear forecasting model, the non-linear forecasting model, and the hybrid model. In this research, the linear forecasting model used is the ARIMAX model as classical time series model, the non-linear model used is DNN as a machine learning model, and the combined model used is the hybrid ARIMAX-DNN. The selection of the best model is based on the smallest RMSE and sMAPE values. This research is expected to be able to assist in planning to meet the demand for currency in East Java in the future.

The rest of the paper is organized as follows. Section 2 reviews the methodology considered in this paper. Next, we describe the dataset and variable in section 3 and

discuss its results in section 4. Conclusion is given by section 5.

## 2. Research Methods

### 2.1 Time Series Regression

One of the time series models is a time series regression. In general, this model is the same as the linear regression model (Shumway & Stoffer, 2006). The predictor variables used are dummy variables in the form of trends, seasonality, and variations of the Eid al-Fitr calendar as in equation 1.

$$Y_t = \beta_1 t + \sum_{i=1}^S \alpha_i M_{i,t} + \sum_{j=1}^G \gamma_j V_{j,t} + \varepsilon_t, \tag{1}$$

where  $t$  is the trend dummy,  $M_{S,t}$  is a seasonal dummy,  $V_{G,t}$  is a dummy variation of the Eid al-Fitr calendar, and  $\varepsilon_t$  is the residual of the time series regression model that does not necessarily meet the IIDN assumptions  $(0, \sigma_a^2)$ . Figure 1 shows the flowchart of the time series regression model.

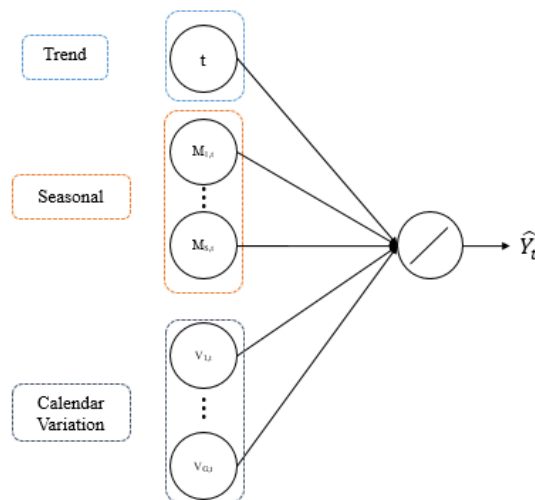


Figure 1: The Flowchart of Time Series Regression.

### 2.2 Autoregressive Intergrated Moving Average with Exogeneous Variable (ARIMAX)

ARIMAX is the development of the ARIMA model where there are additional exogenous variables so that multiple regression can be carried out.  $\varepsilon_t$  is a time series regression residual that does not meet the white noise assumption, where  $\varepsilon_t$  follows the ARIMA equation, namely  $\varepsilon_t = \frac{\theta_q(B)}{\phi_p(B)} a_t$ . If the residuals from the time series regression models do not meet the white noise assumptions, residual modeling is carried out ( $\varepsilon_t$ ) using the ARIMA model. The general model of ARIMAX that contains trend, seasonal, and calendar variation patterns is stated as equation 2 (Suhartono et al., 2015). The flowchart of the ARIMAX model is shown in Figure 2.

$$Y_t = \beta_1 t + \sum_{i=1}^S \alpha_i M_{i,t} + \sum_{j=1}^G \gamma_j V_{j,t} + \frac{\theta_q(B)}{\phi_p(B)} a_t, \tag{2}$$

where  $\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ ,  $a_t$  are the residuals of the ARIMAX

model, and  $\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$ .

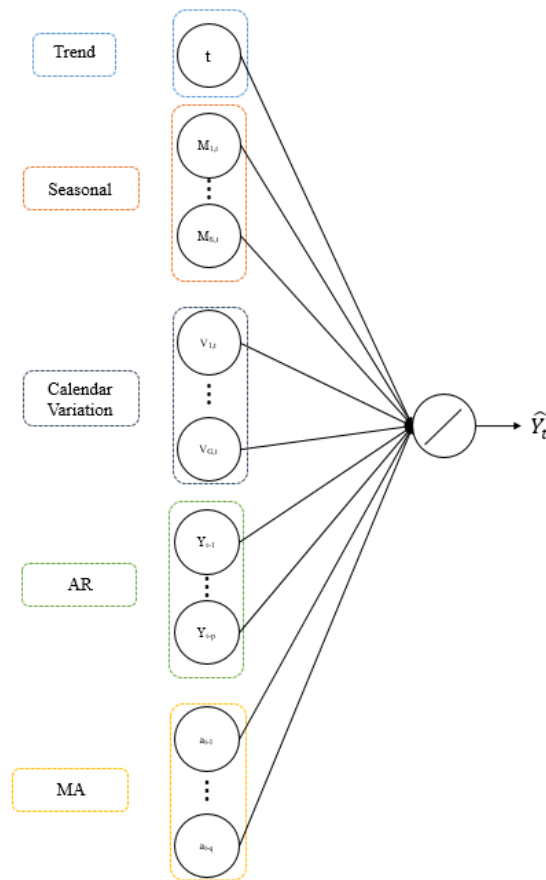


Figure 2: The Flowchart of ARIMAX Model.

### 2.3 Deep Neural Network (DNN)

Deep Neural Network (DNN) is a development of a Neural Network (NN). DNN has more deep layers than NN. This model is expected to recognize more complex processes. DNN architecture can be shown in Figure 3. The input of the DNN model is dummy and lag variables. The selection of the input lag is based on the significant PACF lag on stationary data (Crone & Kourentzes, 2009). The input and hidden layers have a bias of one (Zhang & Qi, 2005). The greater the number of neurons used in the hidden layer, it will increase the model accuracy. But also allow the out-sample prediction value to decrease (Faraway & Chatfield, 1998). Therefore, the number of neurons used in the hidden layer was tried from one to five. The general model used in DNN can be generalized from a single layer FFNN and can be shown in Equation 3 (Suhartono et al., 2019).

$$\hat{Y}_{(t)} = f^o \left[ b^o + \sum_{i=1}^p w_i^o f_i^{h2} \left( b_i^{h2} + \sum_{j=1}^q w_{ji}^{h2} f_j^{h1} \left( b_j^{h1} + \sum_{k=1}^r w_{kj}^{h1} X_{k(t)} \right) \right) \right] \quad (3)$$

with  $X_{k(t)}$  is the input variable,  $\hat{Y}_{(t)}$  is the estimated value of the output variable,  $f^o$  is an activation function on neurons in the output layer.  $b^o$  is bias in neurons,  $w_i^o$  is the weight of the neurons to the output layer,  $w_{ji}^{h2}$  is the weight of neurons to hidden layer 2,  $w_{kj}^{h1}$  is the weight of neurons to the hidden layer 1.

The algorithm most often used for time series analysis is to fit the function

parameters, especially to change the weights associated with neurons in the hidden layer, is the backpropagation algorithm with the gradient descent method (Dordonnat et al., 2008). The activation function used in the hidden layer is the sigmoid function. This function is good for use in neural networks that use backpropagation because it can minimize calculations (Karlik & Olgac, 2011). The most widely used activation function for the output layer is the linear function because non-linear activation functions can distort the specified output (Khashei & Bijari, 2010).

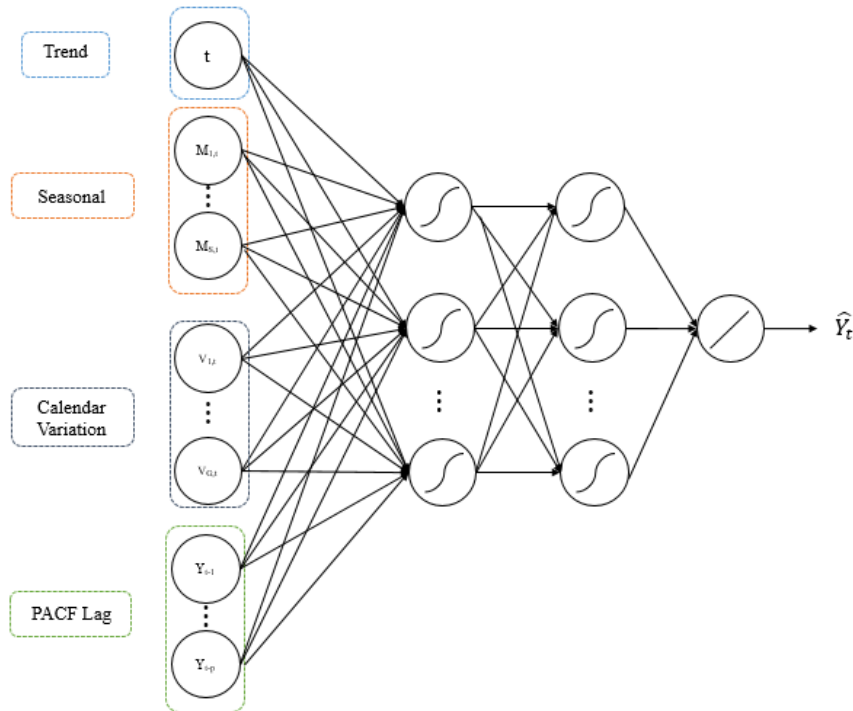


Figure 3: Architecture Deep Neural Network.

## 2.4 Hybrid ARIMAX- DNN

Combined time series models have a better forecast accuracy value than individual models. Based on the M4-Competition research, it shows that the combined method on average can increase the accuracy of forecasting compared to the individual model (Makridakis et al., 2018b). ARIMAX model can capture linear patterns, but cannot capture non-linear patterns. Therefore, this study combines ARIMAX and Deep Neural Network (DNN) methods in series with the hope of being able to capture linear and non-linear patterns. Series hybrids consist of linear and non-linear components where these components are processed sequentially. The general model of time series hybrids is as Equation 4 (Zhang, 2003). Figure 4 shows the flowchart of the hybrid model

$$Y_t = Y_t^{(l)} + Y_t^{(n)} + e_t \quad (4)$$

with  $Y_t^{(l)}$  is a linear component and  $Y_t^{(n)}$  is a non-linear component.

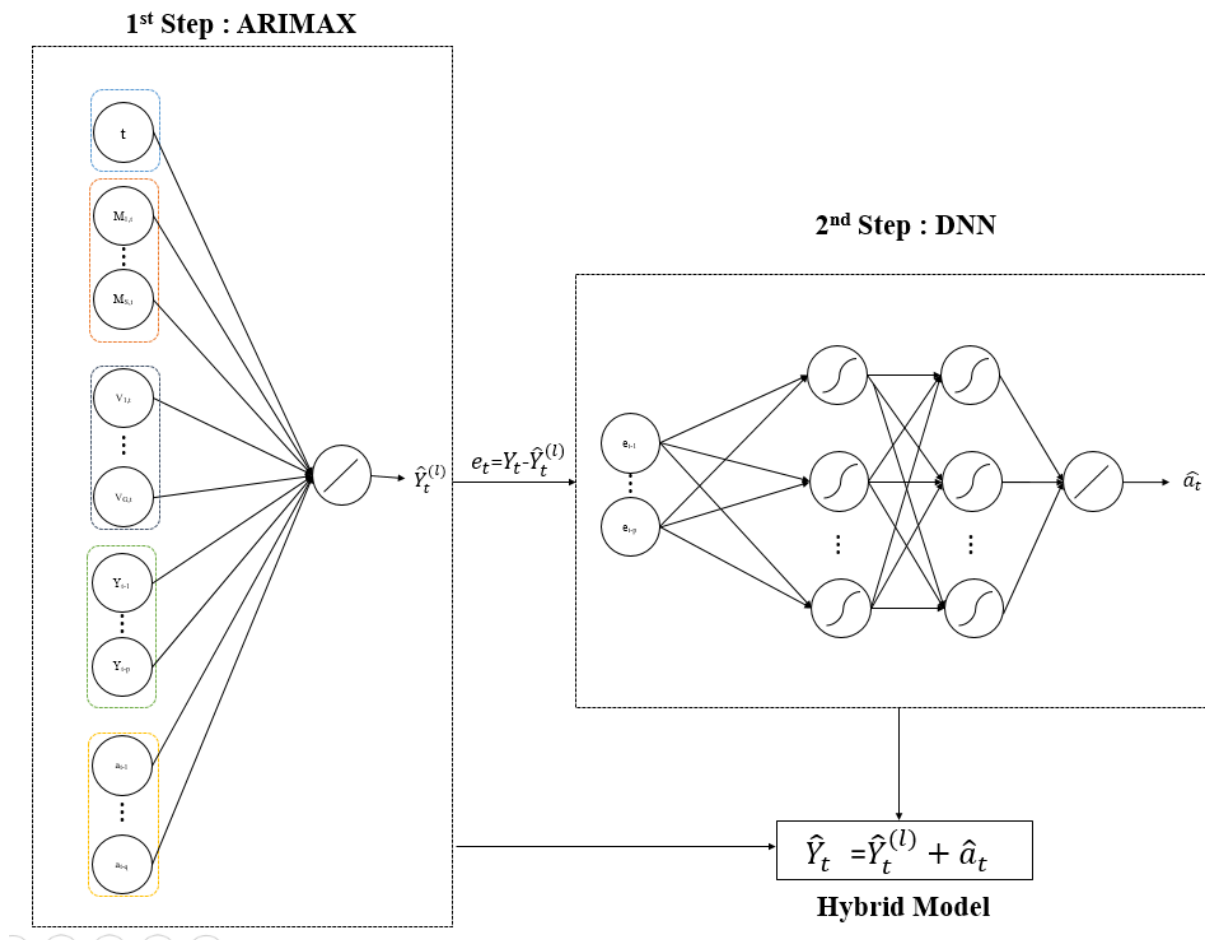


Figure 4: The Flowchart of Hybrid Model.

### 2.5 Model Evaluation

This study uses three models, namely ARIMAX, DNN, and hybrid ARIMAX-DNN model. The right model for in-sample does not necessarily result in a high value of model merit on the out-sample data (Makridakis et al., 2018a). Therefore, in this study choosing the best model based on out-sample data. Often, RMSE is preferred over MSE if the data used are of the same scale. Generally, RMSE is used very often because it is relevant to statistical modeling (Hyndman & Koehler, 2006). The RMSE equation is shown in Equation 5.

$$RMSE = \sqrt{\frac{1}{L} \sum_{l=1}^L (Y_{n+l} - \hat{Y}_n(l))^2} \tag{5}$$

However, RMSE is very sensitive if there are outlier data. So, another measure of the goodness of the model is used, namely sMAPE. sMAPE is more stable if there are outlier data (Makridakis, 1993). Using the sMAPE model's measure avoids large error values when actual data is close to 0 (Makridakis & Hibon, 2000). The sMAPE equation is shown in Equation 6.

$$sMAPE = \left( \frac{1}{L} \sum_{l=1}^L \frac{2|Y_{n+l} - \hat{Y}_n(l)|}{(|Y_{n+l}| + |\hat{Y}_n(l)|)} \right) \times 100\% \tag{6}$$

### 3. Dataset and Methodology

#### 3.1 Dataset

The data used in this study is secondary data that is the inflow and outflow data of Bank "X" in East Java with a monthly period from January 2010 to June 2019. The data will be divided into in-sample and out-sample data. The in-sample data used is January 2010 to December 2017, while the out-sample data is from January 2018 to June 2019.

#### 3.2 Research Variable

The research variables used in this study consisted of 8 banknotes at the inflow and outflow. Table 1 is the variables that will be used in this study.

Table 1: Research Variable.

Variable	Inflow	Variable	Outflow
$Y_{1,t}^{(i)}$	Total	$Y_{9,t}^{(i)}$	Total
$Y_{2,t}^{(i)}$	IDR 100,000.00	$Y_{10,t}^{(i)}$	IDR 100,000.00
$Y_{3,t}^{(i)}$	IDR 50,000.00	$Y_{11,t}^{(i)}$	IDR 50,000.00
$Y_{4,t}^{(i)}$	IDR 20,000.00	$Y_{12,t}^{(i)}$	IDR 20,000.00
$Y_{5,t}^{(i)}$	IDR 10,000.00	$Y_{13,t}^{(i)}$	IDR 10,000.00
$Y_{6,t}^{(i)}$	IDR 5,000.00	$Y_{14,t}^{(i)}$	IDR 5,000.00
$Y_{7,t}^{(i)}$	IDR 2,000.00	$Y_{15,t}^{(i)}$	IDR 2,000.00
$Y_{8,t}^{(i)}$	IDR 1,000.00	$Y_{16,t}^{(i)}$	IDR 1,000.00

Table 2: List of Eid al-Fitr 2010-2020.

Year	Date Eid al-Fitr	Week- $j$	Dummy Variable		
			$V_{j,t-1}$	$V_{j,t}$	$V_{j,t+1}$
2010	10-11 September	2	August	September	October
2011	30-31 August	4	July	August	September
2012	19-20 August	3	July	August	September
2013	08-09 August	2	July	August	September
2014	28-29 July	4	June	July	August
2015	17-18 July	3	June	July	August
2016	06-07 July	1	June	July	August
2017	26-27 June	4	May	June	July
2018	15-16 June	3	May	June	July
2019	3-4 June	1	May	June	July
2020	23-24 May	4	April	May	June

Table 1 shows the inflow and outflow variables for KPwi with  $i = 1, 2, 3, 4$  that KPw1 in Jember, KPw2 in Kediri, KPw3 in Malang, and KPw4 in Surabaya. The pattern of calendar variations in the data follows the occurrence of Eid al-Fitr with the provisions that the 1st week is the 1st to the 7th, the 2nd week is the 8th to the 15th, the 3rd week is the 16th to the 23rd and the 4th week is the 24th to the end. Data on the occurrence of Eid al-Fitr from 2010 to 2020 are presented in Table 2.

### 3.3 Step of Analysis

This research was carried out in several stages to predict the inflow and outflow in East Java, as follows.

1. Describe the data characteristics of the inflow and outflow for each KPw in East Java Province through descriptive statistics and visually with diagrams.
2. Divide the data into two, namely in-sample data and out-sample data. In-sample data is from January 2010 to December 2017, while out-sample data is from January 2018 to June 2019.
3. Perform modeling and forecasting for all banknote fractions and the total inflow and outflow in each KPw using the ARIMAX method with the following steps.
  - a. Regressing the inflow and outflow with dummy variables of trends, seasonality, variations in the calendar month of Eid, and one month after Eid.
  - b. Perform a diagnostic check on the time series regression residuals. If it meets the white noise assumption, then it is followed by forecasting using the time series regression model. However, if it does not meet the assumptions then proceed to the next step.
  - c. Perform ARIMA modeling by identifying through the ACF and PACF orders the time series regression residuals.
  - d. Perform a diagnostic check again on the residual ARIMAX model.
  - e. Perform forecasting with the ARIMAX model and get the RMSE and sMAPE values.
4. Modeling and forecasting inflow and outflow data for each fraction using the Deep Neural Network method with the following steps:
  - a. Determine input variables based on trend component variables, seasonality, and calendar variations in ARIMAX, and lag response variables based on significant PACF plots.
  - b. Preprocessing data, namely significant PACF lag using normalization. Preprocessing needs to be done and the data will be worth between 0 and 1 [12]
  - c. Determine the number of inputs, the number of neurons in the hidden layer, and the activation function.
  - d. Perform forecasting with the Deep Neural Network architecture and get the RMSE and sMAPE values.
  - e. Postprocessing data.
5. Modeling and forecasting inflow and outflow data for each fraction using the series hybrid method as follows.
  - a. Perform ARIMAX modeling on inflow and outflow data.
  - b. Normalizing the residual lag that will be used as DNN modeling input.



- c. Using two types of input lag, the first input lag is to use the AR order in the ARIMAX model, and the second input lag is lag 1, 12, 35 as justification for trends, seasonality, and calendar variations.
  - d. Combines ARIMAX and DNN models into ARIMAX-DNN hybrid models.
  - e. Postprocessing data.
6. Comparing ARIMAX, Deep Neural Network, and ARIMAX-DNN hybrid series models and selecting the best model based on RMSE and sMAPE values in the out-sample data.

## 4. Result and Discussion

### 4.1 Inflow and Outflow Characteristics in East Java Province

Descriptive analysis was carried out to find an overview of inflow and outflow data from January 2010 to June 2019 in 4 KPws spread across East Java Province, namely KPw Jember, KPw Kediri, KPw Malang, and KPw Surabaya. The inflow and outflow characteristics of each KPw are described in Figure 5. Based on Figure 5, the highest inflow of total banknotes in the four KPws in East Java Province occurred in January, June, July, and August. The highest average total outflow of banknotes was in KPw Jember, KPw Kediri, KPw Malang, and KPw Surabaya, namely in May, June, July, and December.

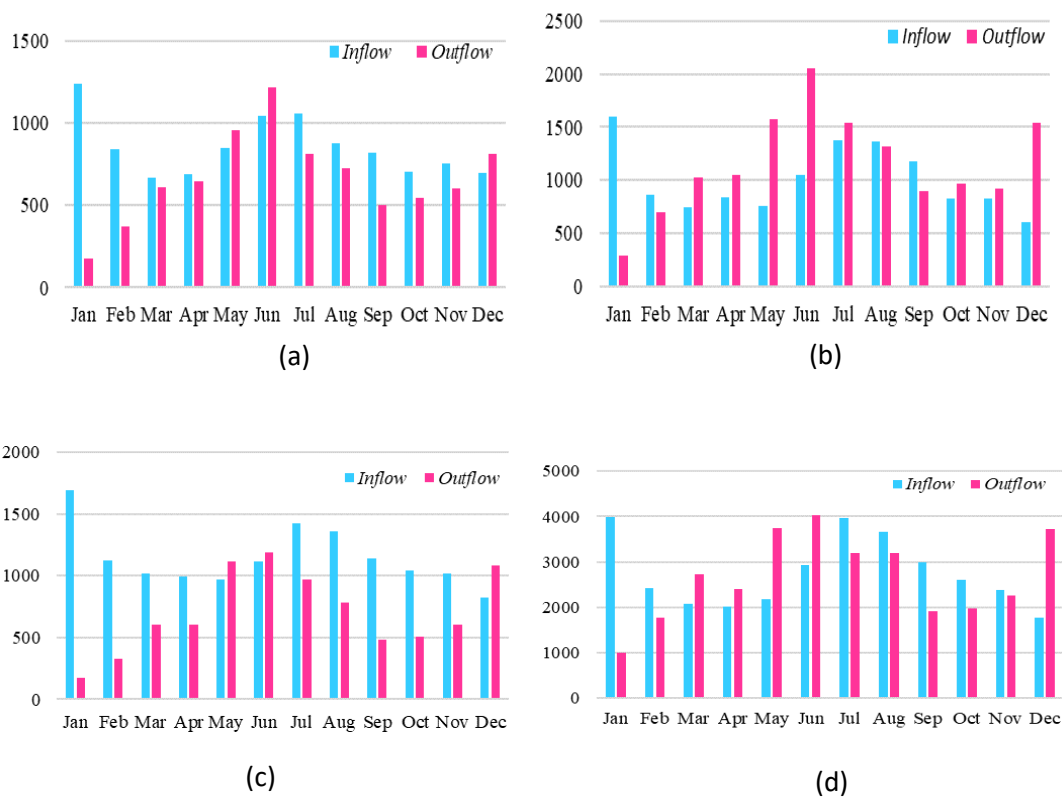


Figure 5: Average Total Banknotes Inflow and Outflow in a Month in (a) KPw Jember, (b) KPw Kediri, (c) KPw Malang and (d) KPw Surabaya.

The bar chart in Figure 6 shows a significant effect of Eid al-Fitr on the movement of inflow and outflow of total banknotes in each KPw in East Java Province. Previous research also shows the influence of Eid al-Fitr on the week of Eid al-Fitr. The average inflow during and one month after Eid is shown in Figure 6.

Figure 6 shows the average increase in the total inflow of paper shards in the four KPws in East Java Province from January 2010 to June 2019. The increase in inflow based on the week of Eid al-Fitr has the following pattern.

1. If Eid al-Fitr occurs in the first week, the average inflow has a very high value in the month of Eid al-Fitr and decreases one month after Eid.
2. If Eid al-Fitr occurs in the second week, then the average inflow has a higher value than one month after Eid.
3. If Eid al-Fitr occurs in the third week, the average inflow will have a value that is almost the same as one month after Eid.
4. If Eid al-Fitr occurs in the fourth week, the average inflow will be high one month after the Eid al-Fitr.

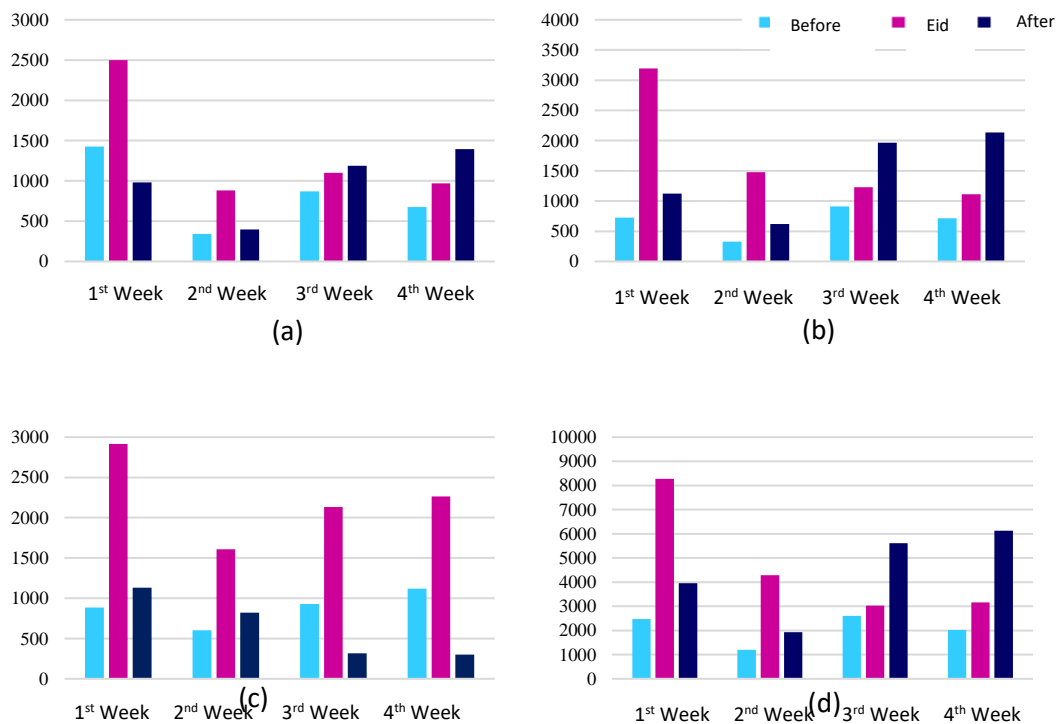


Figure 6: Average Total Banknotes Inflow According to Week of Eid Al-Fitr in (a) KPw Jember, (b) KPw Kediri, (c) KPw Malang and (d) KPw Surabaya.

The pattern of increasing outflow is different from the pattern of increasing inflow. The outflow is influenced by the month of Eid al-Fitr and one month before the onset of Eid because people tend to withdraw money before celebrating Eid. For more details, it is shown in Figure 7.

Based on Figure 7, it is known that the average increase in the total outflow in the four KPws in East Java Province from January 2010 to June 2019. The increase in outflow based on the week of Eid al-Fitr has the following pattern.

1. If Eid al-Fitr occurs in the first week, the average outflow has a very low value in the month of Eid al-Fitr, and one month before the occurrence of Eid al-Fitr has a very high value.
2. If Eid al-Fitr occurs in the second week, the average outflow is almost the same as the one month before Eid.
3. If Eid al-Fitr occurs in the third week, the average outflow will have a higher value than the one month before Eid.
4. If Eid al-Fitr occurs in the fourth week, the average outflow will be high at the time of Eid al-Fitr.

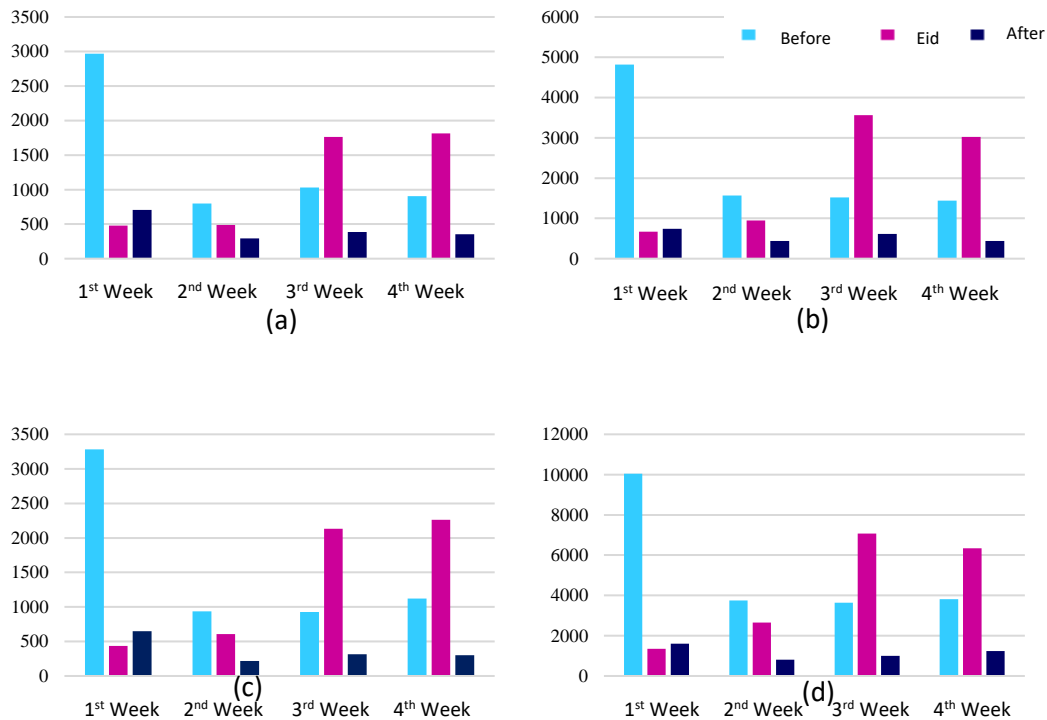


Figure 7: Average Outflow of Total Banknotes According to Eid Al-Fitr Week in (a) KPw Jember, (b) KPw Kediri, (c) KPw Malang and (d) KPw Surabaya.

#### 4.2 Inflow and Outflow Modeling with ARIMAX in East Java Province

The step taken before modeling the inflow and outflow data with ARIMAX is modeling with time series regression (TSR). Time series regression was performed with predictor variables, namely dummy variables consisting of trend dummy, month dummy, and calendar variation dummy. After the time series regression is carried out, the parameter estimation results are obtained, and the Ljung Box test is carried out to check the white noise assumption. Residual from TSR in some inflow and outflow has not met the white noise assumption. Furthermore, ARIMA is used in modeling residuals of time series regression to meet white noise assumptions. This model is called the ARIMAX. So, the model that meets the assumption of white noise is obtained.

$$\begin{aligned}
 Y_t = & 4.1101t + 204.502M_{1,t} + 96.6047M_{2,t} + 50.658M_{3,t} + 36.8312M_{4,t} + 93.4497M_{5,t} + \\
 & 70.6233M_{6,t} + 80.4717M_{7,t} + 53.5539M_{8,t} + 33.5816M_{9,t} + 44.0327M_{10,t} + 61.8094M_{11,t} \\
 & + 33.9793M_{12,t} + 248.450V_{1,t} - 2,44008V_{1,t+1} + 201.814V_{2,t} + 12.0212V_{2,t+1} + 142.363V_{3,t} - \\
 & 1.51067V_{3,t+1} - 76.1263V_{4,t} + 290.622V_{4,t+1} + \frac{1}{(1-0.37929B)} a_t.
 \end{aligned}
 \tag{7}$$

Comparison of the goodness of the model based on the smallest RMSE and sMAPE values with significant variables in the ARIMAX model and the best model is obtained with complete variables. The equation of the ARIMAX model (1,0,0) with complete variables for the Rp 50,000.00 inflow in KPw Jember is as follows in Equation 7. Table 3 shows the best ARIMAX or time series regression models for each data inflow and outflow.

Table 3: ARIMAX Inflow and Outflow Modelling in East Java.

Variable		Model	Variable		Model
Inflow at KPw Jember	Total	Time Series Regression	Total	ARIMAX (2,0,0)	
	100,000	Time Series Regression	100,000	ARIMAX ([1,2,4], 0,0)	
	50,000	ARIMAX (1,0,0)	50,000	ARIMAX ([1,2,12], 1,0)	
	20,000	ARIMAX ([1,11], 0,0)	Inflow at KPw Malang	20,000	Time Series Regression
	10,000	Time Series Regression	10,000	Time Series Regression	
	5,000	ARIMAX ([1,11,12], 0,0)	5,000	Time Series Regression	
	2,000	ARIMAX ([1,10], 0,0)	2,000	ARIMAX (1,0,0)	
	1,000	ARIMAX (0,0, [1,2,8,9])	1,000	ARIMAX (1,0,0)	
Outflow at KPw Jember	Total	ARIMAX ([3], 0,0)	Total	ARIMAX (2,0,0)	
	100,000	ARIMAX ([3], 0,0)	100,000	ARIMAX (2,0,0)	
	50,000	ARIMAX ([3,10,12], 0,0)	50,000	Time Series Regression	
	20,000	ARIMAX (0,0,1)	Outflow at KPw Malang	20,000	ARIMAX ([1,11,12], 0,0)
	10,000	ARIMAX (0,0,1)	10,000	ARIMAX ([1,11], 0,0)	
	5,000	ARIMAX (0,0,1)	5,000	Time Series Regression	
	2,000	ARIMAX (0,0,1)	2,000	Time Series Regression	
	1,000	Time Series Regression	1,000	Time Series Regression	

Variable		Model	Variable		Model
Inflow at KPw Kediri	Total	Time Series Regression	Total	ARIMAX ([2,4], 0,0)	
	100,000	Time Series Regression	100,000	ARIMAX ([4], 0,0)	
	50,000	Time Series Regression	50,000	ARIMAX (3,1,0)	
	20,000	Time Series Regression	Inflow at KPw Surabaya	20,000	ARIMAX (1,0,0)
	10,000	Time Series Regression	10,000	ARIMAX ([1,11], 0,0)	
	5,000	ARIMAX ([1,11], 0,0)	5,000	ARIMAX ([1,2,11], 0,0)	
	2,000	ARIMAX ([1,2,12], 0,0)	2,000	ARIMAX ([1,4], 0,0)	
	1,000	ARIMAX (1,0,0)	1,000	ARIMAX (0,0, [1,2,10])	
Outflow at KPw Kediri	Total	Time Series Regression	Total	ARIMAX ([3], 0,0)	
	100,000	Time Series Regression	100,000	ARIMAX ([3], 0,0)	
	50,000	ARIMAX ([17,22], 0,0)	50,000	ARIMAX ([3], 0,0)	
	20,000	Time Series Regression	Outflow at KPw Surabaya	20,000	Time Series Regression
	10,000	ARIMAX ([1,11], 0,0)	10,000	Time Series Regression	
	5,000	Time Series Regression	5,000	ARIMAX (0,0,1)	
	2,000	Time Series Regression	2,000	ARIMAX ([1,23], 0,0)	
	1,000	ARIMAX (0,0, [12])	1,000	Time Series Regression	

#### 4.3 Inflow and Outflow Modeling with Deep Neural Network in East Java Province

The first step in modeling inflow and outflow with a Deep Neural Network (DNN) is to determine the input. Inputs to DNN are significant PACF lag, trend dummy, month dummy, and calendar variation dummy. The architecture used is two hidden layers with the number of each layer being one to five neurons. The activation function used in the hidden layer is sigmoid, and the activation function used in the output layer is linear. Meanwhile, the weighting value is obtained using the backpropagation algorithm by replicating ten times to get the best weight value. The input used is a significant lag from the PACF data that has been stationary.

Table 4: DNN Inflow and Outflow Modelling in East Java.

Variable		Inflow		Outflow			
		Inflow Input Lag	Neurons on Hidden Layer		Outflow Input Lag	Neurons on Hidden Layer	
			1	2		1	2
KPw Jember	Total	1,2,3,12,13,14,15	2	4	1,2,3,12,13,14,15	5	2
	100,000	1,2,3,12,13,14,15	4	3	1,2,3,12,13,14,15	5	2
	50,000	1,2,12,13,14	3	5	1,2,3,12,13,14,15	1	1
	20,000	12,24	3	5	12,24	3	2
	10,000	2,10,12,14,22	3	2	10,11,12,22,23	4	5
	5,000	11,12,23	3	5	12,24	2	4
		1,2,3,11,12,13,14,1					
	2,000	5,23,24	3	2	12	4	4
	1,000	1,3,4,6,7,12,13,15,16,18,19	4	1	12,24	2	2
KPw Kediri	Total	1,2,3,12,13,14,15	1	4	1,2,3,4,5,12,13,14,15,16,17	5	5
	100,000	1,2,3,12,13,14,15	2	4	1,2,3,4,5,12,13,14,15,16,17	5	5
	50,000	1,2,3,12,13,14,15	1	4	1,2,3,4,5,12,13,14,15,16,17	3	2
	20,000	11,12,23,24	5	1	12,24	3	1
	10,000	2,10,12,14,22,24	5	2	10,11,12,22,23,35	1	3
	5,000	2,12,14,23,24,35	4	3	12,34	2	2
		1,2,3,11,12,13,14,1					
	2,000	5,23,24	3	4	11,12,23,24	3	4
	1,000	1,12,13,14	1	5	12,24,35,36	5	4
KPw Malang	Total	1,2,3,12,13,14,15,34,35	1	3	1,2,3,4,12,13,14,15,16	1	1
	100,000	1,2,3,12,13,14,15,34,35	1	2	1,2,3,4,12,13,14,15,16	5	4
	50,000	1,2,3,6,7,12,13,14,15,18,19	4	2	1,2,3,4,12,13,14,15,16	3	5
	20,000	12,21,33,35	1	1	12,24,34	4	4
	10,000	2,10,11,12,14,22,23,24	1	1	10,11,12,22,23	5	2
	5,000	1,11,12,13,22,23,34	1	2	12,13,25,34,35	2	4
	2,000	1,2,11,12,13,14,23,24	5	3	12,22,23,34,35	2	1
	1,000	1,2,3,4,12,13,14,15,16	4	3	12,24	1	5
	KPw Surabaya	Total	1,2,3,4,12,13,14,15,16	2	5	1,2,3,4,12,13,14,15,16	2

Variable	Inflow		Outflow			
	Inflow Input Lag	Neurons on Hidden Layer		Outflow Input Lag	Neurons on Hidden Layer	
		1	2		1	2
100,000	1,2,3,4,12,13,14,15,16	4	3	1,2,3,4,12,13,14,15,16	2	5
50,000	1,2,3,4,12,13,14,15,16	5	4	1,2,3,4,12,13,14,15,16	3	1
20,000	1,10,11,12,13,22,23,24,35,36	3	2	12,13,24,25	2	2
10,000	2,10,11,12,14,22,23,24	1	2	10,11,12,22,23	2	1
5,000	2,12,14,23,35	2	1	1,12,13,22,24,34	3	5
2,000	1,2,12,13,14	1	5	1,12,13,22,23,34,35	2	4
1,000	1,2,3,12,13,14,15	3	4	12,24,36	1	2

The selection of the best architecture is based on the smallest RMSE and sMAPE values in the out-sample data. The neurons used are 1 to 5 in the first hidden layer and second hidden layer. So, there are 25 combinations. The optimum architectural results for the Rp50,000 inflow data in KPw Jember are shown in Figure 8. And the same way is done for all fractions in each KPw shown in Table 4.

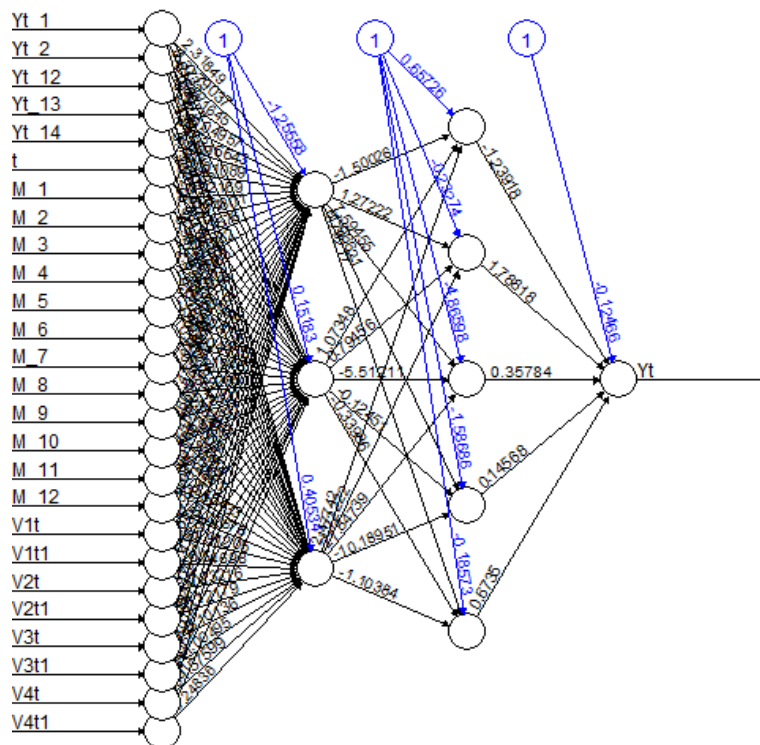


Figure 8: DNN Optimum Architecture for the Rp50,000 Inflow Data in KPw Jember.

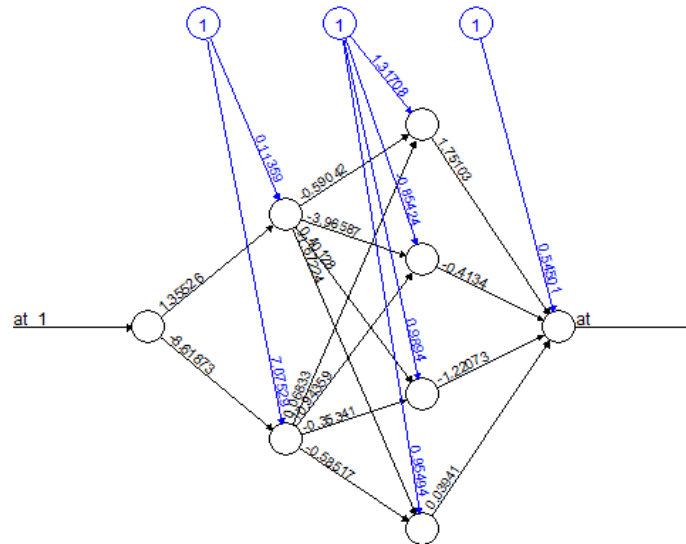


Figure 9: Hybrid ARIMAX-DNN Optimum Architecture for the Rp50,000 Inflow Data in KPw Jember.

#### 4.4 Inflow and Outflow Modeling with Hybrid ARIMAX-Deep Neural Network in East Java Province

After the data is modeled with ARIMAX and the residual value is obtained, then hybrid ARIMAX-DNN modeling will be carried out. The ARIMAX residual will be modeled with DNN and the DNN forecast results will be added with the ARIMAX forecast. There are two types of input used in ARIMAX residual modeling using the DNN method, the first input is the AR model in the ARIMAX model and the second input is lag 1, 12 and 35 to eliminate the effects of trend, seasonal and calendar variations (Faraway & Chatfield, 1998).

Table 5: Hybrid ARIMAX-DNN Inflow and Outflow Modelling in East Java.

Variable		Inflow		Outflow			
		Input Lag	Neurons on Hidden Layer		Input Lag	Neurons on Hidden Layer	
			1	2		1	2
KPw Jember	Total	1,12,35	4	1	1,12,35	2	4
	100,000	1,12,35	5	1	3	2	4
	50,000	1	2	4	1,12,35	5	4
	20,000	1,12,35	4	3	1,12,35	2	1
	10,000	1,12,35	4	1	1,12,35	2	5
	5,000	1,12,35	2	4	1,12,35	5	2
	2,000	1,12,35	2	1	1,12,35	4	3
1,000	1,12,35	4	3	1,12,35	4	1	



Variable		Inflow			Outflow		
		Input Lag	Neurons on Hidden Layer		Input Lag	Neurons on Hidden Layer	
			1	2		1	2
KPw Kediri	Total	1,12,35	5	3	1,12,35	4	1
	100,000	1,12,35	4	3	1,12,35	4	2
	50,000	1,12,35	5	4	1,12,35	5	3
	20,000	1,12,35	1	5	1,12,35	1	2
	10,000	1,12,35	2	2	1,12,35	4	3
	5,000	1,11	4	2	1,12,35	4	5
	2,000	1,12,35	5	2	1,12,35	3	2
	1,000	1,12,35	2	2	1,12,35	5	3
KPw Malang	Total	1,12,35	2	2	1,12,35	5	5
	100,000	1,12,35	4	2	1,12,35	4	5
	50,000	1,12,35	1	3	1,12,35	2	4
	20,000	1,12,35	4	5	1,12,35	4	5
	10,000	1,12,35	5	5	1,12,35	2	4
	5,000	1,12,35	2	5	1,12,35	4	2
	2,000	1,12,35	5	4	1,12,35	4	3
	1,000	1,12,35	4	4	1,12,35	4	1
KPw Surabaya	Total	1,12,35	4	5	2,4	2	4
	100,000	4	1	5	4	2	4
	50,000	1,12,35	2	5	1,12,35	4	2
	20,000	1,12,35	1	4	1,12,35	4	2
	10,000	1,12,35	2	5	1,12,35	4	1
	5,000	1,12,35	1	2	1,12,35	5	1
	2,000	1,12,35	5	5	1,12,35	5	3
	1,000	1,12,35	1	1	1,12,35	2	3

The optimum architecture selection is based on the smallest RMSE and sMAPE values in the out-sample data. The first input is tested based on the ARIMAX model (1,0,0) which means that the input is only lag 1, while the second input is lag 1, 12, and 35. Figure 9 shows the optimum architecture for the Rp50,000 inflow data in KPw Jember, where there are two neurons in the first hidden layer and the second four neurons in hidden layer 2. The same way is done for the total inflow and outflow banknotes and banknotes in KPw Jember and other KPw with the results in Table 5.

#### 4.5 Comparison of Inflow and Outflow Forecasting Results in East Java

After analyzing the ARIMAX, DNN, and hybrid ARIMAX-DNN models, the forecast results are compared. Comparison of forecast results is carried out to obtain the best method. The best method was selected based on the smallest RMSE and sMAPE

values in the out-sample data for each fraction. Figure 10 presents the hybrid ARIMAX-DNN method that is the best method for predicting inflow and outflow in all KPws in East Java Province, with as many as 32 fractions. DNN method is the best method for inflow data of 14 fractions and outflow data of 12 fractions. The ARIMAX method is only better for predicting six banknotes. So, it can be concluded that the hybrid and non-linear models tend to be better at modeling the inflow and outflow of banknotes in East Java. It shows that time-series data tend not to be sufficiently modeled by one model (Khashei & Bijari, 2010). The results of this study are in line with the M4-Competition research, namely the hybrid method on average can improve forecasting accuracy compared to individual models (Makridakis et al., 2018b).

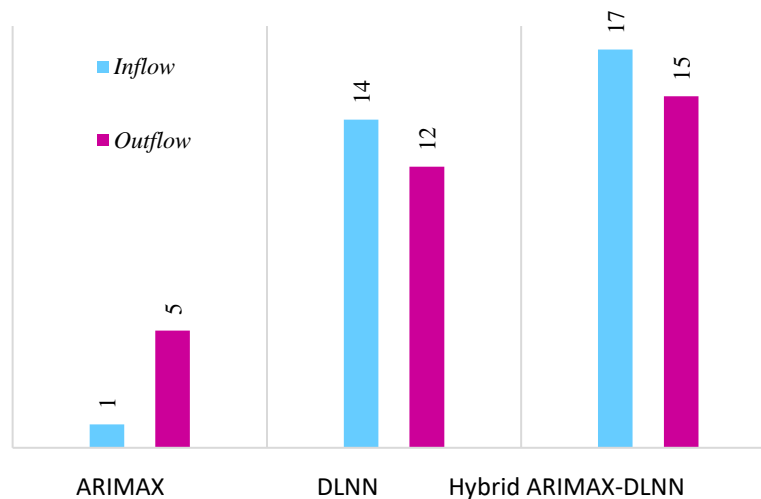


Figure 10: Best Methods of Inflow and Outflow in East Java.

## 5. Conclusion

Based on the results of the analysis and discussion, it was found that the inflow value increases at certain times, namely during Eid al-Fitr and one month after Eid. Meanwhile, the outflow value has increased during Eid al-Fitr and one month before Eid. Eid al-Fitr holidays that occur on different weeks will result in differences in increased inflow and outflow in all KPws in East Java Province. It shows that the data has a calendar variation effect. Based on the comparison of forecast results using the ARIMAX, DNN, and hybrid ARIMAX-DNN methods, it shows that hybrid ARIMAX-DNN is the best method in predicting 32 fractions. Furthermore, the DNN method is the best method for predicting 26 fractions and other fractions predicted by the ARIMAX method. It shows that hybrid models tend to be better at modeling the inflow and outflow of banknotes in East Java. The hybrid model can capture the trend, seasonal, calendar variations, and nonlinear patterns from data. Suggestions for further research are to consider using outlier detection to overcome the unfulfilled assumptions of normal distribution. For the DNN method, other activation functions should be tried in the hidden layer so that comparisons can be made and find out which activation function is more appropriate. Another approach based on the spatial-temporal model can also be proposed for forecasting inflow and outflow simultaneously at four locations in East Java.

## Acknowledgments

This research was supported by DRPM under the scheme of “Penelitian Unggulan Perguruan Tinggi”, project No. 1226/PKS/ITS/2020. The authors thank the General Director of DIKTI for funding and anonymous referees for their useful suggestions.

## References

- [BI], Bank Indonesia. (2018). *Regional Economic and Financial Studies of East Java Province*. Surabaya: Bank Indonesia Representative Office, East Java Province.
- Crone, S. F., & Kourentzes, N. (2009). Input-variable specification for neural networks—an analysis of forecasting low and high time series frequency. *2009 International Joint Conference on Neural Networks*, 619–626. IEEE.
- Dordonnat, V., Koopman, S. J., Ooms, M., Dessertaine, A., & Collet, J. (2008). An hourly periodic state space model for modelling French national electricity load. *International Journal of Forecasting*, 24(4): 566–587.
- Faraway, J., & Chatfield, C. (1998). Time series forecasting with neural networks: a comparative study using the air line data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(2): 231–250.
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4): 679–688.
- Karlik, B., & Olgac, A. V. (2011). Performance analysis of various activation functions in generalized MLP architectures of neural networks. *International Journal of Artificial Intelligence and Expert Systems*, 1(4): 111–122.
- Khashei, M., & Bijari, M. (2010). An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with Applications*, 37(1): 479–489.
- Lee, M. H., Hamzah, N., & others. (2010). Calendar variation model based on ARIMAX for forecasting sales data with Ramadhan effect. *Proceedings of the Regional Conference on Statistical Sciences*, 349–361.
- Makridakis, S. (1993). Accuracy measures: theoretical and practical concerns. *International Journal of Forecasting*, 9(4): 527–529.
- Makridakis, S., & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting*, 16(4): 451–476.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018a). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS One*, 13(3): e0194889.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018b). The M4 Competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, 34(4): 802–808.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61: 85–117.

- Shumway, R. H., & Stoffer, D. S. (2006). *Time Series Analysis and Its Applications with R Examples*. Pittsburg: Springer.
- Suhartono, D. E. A., Prastyo, D. D., Kuswanto, H., & Lee, M. H. (2019). Deep Neural Network for Forecasting Inflow and Outflow in Indonesia. *Sains Malaysiana*, 48(8): 1787–1798.
- Suhartono, Lee, M. H., & Prastyo, D. D. (2015). Two levels ARIMAX and regression models for forecasting time series data with calendar variation effects. *AIP Conference Proceedings*, 1691(1), 050026. AIP Publishing LLC.
- Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50: 159–175.
- Zhang, G. P., & Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, 160(2): 501–514.