Jurnal Matematika UNAND
Vol. **15** No. **3** Hal. 302 – 318
Edisi Juli 2026
ISSN : 2303–291X
e-ISSN : 2721–9410

https://doi.org/10.25077/jmua.15.3.302-318.2026

CALENDAR VARIATION TIME SERIES ANALYSIS FOR FORECASTING THE NUMBER OF DOMESTIC PASSENGERS OF JUANDA INTERNATIONAL AIRPORT SURABAYA

PUTRIAJI HENDIKAWATI*, ANISA OKTAVIANI
Applied and Computational Statistics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang

Received May 3, 2025, Received in revised form January 27, 2026
Accepted May 2, 2026 Available online July 31, 2026

Abstract. This study focuses on forecasting the number of domestic transportation passengers at Juanda International Airport using a time series model by including exogenous variables of calendar variation effects of Eid Al-Fitr and Eid Al-Adha holidays. Data of the number of domestic passengers at Juanda International Airport from the Statistics Indonesia (BDS) was analyzed for the period January 2014 to May 2024. The models used include Dummy Regression, Trend Regression, ARIMA, SARIMA, ARIMAX, and SARIMAX. The analysis results show that the ARIMAX model produces the most accurate prediction. Juanda Airport passenger data is optimal in the ARIMAX ([1,11],1,1) model with the exogenous variable Eid al-Fitr dummy which produces MAPE:7.49; RMSE: 41,133.26; and MAE: 34.119,06. The results of the analysis show that including exogenous variables in the form of calendar variations in the forecasting model can improve prediction accuracy.

Keywords: Calendar variation, Time series, Domestic passengers, Airport

1. Introduction

Air transportation has become a supporting factor for community mobility and economic growth in Indonesia with a geographical condition consisting of thousands of islands. Indonesia, as the world's largest archipelago, faces significant challenges in managing the mobility of a population spread across more than 17,000 islands. Juanda International Airport in Sidoarjo, East Java, one of the busiest airports in Indonesia, not only serves the air transportation needs in the East Java region, but also has the function of connecting eastern Indonesia with major economic centers on the island of Java and beyond. Over the past few years, the growth in the number of passengers at Juanda International Airport has increased significantly. This

*Corresponding author

growth has been driven by various factors, including increased economic activity in East Java, the development of the tourism sector, and increased public awareness of the importance of fast and efficient mobility.

Data of the Central Bureau of Statistics (BPS) from year to year shows significant fluctuations in the number of passengers at Juanda International Airport, especially during certain periods such as Eid al-Fitr and Eid al-Adha. In Indonesia, Eid al-Fitr and Eid al-Adha are two important moments where the number of air travel tends to increase significantly. During this period, many people travel back to their hometowns to celebrate the holidays with their families. The National Police report shows an increase in traffic accidents during the homecoming season, with a total of 138,673 cases in 2022, of which 29,679 cases occurred in East Java, making sea and air transportation a safer option. BPS recorded a significant upward trend in passengers at Juanda International Airport. The number of passengers reached 7,719,572 people in 2019, decreased dramatically to 2,283,671 in 2020 due to the COVID-19 pandemic, but increased again to 4,500,000 in 2022 [1], [2], [3]. This increase in passenger demand has led to congestion at airports, flight delays, and increased pressure on airport infrastructure. Faced with these challenges, accurate forecasting of passenger numbers is important for Juanda International Airport managers, airlines, and policy makers.

Forecasting aims to predict the future as accurately as possible, taking into account all available information, including historical data and knowledge of future events that might affect the forecast [4]. The Autoregressive Integrated Moving Average (ARIMA) model is one of the most widely used forecasting methods in time series data analysis. This model is known for its ability to identify and model patterns in historical data to predict future values. When faced data with seasonal elements, the Seasonal ARIMA (SARIMA) model is one option. An empirical investigation carried out and found that the ARIMA and SARIMA model offers a simple and adaptable way to execute forecasting, with accuracy comparable to other more complex prediction techniques. However, in the face of the added complexity of calendar variations and other external factors, ARIMA or SARIMA models are not good enough to predict Tanjung Perak passenger [5], fish production [6], steamflow [7], application to employment [8], electricity demand [9], and Covid-19 epidemic [10]. Therefore, the ARIMAX (ARIMA with Exogenous Variables) model was chosen as a solution to predict data affected by calendar variations. ARIMAX not only retains the ability of ARIMA in analyzing time series data, but also allows the addition of exogenous variables that can affect the number of passengers so that the model can account for passenger spikes that occur during those periods.

Calendar variation is an important factor that needs to be taken into account in modeling data, especially time series data that is affected by national holiday events. Calendar variation refers to seasonal fluctuations that occur during certain periods of the year, such as during religious holidays and vacation seasons. Several studies have considered the effect of calendar variation in modeling data and produced more accurate predictions. Forecasting the number of tourist visitors [11] and search trends for the keyword on Google Trends data [12] showed a high level of accuracy. Another study found that ARIMAX was more effective in forecasting the number of passengers at Makassar Port, with results showing seasonal patterns related to holidays [13]. The development of the SARIMAX model also provides positive results, as seen in the research for the prediction of TransJakarta passengers [14], rainfall forecasting [15], and passanger of Tanjung Priok port [16]. Meanwhile, [17] showed that the ARIMAX model has a higher accuracy rate than the ARIMA and SARIMA models in predicting data affected by calendar variations, with an accuracy rate of up to 90%, and recommended the use of ARIMAX for future research due to its ability to consider external factors such as calendar variations.

Calendar variations also play an important role in influencing passenger numbers at Juanda International Airport. This study aims to analyze and predict the number of domestic transportation passengers in Juanda International Airport using the time series method by considering the effect of holiday calendar variations. The calendar variation variable used to increase the accuracy of the forecasting results. The results are expected to show the importance of considering the holiday calendar effect on time series data and the results can provide useful information for Juanda International Airport managers in capacity planning and service improvement.

2. Some Concepts

2.1. *Calender Variation*

Calendar variation can be seen as a recurring pattern that has different periods due to changes in the length of the month each month due to the influence of events that occur on different dates each year. There are two calendar variations, namely the trading day effect and the holiday effect [18], [19], and [20]. Bell [20] examined modeling time series data with calendar variation. Holidays are caused by important cultural or religious celebrations such as Eid al-Fitr, Easter, and Chinese New Year, which can affect business activities and customer behavior. In Indonesia, a predominantly Muslim country, the influence of the calendar is also seen during religious holidays, which is most noticeable during the Eid al-Fitr i and Eid al-Adha holidays. The Islamic (Hijri) calendar uses the circulation of the moon as a reference, so the beginning of the day and date in the Hijri Calendar is different from the day and date in the Gregorian calendar. As a result, a year in the Hijri Calendar is about eleven days shorter than a year in the Gregorian calendar. In this study, the exogenous variable used is a dummy variable, namely the effect of calendar variation on Eid al-Fitr and Eid al-Adha.

2.2. *Dummy Regression*

Dummy regression is a technique in linear regression that includes categorical variables as predictors. Dummy variables are usually used to represent different groups of categories, or conditions [21]. Dummy regression uses categorical variables, where categorical variables with k categories are represented using $k-1$ dummy variables. Each dummy variable has a value of 0 or 1 to indicate whether an observation belongs to a particular category or not. Linear regression model equation with the use of dummy variables.

$$Y_i = \beta_0 + \beta_1 X_1 i + \cdots + \beta_k X_k i + y_1 D_1 i + y_2 D_2 i + \cdots + y_m D_m i + \epsilon_i, \qquad (2.1)$$

with Y_i dependent variable (respondent) n the i^{th} observation, β_0 intercept (constant) of the model, $\beta_1, \beta_2, \cdots, \beta_k$ coefficient for continuous independent variable $X_1, X_2, \cdots, X_k$, $D_{1i}, D_{2i}, \cdots, D_{mi}$ dummy variable on the i^{th} observation, This dummy variable takes the value 0 or 1 to indicate membership in a particular category, $Y_1, Y_2, \cdots, Y_m$ the coefficient for the dummy variable, which indicates the average difference in the dependent variable between different categories, and ϵ_i error or residual at the i^{th} observation.

2.3. *Trend Regression*

Trend regression is a method used in time series analysis to identify and quantify long-term trends in data. Modeling in the context of time series, the occurrence of a trend refers to a systematic change in the value of the dependent variable over time. Trend regression helps in separating the trend component from seasonal and random variations in the data [22]. Trend regression is described by the following equation:

$$Y_t = \beta_0 + \beta_1 t + \epsilon_t, \qquad (2.2)$$

with Y_t the value of the dependent variable at time t, β_0 the intercept or constant of the model, which indicates the initial value of the trend, β_1 trend coefficient, which measures the rate of change or slope of the trend over time, t time or period often expressed in units of time such as months, years, or quarters, and ϵ_t error or residuals at time t, which includes random variation not explained by the trend model.

2.4. *ARIMA*

The Autoregressive Integrated Moving Average (ARIMA) model combines three processes, namely autoregressive, differencing, and moving average. The general format for the ARIMA model is (p,d,q), where p is the order of the AR operator, d is the differencing order, and q is the order of the MA operator [23]. The ARIMA model is expressed with the following equation:

$$\varnothing_p(B)(1-B)^d \hat{Z}_t = \mu + \theta_q(B) a_t, \qquad (2.3)$$

with $\varnothing_p$ coefficient of p^{th} autoregressive parameter, θ_q coefficient of the q^{th} autoregressive parameter, B backshift operator, D differencing, μ constant, a_t error value at time t, p autoregressive degree, d level of differencing process, q degree of moving average.

2.5. *SARIMA*

The Seasonal Autoregressive Integrated Moving Average or (SARIMA) model combines non-seasonal and seasonal factors in one multiplicative model. The SARIMA notation is expressed as ARIMA (p,d,q) x $(P,D,Q)^S$, where p,d,q indicates the order for the seasonal component, and S is the seasonal period. This model can be written with the following equation:

$$\phi_p(B)\phi(B^s)(1-B)^d(1-B^s)^d\hat{Z}_t = \mu + \theta_q(B)\Theta_Q(B^s)\alpha_t, \tag{2.4}$$

with $\hat{Z}_t$ observation at time t, $\phi_p(B)$ non-seasonal AR component of order p, $\phi_p(B^s)$ p-order seasonal AR component, $(1-B)^d$ non-sensasional differencing of order d, $(1-B^s)^D$ seasonal differencing of order d, $\theta_q(B)$ non-seasonal MA component of order q, $\Theta_Q(B^s)$ seasonal MA component of order q, α_t residuals at time t.

2.6. *ARIMAX*

Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) is a time series model that extends the ARIMA model by including one or more exogenous variables [24]. One of the frequently used exogenous variables is the calendar variation effect variable, which takes into account the influence of certain events such as holidays or trading days on time series data. In this study, the exogenous variable used is a dummy variable, namely the effect of calendar variations on Eid al-Fitr and Eid al-Adha holidays. The ARIMAX method has several models, namely AR, MA, ARMA, ARIMA, and ARIMAX. In general, the ARIMAX model equation can be expressed as follows [25].

$$(1-B)^d\phi_p(B)\hat{Z}_t = \mu + \theta_q(B)\alpha_t + \beta_1 X_{1t} + \cdots + \beta_k X_{kt}, \tag{2.5}$$

with Z_t response variable, α_p p^{th} autoregressive parameter, Y_t q^{th} autoregressive parameter, $\beta_1, \cdots, \beta_1$ coefficients of exogenous variables.

The stages in ARIMAX are the same as those in ARIMA modeling, which include parameter estimation and parameter significance testing, residual diagnostic checks, and forecasting the ARIMAX model by taking into account the effects of calendar variation.

2.7. *SARIMAX*

The SARIMAX model is an extension of the ARIMAX model that contains seasonal elements. In general, the equation for the SARIMAX model can be written as follows [26].

$$\phi_p(B)\phi_p(B^s)\phi_p(1-B)^d(1-B^s)^D\hat{Z}_t = \mu+\theta_q(B)\Theta_Q(B^s)\alpha_t+\alpha_1 X_{1,t}+\alpha_2 X_{2,t}+\cdots+\alpha_k X_{k,t} \tag{2.6}$$

with $\hat{Z}_t$ observation value at time t, ϕ_p, ϕ_P p^{th} AR parameter, B backshift operator, θ_q, Θ_Q p^{th} MA parameter, α_t residuals at time t, μ' constant.

2.8. *Model Accuracy Evaluation*

Model evaluation is a crucial stage in time series analysis that aims to identify the best model among the various model options obtained [16]. This process compares the performance of each model to find the most accurate and efficient by looking at the error value, the most accurate model is the model with the smallest error. This aims to produce future predictions that are optimal and close to actual data. The accuracy of the forecasting model in this study is evaluated using the error measures RMSE, MAPE, and MAE [4].

(a) Root Mean Square Error (RMSE)
RMSE or root mean square error is used to measure the error of the model's estimated value expressed as the root mean square error. The RMSE formula is expressed by the equation:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Z_t - \hat{Z}_t)^2}. \tag{2.7}$$

RMSE is used to compare several estimation models of the same time series. A good model has a small RMSE, because the model will be more suitable or closer to the existing data. Models with a smaller RMSE tend to have a smaller forecast error variance, which means that the variation in the value produced by a forecasting model is close to the variation in the observation value. The smaller the RMSE value, the closer the predicted and observed values.

(b) Mean Absolute Percentage Error (MAPE)
MAPE is used to measure the error of the estimated value of the model expressed in the form of an absolute average percentage error. The formula for determining the MAPE value is expressed by the equation:

$$MAPE = 100 \times \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Z_t - \hat{Z}_t}{Z_t}\right|. \tag{2.8}$$

MAPE is useful for providing information on how much the forecasting error is compared to the actual value. MAPE measures the accuracy of the forecast produced by a model. The smaller the percentage error value in MAPE, the more accurate the forecasting results.

(c) Mean Absolute Error (MAE)
Mean Absolute Error evaluater the average absolute difference between actual and forecasted values in a statistical model. The MAE formula is expresses by the equation:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|Z_t - \hat{Z}_t\right| \tag{2.9}$$

MAE gives results according to the original data units so that they are easy to interpret. MAE gives equal weight to all errors including outliers.

2.9. *Data and Research Variables*

This study uses secondary data, namely monthly data on the number of airplane passengers departing domestic flights at Juanda International Airport. This data covers the period from January 2014 to May 2024, with a total of 125 observations obtained from the official website of the Statistics Indonesia (BPS) at www.bps.go.id. This study integrates dummy variables to capture the calendar variation effects of Eid al-Fitr and Eid al-Adha holidays that affect passenger travel demand, which often trigger a surge in demand, as well as holiday seasons and other peak travel periods such as school holidays or major festivals. This research involves dividing the dataset into training and testing data with a ratio of 70:30. The variable Number of Passenger Departures is the response variable, while the exogenous variables are calendar variations in the form of Eid al-Fitr and Eid al-Adha holidays. Determination of exogenous variables is using (i) a dummy variable to declare the Eid al-Fitr and Eid al-Adha holiday and (ii) ratio of calendar variations refer to [27].

2.10. *Data Analysis*

This research was conducted in several stages as follows.

2.10.1. *Descriptive Analysis*

Descriptive analysis was conducted on the number of domestic passengers at Juanda International Airport. This stage provides an overview of data characteristics, including, mean, median, mode, standard deviation, minimum value, and maximum value. This information helps researchers to understand the distribution of the data.

2.10.2. *Data Modelling*

(i) Stationary Test in Variance
Stationarity in variance was identified using the Bartlett test. If the data is not stationary, then data transformation is performed.

(ii) Stationary Test in Mean
The stationarity check in the mean is done through the ADF (Augmented Dickey-Fuller) test. If the data is not yet stationary at the level, then differencing is performed.

(iii) Identification Model
Form ACF and PACF plots on data that has been stationary in variance and mean.

(iv) Model Parameter Estimation
In parameter estimation to determine the suspected model of dummy regression and trend regression based on a probability value ≤ 0.05. Meanwhile, the ARIMA and SARIMA models are based on significant lags and auto ARIMA results. ARIMAX and SARIMAX require the addition of dummy variables in the modeling process. However, the determination of the estimation model and model selection for ARIMAX and SARIMAX still use the same approach as ARIMA and SARIMA, namely based on significant lags, auto ARIMA, and the lowest AIC value.

(v) Diagnostic Test
Perform residual diagnostic tests, including white noise test, normality test, and heteroscedasticity test on the selected model.

(vi) Best Model Selection
Determine the best model that meets the diagnostic test and has the lowest AIC value.

(vii) Model Accuracy Evaluation
Evaluate the accuracy of each model by looking at the smallest RMSE, MAPE, and MAE values.

(viii) Forecasting
Forecasting the number of aircraft passengers at Juanda International Airport for the period January 2024 – June 2024 .

3. Result and Discussion

3.1. *Research Results*

3.1.1. *Descriptive Analysis of the Number of Domestic Passenger of Juanda International Airport*

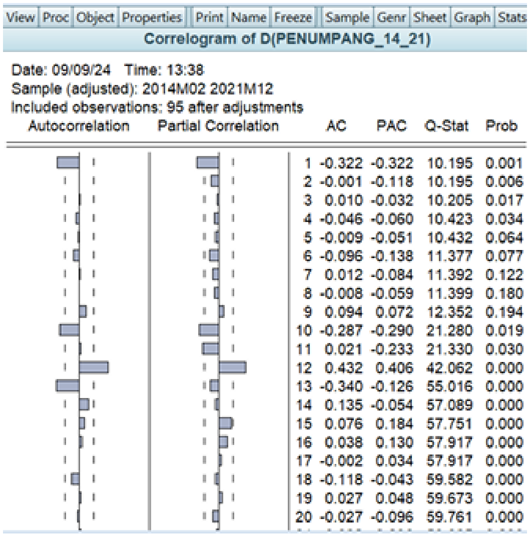
Figure. 1. Domestic Aircraft Passenger Data of Juanda International Airport (Jan 2014 - May 2024)

The data used in this study includes 125 months of airplane passenger data. The data is divided with a proportion ratio of 70:30 which is trainng data starting from January 2014 to December 2021 and testing data from January 2022 to May 2024. Figure 1 shows that the data on the number of airplane passengers is quite stable, up and down regularly every year. However, in 2020 to 2021, there was a drastic decline due to the COVID-19 pandemic. The average number of domestic passengers at Juanda International Airport is 488,805 passengers per month and the median is 521,681. The range of passenger numbers ranged from 5,397 to 899,548 passengers, with a standard deviation of 186,574.8 indicating large variations between months. The negative skewness (-0.504) and kurtosis of 2.664 indicate a near-normal distribution, and the Jarque-Bera test results support a near-normal distribution (probability 0.052). The data shows significant fluctuations in the number of passengers, especially during vacation and holiday events.

3.1.2. *Data Modelling of Juanda International Airport Domestic Passenger Data*

(1) Stationarity Test
The initial step in identifying time series modeling is to test the stationarity of the data on variance and mean. Stationarity testing is carried out using the Dicky Fuller (ADF) test which shows that the data is not yet stationary and requires one differencing process.

(2) Model Identification
The ACF plot in Figure 2 shows cut-offs at lags 1, 10, 12, and 13, indicating the presence of moving average (MA) elements, while the PACF shows cut-offs at lags 1, 10, 11, and 12, indicating the strong influence of the autoregressive (AR) component at those lags.



	AC	PAC	Q-Stat	Prob
1	-0.322	-0.322	10.195	0.001
2	-0.001	-0.118	10.195	0.006
3	0.010	-0.032	10.205	0.017
4	-0.046	-0.060	10.423	0.034
5	-0.009	-0.051	10.432	0.064
6	-0.096	-0.138	11.377	0.077
7	0.012	-0.084	11.392	0.122
8	-0.008	-0.059	11.399	0.180
9	0.094	0.072	12.352	0.194
10	-0.287	-0.290	21.280	0.019
11	0.021	-0.233	21.330	0.030
12	0.432	0.406	42.062	0.000
13	-0.340	-0.126	55.016	0.000
14	0.135	-0.054	57.089	0.000
15	0.076	0.184	57.751	0.000
16	0.038	0.130	57.917	0.000
17	-0.002	0.034	57.917	0.000
18	-0.118	-0.043	59.582	0.000
19	0.027	0.048	59.673	0.000
20	-0.027	-0.096	59.761	0.000

Figure. 2. ACF and PACF Plots

(3) Parameter Estimation Model

(a) Dummy Regression
Dummy regression parameter estimation is conducted by including dummy variables and ratios representing the Eid al-Fitr and Eid al-Adha holidays. Dummy variables are used as binary indicators (0 or 1) indicating the presence or absence of these holidays, while ratios or variable weights are used to adjust the effect of these holidays based on the intensity or duration of each event. The best model is done by comparing several alternative models as shown in Table 1. Table 1 shows that some parameters are significant, as the probability of some variables is ≤ 0.05.

(b) Trend Regression
Trend Regression parameter estimation is performed to evaluate the im-

Table 1. Estimated Parameter Dummy Regression Models

Exogenous Variables	Coeff	Prob	AIC	Description
Eid al-Fitr Dummy	504505,1	0,0008	29,18	Passed
Eid al-Fitr Ratio	556871,4	0,0057	29,22	Passed
Eid al-Adha Dummy	472433,0	0,0063	29,22	Passed
Eid al-Adha Ratio	556332,1	0,0066	29,22	Passed

pact of long-term trends on the data. By using trend variables, the model can identify consistent patterns of growth or decline over time. The best model is done by comparing several alternative models as shown in Table 2.

Table 2. Estimation and Significance Test of Trend Regression Parameters

Exogenous Variables	Coeff	Prob	AIC	Description
Trend, Eid al-Fitr Dummy	6508,184	0,0000	28,64	Not Passed
	206129,9	0,0790		
Trend, Eid al-Fitr Ratio	6651,600	0,0000	28,66	Not Passed
	215184,0	0,1630		
Trend, Eid al-Adha Dummy	6705,926	0,0000	28,67	Not Passed
	127077,8	0,3443		
Trend, Eid al-Adha Ratio	6678,407	0,0000	28,66	Not Passed
	186220,9	0,2376		
Trend, Eid al-Adha Dummy	6733,813	0,0000	28,68	Not Passed
and Eid al-Adha Ratio	-207608,0	0,6105		
	416562,0	0,3851		
Trend, Eid al-Fitr Dummy	6508,416	0,0000	28,66	Not Passed
and Eid al-Fitr Ratio	209639,6	0,2882		
	-5720,635	0,9823		

Table 2 shows that the parameters are not significant, because the probability of all variables ≥ 0.05.

(c) ARIMA Model

ARIMA model estimation is done after checking data stationarity and model identification using ACF and PACF plots. Determining the best model is done by comparing several alternative models as shown in Table 3.

Table 3 shows that the tested ARIMA models have significant parameters with probability values less than 0.05.

(d) SARIMA Model

The SARIMA model was chosen in consideration of the recurring patterns that occur in the annual Eid al-Fitr and Eid al-Adha holiday events. The SARIMA model estimation is carried out using the ARIMA model reference by adding a seasonal component. Determining the best SARIMA model is done by comparing several alternative models as shown in Table 4.

Table 3. Estimation and Significance Test of ARIMA Parameters

Model	Parameters	Coeff	Prob	AIC	Description
(1,1,0)	AR(1)	0,992617	0,0000	25,925	Passed
	d(1)	0,496254	0,0000		
(1,1,1)	AR(1)	0,982200	0,0000	24,696	Passed
	d(1)	0,500649	0,0000		
	MA(1)	0,928148	0,0000		
([10],1,0)	AR(10)	0,940158	0,0000	27,148	Passed
	d(1)	0,505244	0,0000		
([1,11],1,0)	AR(1)	0,905669	0,0000	24,885	Passed
	AR(11)	0,090422	0,0288		
	d(1)	0,496892	0,0000		
([1,11],1,1)	AR(1)	0,786637	0,0000	24,586	Passed
	AR(11)	0,195039	0,0004		
	d(1)	0,500721	0,0000		
	MA(1)	0,924620	0,0000		
([1,12],1,1)	AR(1)	0,786637	0,0000	24,588	Passed
	AR(12)	0,206711	0,0017		
	d(1)	0,500569	0,0000		
	MA(1)	0,951695	0,0000		

(e) ARIMAX Model
The ARIMAX model was chosen in consideration of the effects of calendar variations in the Eid al-Fitr and Eid al-Adha holidays. The ARIMAX modeling stages are the same as ARIMA but by adding exogenous variables. Determining the best ARIMAX model is done by comparing several alternative models as shown in Table 5.

(f) SARIMAX Model
The SARIMA model was chosen in consideration of the recurring patterns that occur in the annual Eid al-Fitr and Eid al-Adha holiday events by adding the effect of calendar variation of Eid al-Fitr and Eid al-Adha holidays. The estimation and significance test of the SARIMAX model parameters are presented in Table 6.

(4) Diagnostic Test
Based on the estimation results and significance tests of the six models in the previous stage, the best model estimate with the smallest AIC value was selected and further evaluated through the residual diagnostic test. The results of this test can be seen in Table 7.

(5) Selection of the Best Model
The selection of the best model is based on the results of the residual diagnostic test, taking into account the most suitable model that meets the statistical assumptions. The best model of Juanda International Airport Domestic Passenger are presented in Table 8.

(6) Model Accuracy Evaluation
Evaluation of model accuracy is done by comparing the MAPE, RMSE, and MAE values of each of the best models previously obtained. The model with

Table 4. Estimation and Significance Test of SARIMA Parameters

Model	Parameters	Coeff	Prob	AIC	Description
$(1,1,[10])^{12}$ (1,1,[2,12])	d(1)	0,50	0,0000	24,35	Passed
	AR(1)	0,59	0,0000		
	MA(10)	-0,49	0,0002		
	SAR(1)	0,99	0,0000		
	SMA(2)	-0,62	0,0000		
	SMA(12)	0,40	0,0001		
$(1,1,[10,12])^{12}$ (1,1,[2])	d(1)	0,50	0,0000	24,40	Passed
	AR(1)	0,67	0,0000		
	MA(10)	-0,36	0,0085		
	MA(12)	0,44	0,0011		
	SAR(1)	0,99	0,0000		
	SMA(2)	-0,73	0,0000		
$(([1,11],1,[11])^{12}$ (1,1,[2]))	d(1)	0,50	0,0000	24,410	Passed
	AR(1)	0,52	0,0000		
	AR(11)	0,47	0,0000		
	MA(11)	-0,99	0,0004		
	SAR(1)	0,99	0,0000		
	SMA(2)	-0,60	0,0000		
$((1,1,[10])^{12}$ ([1,12],1,[2]))	d(1)	0,50	0,0000	24,415	Passed
	AR(1)	0,99	0,0000		
	MA(10)	-0,51	0,0001		
	SAR(1)	0,54	0,0000		
	SAR(12)	0,30	0,0001		
	SMA(2)	-0,62	0,0001		
$((1,1,[12])^{12}$ (1,1,[2]))	d(1)	0,50	0,0000	24,42	Passed
	AR(1)	0,99	0,0000		
	MA(12)	0,50	0,0001		
	SAR(1)	0,73	0,0000		
	SMA(2)	-0,85	0,0000		
$([1,11],1,1)^{12}$ (1,1,1)	d(1)	0,50	0,0000	24,43	Passed
	AR(1)	0,56	0,0014		
	AR(11)	0,21	0,0095		
	MA(1)	-0,84	0,0000		
	SAR(1)	0,99	0,0000		
	SMA(1)	0,93	0,0000		

Table 5. Estimation and Significance Test of ARIMAX Parameters

Model	Parameters	Coeff	Prob	AIC	Description
([1,11],1,1)	AR(1)	0,869140	0,0000	24,42	Passed
	AR(11)	0,125907	0,0142		
	d(1)	0,496709	0,0000		
	MA(1)	0,597060	0,0000		
	Eid al-Fitr Dummy	-51625,93	0,0000		
([1,11],1,1)	AR(1)	0,848654	0,0000	24,55	Passed
	AR(11)	0,145705	0,0130		
	d(1)	0,494617	0,0000		
	MA(1)	0,645871	0,0145		
	Eid al-Fitr Ratio	-26911,81	0,0036		

Table 6. Estimation and Significance Test of SARIMAX Parameters

Model	Parameters	Coeff	Prob	AIC	Description
$(1,1,[10])(1,1,[2,12])^{12}$	d(1)	0,49	0,0000	24,34	Passed
	AR(1)	0,99	0,0000		
	MA(10)	-0,45	0,0001		
	SAR(1)	0,57	0,0000		
	SMA(2)	-0,57	0,0000		
	SMA(12)	0,32	0,0007		
	Eid al-Fitr Dummy	-20956,38	0,0378		
$(1,1,[10])(1,1,[2,12])^{12}$	d(1)	0,49	0,0000	24,33	Passed
	AR(1)	0,99	0,0000		
	MA(10)	-0,41	0,0037		
	SAR(1)	0,57	0,0000		
	SMA(2)	-0,58	0,0001		
	SMA(12)	0,46	0,0003		
	Eid al-Fitr Ratio	-13770,3	0,0436		

Table 7. Diagnostic Test of Dummy Regression

Model	Exogenous Variable	White Noise	Heteroscedasticity	Normality
Dummy Regression	Eid al-Fitr Dummy	No	Heteroscedasticity	No
Trend Regression		Does not pass the significance test		
ARIMA([1,11],1,1)	-	No	Homoskedasticity	Close to normal
SARIMA $(1,1,[10])(1,1,[2,12])^{12}$	-	Yes	Homoskedasticity	Yes
ARIMAX([1,11],1,1)	Eid al-Fitr Dummy	Yes	Homoskedasticity	Yes
SARIMAX $(1,1,[10])(1,1,[2,12])^{1}2$	Eid al-Fitr Ratio	Yes	Homoskedasticity	Yes

Table 8. The Best Model

Model	Exogenous Variable	AIC	Diagnostic Test
Dummy Regression	Eid al-Fitr Dummy	29.18	No
Trend Regression	Eid al-Fitr Dummy	28.64	No
ARIMA([1,11],1,1)	-	24.59	Yes
SARIMA$(1,1,[10])(1,1,[2,12])^{12}$	-	24.35	Yes
ARIMAX([1,11],1,1)	Eid al-Fitr Dummy	24.42	Yes
SARIMAX $(1,1,[10])(1,1,[2,12])^{12}$	Eid al-Fitr Ratio	24.33	Yes

the lowest MAPE, RMSE, and MAE values in all three metrics is selected as the best. This selected model is then used for forecasting in the next stage. The evaluation metrics of the six best models are presented in Table 9.

Based on the evaluation metrics in Table 8, the ARIMAX ([1,11],1,1) model with Eid dummy shows the best overall performance. This model has the lowest MAPE, RMSE, and MAE values. These values confirm that this model provides predictions with the lowest error rate compared to other models. Therefore, the ARIMAX ([1,11],1,1) model with Eid dummy is chosen as the best model and

Table 9. Evaluation Metrics

Model	Exogenous Variables	Model Evaluation Metrics		
		MAPE	RMSE	MAE
Dummy Regression	Eid al-Fitr Dummy	86,81	364410,6	338140,1
Trend Regression				
ARIMA ([1,11],1,1)	-	52,25	71036,44	54088,37
SARIMA $(1,1,[10])(1,1,[2,12])^{12}$	-	40,63	64168,77	49175,70
ARIMAX ([1,11],1,1)	Eid al-Fitr Dummy	7,49	41133,26	34119,06
SARIMAX $(1,1,1[10])(1,1,[2,12])^{12}$	Eid al-Fitr Ratio	11,07	53306,27	45579,46

is more suitable for forecasting the number of domestic passengers at Juanda International Airport.

(7) Forecasting the Number of Domestic Passengers at Juanda International Airport
Forecast results for the next six months from January to June 2024 show an increasing trend in the number of domestic passengers at Juanda International Airport, namely 51,404, 60,280, 51,644, 54,194, 58,316, and 76,155 respectively. These results can be used as a reference for passenger predictions in the following months.

3.2. *Discussion*

Surabaya, as the capital of East Java province and the second largest city in Indonesia, has experienced rapid economic growth, which in turn has led to increased demand for air transportation services. Juanda International Airport, which serves millions of passengers annually, is an important indicator of economic development in East Java. The dynamic growth of passenger numbers at Juanda International Airport is inseparable from various external factors. One of the biggest challenges in inter-regional transportation activities is the impact of external factors such as the COVID-19 pandemic, which has not only shaken the aviation industry but also the transportation industry globally. At its peak in 2019, Juanda International Airport recorded a very high number of passengers, but this number dropped dramatically in the following years due to travel restrictions and changes in public behavior that are more cautious about air travel. This decline in passenger numbers not only affects airport and airline revenues, but also poses challenges for future airport capacity planning and management.

This research focuses on forecasting the number of domestic transportation passengers at Juanda International Airport using a time series model that considers calendar variations of the Eid al-Fitr and Eid al-Adha holidays. Modeling is done Dummy Regression, Trend Regression, ARIMA, SARIMA, ARIMAX, and SARIMAX. Each model was tested with various parameters and combinations of dummy variables representing Eid al-Fitr and Eid al-Adha, both in the form of binary indicators and in the form of ratios. The use of complex models, such as ARIMAX and SARIMAX, aims to capture specific calendar variations so as to illustrate the effect of holiday calendar variations on passenger numbers.

Some forecasting models show the results of predicting the number of passengers to be less accurate with a fairly high prediction error. This is partly due to changes in the data on the number of airplane passengers which fell drastically during the Covid-19 pandemic. However, the ARIMAX ([1,11],1,1) model with Eid al-Fitr dummy shows the best performance compared to other models. Compared to other models, this model can minimize the prediction error as seen in Table 9. The surge of passengers coinciding with Eid al-Fitr, which reflects the annual pattern of homecoming and return flows, can be better anticipated using the ARIMAX model. These two holidays are the peak points of passenger movement every year, so the selection of a model that takes into account calendar effects is expected to provide more accurate forecasting results. It can be seen that the variation of the Eid holiday has a significant influence because this holiday often triggers a sharp increase in the number of passengers due to the tradition of going home. During Eid al-Fitr, there is a high movement of passengers at the airport as the majority of people in East Java and surrounding areas travel home or on vacation. Therefore, this ARIMAX model not only shows good prediction accuracy, but is also relevant in airport operational planning, such as setting flight capacity, manpower, and additional services during passenger-intensive periods. The model shows that including calendar variations in the analysis can provide sharper information and reduce the prediction error rate,allowing airports to better prepare strategies when facing passenger surges.

The results show that the ARIMAX model that includes the exogenous variable of holiday variations consistently provides more accurate forecasting results compared to other models that do not consider calendar variations. ARIMAX proved to be superior in capturing fluctuations influenced by calendar variations. Other models, such as Dummy Regression and Trend Regression, have limitations in capturing complex seasonal patterns, especially when calendar holiday variations exert a large influence. The ARIMA and SARIMA models yet provide less accurate prediction results because they cannot accommodate exogenous variables optimally like ARIMAX.

3.3. *Conclusion*

This study shows that the ARIMAX model with dummy variables for holidays such as Eid al-Fitr and Eid al-Adha is the most effective for predicting the number of passengers at Juanda International Airport. The ARIMAX ([1,11],1,1) model with Eid al-Fitr dummy produces accurate predictions for passenger spikes during Eid al-Fitr. Overall, the ARIMAX model proved to be more accurate than other models because it considers calendar variations, assisting managers in planning capacity and improving operational readiness during peak passenger periods. The results of this study indicate the importance of using ARIMAX models in situations where forecasting is affected by calendar variations.

Bibliography

[1] BPS, 2023a, *Kota Surabaya Dalam Angka 2023*, Bappeda Potensi Wilayah, Vol. 4, Issue 1, http://bappeda.jatimprov.go.id/bappeda/wp-content/uploads/potensi-kab-kota-2013/kota-surabaya-2013.pdf

[2] BPS, 2023c, *STATISTIK TRANSPORTASI LAUT 2022*, Vol. **8**. http://repositorio.unan.edu.ni/2986/1/5624.pdf

[3] BPS, 2023b, *STATISTIK INDONESIA 2023*, Statistik Indonesia 2023, Vol. 1101001. https://www.bps.go.id/publication/2020/04/29/e9011b3155d45d70823c141f/statistik-indonesia-2020.html

[4] Hyndman, R.J. & Athanasopoulos, G., 2021, *Forecasting: Principles and Practice*. 3rd Edition, Otexts Publishing.

[5] Putri, S., & Sofro, A., 2022, Peramalan Jumlah Keberangkatan Penumpang Pelayaran Dalam Negeri di Pelabuhan Tanjung Perak Menggunakan Metode ARIMA dan SARIMA, *MATHunesa: Jurnal Ilmiah Matematika*, Vol. **10** Issue 1: 61 – 67, https://doi.org/10.26740/mathunesa.v10n1.p61-67

[6] Putra, E.F., Asdi, Y. & Maiyastri, M., 2019, Peramalan Dengan Metode Pemulusan Eksponensial Holt-Winter Dan SARIMA (Studi Kasus: Jumlah Produksi Ikan (Ton) di Kota Sibolga Tahun 2000-2017), *Jurnal Matematika UNAND*, Vol. **8**(1): 75 – 83

[7] Adnan, R.M., Yuan, X., Kisi, O. & Yuan, Y., 2017, Streamflow forecasting of Astore River with Seasonal Autoregressive Integrated Moving Average Model, *Eur Sci J*, Vol. **13**(12): 145 – 156

[8] Wang, X. and Liu, Y., 2009, ARIMA Time Series Application to Employment Forecasting, *In 2009 4th International Conference on Computer Science & Education*, pp. 1124 – 1127, IEEE

[9] Wang, Y., Wang, J., Zhao, G. & Dong, Y., 2012, Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: A case study of China, *Energy Policy*, Vol. **48**: 284 – 294

[10] Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S. & Ciccozzi, M., 2020, Application of the ARIMA Model On The COVID-2019 Epidemic Dataset, *Data in Brief*, Vol. **29**: 105340

[11] Berlinditya, B., & Noeryanti, N. , 2019, Pemodelan Time Series Dalam Peramalan Jumlah Pengunjung Objek Wisata Di Kabupaten Gunung Kidul Menggunakan Metode ARIMAX Efek Variasi Kalender, *Jurnal Statistika Industri Dan Komputasi*, Vol. **4**, Issue 1: 81 – 88

[12] Dani, A. T. R., Wahyuningsih, S., Putra, F. B., Fauziyah, M., Wigantono, S., Sandariria, H., A'yun, Q. Q., & Zen, M. A., 2023, Aplikasi Model ARIMAX dengan Efek Variasi Kalender untuk Peramalan Trend Pencarian Kata Kunci "Zalora" pada Data Google Trends, *Inferensi*, Vol. **6**, Issue 2: 107 https://doi.org/10.12962/j27213862.v6i2.15793

[13] Farih, L., Fauziah, I., & Wijaya, M., 2019, Prediction of The Number of Ship Passengers in The Port of Makassar using ARIMAX Method in The Presence of Calendar Variation, *InPrime: Indonesian Journal of Pure and Applied Mathematics*, Vol. **1**, Issue 1: 57 – 67, https://doi.org/10.15408/inprime.v1i1.12786

[14] Qomariyah Virati, M., Paulus Pamanik, D., & Pramana, S., 2020, Forecasting Number of Passengers of TransJakarta using SARIMAX Method, *Open Access J Data Sci Appl*, Vol. **3**, Issue 1: 31 – 37, https://doi.org/10.34818/JDSA.

2020.3.45

[15] Latief, N. H., Nur'eni, N., & Setiawan, I., 2022, Peramalan Curah Hujan di Kota Makassar dengan Menggunakan Metode SARIMAX, *STATISTIKA Journal of Theoretical Statistics and Its Applications*, Vol. **22**, Issue 1: 55 – 63, https://doi.org/10.29313/statistika.v22i1.990

[16] Kartiningtyas, A. N., Zukhronah, E., & Sugiyanto, 2019, Model ARIMAX untuk Meramalkan Banyak Penumpang dari Pelayaran Dalam Negeri di Pelabuhan Tanjung Priok, *Prosiding Sendika*, Vol. **5**, Issue 2: 6 – 15

[17] Syaharuddin, Akmala, Q. S., & Sucipto, L., 2022, Metode ARIMA, ARIMAX, dan SARIMA: Sebuah Meta-Analisis Perbedaan Tingkat Akurasi Peramalan Data Time Series, *Jurnal Informatika Kaputama (JIK)*, Vol. **6**, Issue 3: 502 – 509

[18] Liu, L.M., 1986, Identification of Time Series Models in The Presence of Calendar Variation, *International Journal of Forecasting*, Vol. **2**(3): 357 – 372

[19] Hillmer, S.C., 1982, Forecasting Time Series with Trading Day Variation, *Journal of Forecasting*, Vol. **1**(4): 385 – 395

[20] Bell, W.R., dan Hillmer, S.C., 1983, Modeling Time Series with Calendar Variation, *Journal of the American Statistical Association*, Vol. **78**(383): 526 – 534

[21] Fox, J., 2016, Applied Regression Analyis and Generated Linear Models 3rd Edition, *Journal of Chemical Information and Modeling*, Vol. **53**, Issue 9

[22] Box, G.E.P., Jenkins, G.M., 2015, *Time Series Analysis: Forecasting And Control*, 5th Edition, John Wiley and Sons Inc., Hoboken, New Jersey

[23] Suryani, R. V., Rismawan, T., & Ruslianto, I., 2023, Penerapan Metode Arima Untuk Memprediksi Pemakaian Bandwidth di Universitas Tanjungpura, *Coding Jurnal Komputer dan Aplikasi*, Vol. **10**(03): 421 – 432

[24] Makridakis, S., Wheelwright, S. C., & Hyndman, R. J., 1998, *Forecasting: Methods and Applications*, 3rd Edition, Wiley, https://doi.org/10.2307/2287014

[25] Hendikawati, P., Subanar, S., Abdurakhman, A., & Tarno, T., 2022, Effects of Calendar Variations on the Indonesia Stock Exchange: an Empirical Study of Potential Stocks, *Statmat: Jurnal Statistika dan Matematika*, Vol. **4**(1): 39 – 60, https://doi.org/10.32493/sm.v4i1.14921

[26] Nasirudin, F., & Dzikrullah, A.A., 2023, Pemodelan Harga Cabai Indonesia dengan Metode Seasonal ARIMAX, *Jurnal Statistika dan Aplikasinya*, Vol. **7**(1): 105 – 115, https://doi.org/10.21009/jsa.07110

[27] Shuja, N., Lazim, M. A., & Wah, Y. B., 2007, *Moving Holiday Effects Adjustment for Malaysian Economic Time Series*, Department of Statistics.