BAREKENG: Journal of Mathematics and Its Applications
December 2024 Volume 18 Issue 4 Page 2191–2202
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss4pp2191-2202

EID AL-FITR INFLUENCES THE NUMBER OF TRAIN PASSENGERS ON THE SUMATRA ISLAND (CALENDAR VARIATIONS TIME SERIES MODEL)

Muhammad Sjahid Akbar[1*], Dinda Ayu Safira[2], Rahmi Fadhilah[3], Salma Damayanti[4], Riskianto[5], Khem Puthy[6]

[1,2,3,4,5]Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember
Jln. Teknik Mesin 175 Sukolilo Surabaya, Indonesia

[6]Department of Mathematics, Faculty of Science, Royal University of Phnom Penh
Phnom Penh, Cambodia

Corresponding author's e-mail: **m_syahid_a@statistika.its.ac.id**

ABSTRACT

Article History:

Received: 10th January 2024
Revised: 12th May 2024
Accepted: 8th July 2024
Published: 14th October 2024

Keywords:

ARIMAX;
Eid Al-Fitr;
Regression;
Train.

The Train is one of the transportation options for land travel on the island of Sumatra because of its affordable cost, comfort, and fast mobility. The majority of the population of Sumatra Island who are Muslims influenced the sharp increase in the number of train passengers during Eid Al-Fitr due to the large number of residents who returned to their hometowns (Sumatra Island). The time of Eid Al-Fitr will change every year on the Gregorian calendar, but it is always the same if using the Hijri calendar. This study aims to predict the number of train passengers on the island of Sumatra based on calendar variations (Eid Al-Fitr). To overcome this calendar variation, time series modeling will be used with the addition of exogenous variables (ARIMAX). This model consists of a time series regression equation added with a time series model of the residual regression equation of an exogenous variable. The resulting model can forecast the number of train passengers in the next few months and find that Eid Al-Fitr affects the data. Every Eid Al-Fitr, there is an increase in the number of train passengers by 49 passengers. The model obtained is in the good category with MAPE in-sample of 18.12% and out-sample of 9.93%.

How to cite this article:

M. S. Akbar, D. A. Safira, R. Fadhilah, S. Damayanti, Riskianto and K. Puthy., "EID AL-FITR INFLUENCES THE NUMBER OF TRAIN PASSENGERS ON THE SUMATRA ISLAND (CALENDAR VARIATIONS TIME SERIES MODEL)," *BAREKENG: J. Math. & App.*, vol. 18, iss. 4, pp. 2191-2202, December, 2024.

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

Research Article · Open Access

1. INTRODUCTION

Transportation is essential in modern life, with increasing economic standards and high mobility driving a growing demand for transportation services [1]. Among the various modes of transportation, trains have become a popular choice for many people due to their affordability, convenience, speed, and capacity to transport large numbers of passengers [2]. In Indonesia, PT Kereta Api Indonesia (PT KAI), a State-Owned Enterprise, is responsible for providing and maintaining rail transportation services. To ensure passenger comfort, PT KAI continuously works to improve its facilities and infrastructure.

The number of train passengers fluctuates over time, with noticeable spikes during significant periods such as Eid al-Fitr and the year-end holidays in December. These fluctuations necessitate accurate forecasting by PT KAI to ensure adequate train facilities are available to meet passenger demand. The surge in passengers at the end of each year is identified as a seasonal pattern in time series analysis, while the increase during Eid al-Fitr, which follows the Hijri calendar, is an example of calendar variation. This calendar variation poses a unique challenge since Eid al-Fitr shifts by approximately 21 days each year when calculated using the Gregorian calendar.

To effectively analyze and forecast passenger numbers, particularly in the context of calendar variations, the ARIMA method with non-metric exogenous variables, known as ARIMAX [18], is employed. ARIMAX is suitable for data that exhibit calendar variation effects, making it a pertinent choice for this study.

Previous studies have explored similar forecasting challenges. For example, [3] used the Autoregressive Integrated Moving Average (ARIMA) model to forecast train passenger numbers on Sumatra Island, with data spanning from January 2006 to December 2016. The study identified the ARIMA (1,1,1) model as the most accurate for this purpose. Another study by [4] predicted train passenger numbers while accounting for calendar variations using the SARIMAX model. This research analyzed data from 2014 to 2018 and found that the SARIMAX (1,1,1) model provided the best forecasts.

However, these studies did not include exogenous variables capable of capturing both calendar variations and the impact of the Covid-19 pandemic. This research seeks to fill that gap by incorporating exogenous variables that account for calendar variations and the Covid-19 pandemic within the ARIMAX framework. The aim is to enhance the accuracy of forecasts for the number of train passengers and to monitor the residuals of the regression model on Sumatra Island, thus contributing to more effective planning and service provision by PT KAI.

This study not only builds on the work of previous research but also addresses the unique challenges posed by calendar variations and extraordinary events like the Covid-19 pandemic, ensuring a comprehensive approach to passenger forecasting in the rail industry.

2. RESEARCH METHODS

The research methodology consists of detailed explanations regarding the research structure, including data sources, descriptions of variables, and methods of data analysis.

2.1 Data Sources

The data used in this research comes from Badan Pusat Statistik (BPS), which provides comprehensive information on the number of train passengers on Sumatra Island. As defined by Government Regulation Number 33 of 2021[7], it is classified as land transportation used for the public transportation of people and goods at predetermined costs. This regulation, enacted in 2021, plays a crucial role in governing and standardizing the operation of train services throughout Indonesia, including Sumatra. It establishes guidelines for safety, operational procedures, and fare structures to ensure consistent and efficient transportation. The period of this research spans form January 2013 to September 2023, covering data from three main railway regions on the island: North Sumatra, West Sumatra and South Sumatra.

2.2 Research Variable

Data on the number of train passengers on the island of Sumatra in this study is divided into two parts, namely training and testing. Training data is used to build the model, covering January 2013 to January 2023. Testing data is used to determine the forecast accuracy of the model built from training data covering February 2023 to September 2023.

Table 1. **Variable Description**

Variable	**Description**
$\boldsymbol{Y}$	Number of Train Passengers
$\mathbf{d_1}$	Dummy Covid-19
$\mathbf{d_2}$	Dummy Id
$\mathbf{d_3}$	Dummy Idt-1
$\mathbf{d_4}$	Dummy Idt+1
$\mathbf{d_5}$	Dummy Holiday
$\mathbf{d_6}$	Dummy Trend
$\mathbf{d_7}$	Dummy January
$\mathbf{d_8}$	Dummy Lockdown

Variable Definitions:

1. Number of Train Passengers on the island of Sumatra taken from January 2013 until September 2023.
2. A dummy COVID-19 is a dummy variable that indicates before and after the occurrence of COVID-19 (dummy COVID-19). This dummy variable categorizes data influenced and unaffected by COVID-19 overall. The assumption is that when there is COVID-19, the number of passengers decreases, hence the COVID dummy is 0, while dummies other than COVID are 1.
3. A dummy Id is a dummy variable that represents Eid al-Fitr (dummy Id). This dummy variable accounts for calendar variations due to the influence of Eid al-Fitr.
4. A dummy Idt-1 is a dummy variable that reflects one month before Eid al-Fitr (dummy Idt-1). This dummy variable captures train passengers traveling before Eid al-Fitr as people typically depart before the holiday.
5. A dummy Idt+1 is a dummy variable that denotes one month after Eid al-Fitr (dummy Idt+1). This dummy variable captures train passengers traveling after Eid al-Fitr as people return to their hometowns after the holiday.
6. A dummy Holiday is a dummy variable for every school holiday and year-end break (dummy Holiday). This dummy variable captures passengers traveling during school holiday periods.
7. A dummy Trend is a dummy variable for trends occurring within a specific period (dummy Trend). This dummy variable identifies trends within passenger data.
8. A dummy January is a dummy variable for January (dummy January). This dummy variable accounts for increased travel during the New Year period when many people travel to or from their hometowns.
9. A dummy Lockdown is a dummy variable indicating government-imposed lockdown periods (dummy lockdowns). This dummy variable captures data specifically related to COVID-19. During a lockdown, there is no activity (no passengers), hence the lockdown dummy is 0, while the rest are 1, but it only affects COVID.

2.3 Analysis Technique

In this research, the number of rail passengers on Sumatra Island is forecasted using the ARIMAX methodology, which combines the ARIMA method with time series regression. The time series regression component of the model is specifically designed to account for key external variables that influence passenger numbers. These variables include calendar variations, such as the surge in train passengers during Eid al-Fitr, seasonality during school holidays, and long-term trends that reflect the general increase in passengers over time. Additionally, the model incorporates the significant impact of the COVID-19 pandemic, which caused a drastic decline in the number of train passengers. By including these variables in the regression model, the effects of external factors on the number of passengers can be captured. Once these influences are modeled, the ARIMA method is then applied to forecast the data, focusing on the underlying time series structure without the direct effects of calendar variations, seasonality, trends, and COVID-19-related disruptions. This approach allows more accurate and nuanced predictions of future passenger numbers. The following are the steps used in this research:

1. Define dummy variables on data.

Table 2. Dummy Define

Variable	Point to-	Information
$\mathbf{d_1}$	1-87; 113-121	Binary dummy
$\mathbf{d_2}$	8; 19; 31; 43; 54; 66; 78; 89; 101; 113	Binary dummy
$\mathbf{d_3}$	7; 18; 30; 42; 53; 67; 77; 88; 100; 112	Binary dummy
$\mathbf{d_4}$	9; 20; 32; 44; 55; 67; 79; 90; 102; 114	Binary dummy
$\mathbf{d_5}$	7; 12; 19; 24; 31; 36; 43; 48; 55; 60; 67; 72; 79; 84; 91; 96; 103; 108; 115; 120	Binary dummy
$\mathbf{d_6}$	1-86; 121-87	Trend dummy
$\mathbf{d_7}$	1; 13; 25; 37; 49; 61; 73; 85; 97; 109; 121	Binary dummy
$\mathbf{d_8}$	88-92; 103-107	Binary dummy

2. Modeling the effects of calendar, seasonal, trend and COVID variations with the following time series regression equation [18]:

$$Y_t = \beta_0 + \beta_1 \mathrm{d}_{1,\mathrm{t}} + \beta_2 \mathrm{d}_{2,\mathrm{t}} + \beta_3\, \mathrm{d}_{3,t} + \beta_4 \mathrm{d}_{4,t} + \beta_5 \mathrm{d}_{5,t} + \beta_6 \mathrm{d}_{6,\mathrm{t}} + \beta_7 \mathrm{d}_{7,t} + \beta_8 \mathrm{d}_{8,\mathrm{t}} + \varepsilon_t$$

3. Analysis of stationary test variance and mean of residual of time series regression models with Box-Cox and Augmented Dickey-Fuller Test (ADF). If the data is not stationary, then transformation and differencing are performed.
4. If the data is stationary, then determine the order of the ARIMA model based on ACF and PACF.
5. Parameter estimation and diagnostic checks, i.e.,
    a. White Noise Test
    Residual is white noise if there is a sequence of independent, identically distributed random variables with the mean is zero and the variance is constant (σ^2) [8]. The hypotheses used in this test are:

$$H_0: \rho_1 = \rho_2 = \cdots = \rho_k = 0$$
$$H_1: \text{There is at least one value } \rho_k \neq 0$$

The test statistics used in this test are expressed by the following equation:

$$Q^* = n(n+2)\sum_{k=1}^{K}(n-k)^{-1}\hat{\rho}_k^2 \tag{1}$$

with $k = 1,2,3 \dots K$
$\hat{\rho}_k^2$: sample correlation at k-th lag
K : the number of lags being tested
n : The number of observations
α : Significant grade value
The rejection area used is the null hypothesis rejected if the statistical value $(Q^*) >$ critical value (χ^2_{K-1}) or $p_{value} < \alpha$ with [11].

    b. Normality Test
    The residual normal assumption test used is the Kolmogorov-Smirnov method normality test [15]. The assumption of normality is crucial because many statistical methods, including those used in time series analysis, assume that residuals (errors) follow a normal distribution. This assumption ensures the validity of confidence intervals and hypothesis tests, leading to more reliable and accurate inferences. By verifying that the residuals are normally distributed, we confirm that the model is appropriately specified and that the statistical tests applied are valid. The hypotheses used in this test are:

H_0: normally distributed residuals
H_1: residuals are not normally distributed

The test statistics used in this test are expressed in the following equation.

$$D = Sup|S(x) - F_0(x)| \tag{2}$$

with:

$F_0(x)$: Cumulative chance function of normal distribution
$S(x)$: Sample distribution functions
Sup : Maximum value of $|S(x) - F_0(x)|$

The rejection area used is the null hypothesis rejected if the statistical value of the test is greater than the critical value or $p_{value} < \alpha$ with i.e. free degrees used. The critical value of using the Kolmogorov-Smirnov table with free degrees is, where is the number of observations [15].

6. Model ARIMAX

The Autoregressive Integrated Moving Average with Extragenous Variables (ARIMAX) model is an extension of the ARIMA model, designed to handle nonstationary time series data while incorporating exogenous variables to enhance forecasting accuracy [13]. These exogenous variables, often in the form of dummy variables, represent external factors that may influence the time series. For example, dummy variables can be used to capture the effects of seasonal events (like holidays or promotions), policy changes, economic shifts, or the binary events (such as the occurrence of natural disaster). By including these variables, the ARIMAX model can account for sudden shifts or specific periods that impact the dependent variable, leading to a more accurate and context-sensitive forecast [14]. The general equation for the ARIMAX model is as follows:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \cdots + \beta_k X_{k,t} + \frac{1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q}{1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p} e_t \tag{3}$$

with:

Y_t : observation value at t-th time
$X_{i,t}$: dummy variable value (1 or 0)
β_i : i-th variable dummy parameter
ϕ_p : the value of the coefficient on the p-th order
θ_q : the value of the coefficient on the q-th order
e_t : error value at t-th time

7. Forecast from the model obtained in step 6, then calculated the training and testing MAPE to see the accuracy of the best model. Forecasting is very useful in various fields in planning and controlling a system [8][11]. This paper measured by mean absolute percent error (MAPE), with equation [16]:

$$MAPE = \frac{\sum_{t=1}^{n} \left|\frac{a_t - b_t}{a_t}\right|}{n} \times 100\% \tag{4}$$

with:

a_t : actual data
b_t : forecast data at time t
n : number of data

3. RESULTS AND DISCUSSION

This chapter will be explained related to time series modeling using the calendar variation method on train passenger data and monitoring the prediction of the number of passengers on the island of Sumatra. The data used is data from January 2013 to September 2023 obtained from [17]. The graph of the data is as follows:

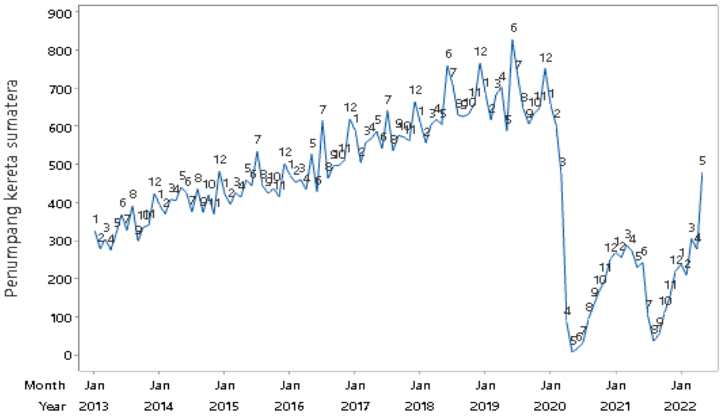


Figure 1. **Time series plot training data of the train passenger in Sumatra**

Based on the data graph, we can see that from 2013 to the end of 2019, a surge in passengers occurred every Eid, which is referred to as the calendar variation effect. In addition, passenger spikes also occur every mid-year and year-end, which is referred to as the seasonal effect. When COVID-19 was declared a pandemic in Indonesia, there was a very significant decline from 2020 to 2021 and then an increase again after COVID-19 was declared over in 2022 until now. In this study, data from January 2013 to January 2023 will be used as training data, and data from February 2023 to September 2023 will be used as testing data.

3.1 Dummy Regression Modeling

Dummy regression modeling is one type of regression modeling with the help of additional variables in the form of artificial variables by giving a value of 0 or 1. Dummy variables are used as described in subsection 2.2.

Here is the regression model obtained:

$$Y_t = 154.9 + 139.9\, d_{1,t} + 48.5\, d_{2,t} + 3.6\, d_{3,t} + 7.3\, d_{4,t} + 63.0\, d_{5,t} + 4.696\, d_{6,t} + 26.4\, d_{7,t} - 152.4\, d_{8,t} + \epsilon_t$$

with the summary model as follows:

Table 3. **Model Summary**

S	R-sq	R-sq(adj)
43.648	94.64%	94.26%

Data source: Minitab software Output, results of times series regression model of the train passenger data in Sumatra

The value of the coefficient of determination from the regression equation is 94.64%, which means that the model has explained the Y variable very well, so the researchers used the regression model equation above to predict how train passengers in Sumatra in the following years. It can be seen that the model has been able to capture events before and after COVID-19 but still has autocorrelation errors. Autocorrelation errors can cause residuals to not be normally distributed, which is very important thing. Therefore, researchers will conduct ARIMA analysis on residual data from regression so that better results are obtained.

3.2 ARIMAX modeling

3.2.1 Model Identification

The plot of training residual data is as follows:

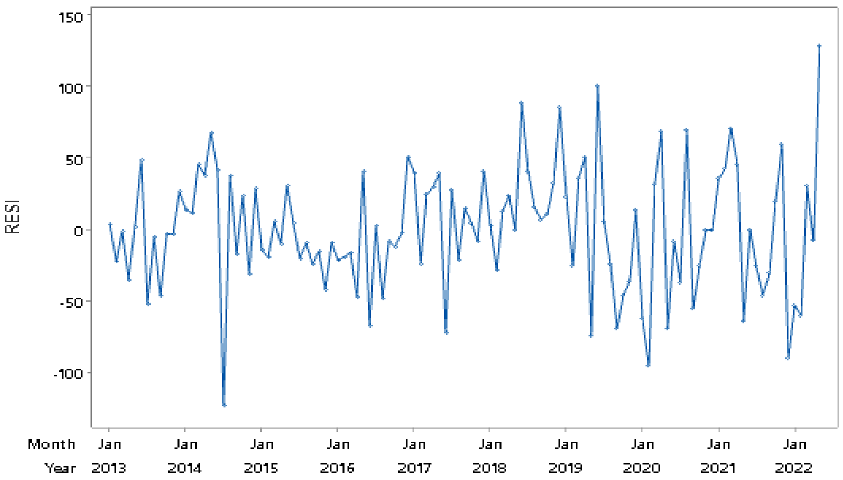
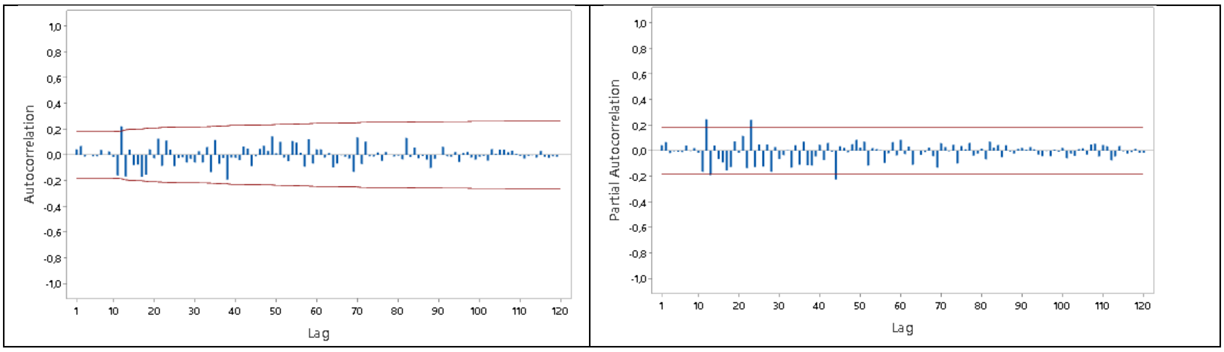


Figure 2. Time series plot residual data of the train passenger in Sumatra

The **Figure 2** shows that the error is stationary with respect to variance due to its constant variance. Likewise, the error is stationary with respect to the mean because of the statistic test. The augmented Dickey-Fuller is -9.9976 and the $p - value$ is 0.000, which rejects the null hypothesis because -9.9976 is less than the critical value of -2.8859, so it can be concluded that the data has been stationary and does not need to be differentiated.

Next, plot ACF and PACF on training data to determine which ARIMA models may be formed as follows:



(a) **(b)**

Figure 3. (a) ACF (b) PACF plot training data of the train passenger in Sumatra

Based on the ACF and PACF plots above, the presumed models may be as follows:

Table 4. ARIMA Models

Models	Parameters	White Noise	Normality
$ARIMA(0\ 0\ 1)(2\ 0\ 1)^{12}$	Not Significant	White noise	Not Normal
$\boldsymbol{ARIMA(1\ 0\ 2)(2\ 0\ 2)^{12}}$	**Significant**	**White noise**	**Normal**
$ARIMA(0\ 0\ 1)(2\ 0\ 2)^{12}$	Significant	Not white noise	Normal
$ARIMA(1\ 0\ 0)(2\ 0\ 2)^{12}$	Significant	Not white noise	Normal
$ARIMA(0\ 0\ 0)(2\ 0\ 2)^{12}$	Not Significant	Not white noise	Not Normal

Data source: Minitab software Output, results of presumed ARIMA models of the train passenger data in Sumatra

Based on **Table 4**, the model that meets these three criteria such us significant parameters, white noise and a normal distribution residual is $\text{ARIMA}(1\ 0\ 2)(2\ 0\ 2)^{12}$.

3.2.2 Parameters Estimator

Result of parameters estimation with model $\text{ARIMA}(1\ 0\ 2)(2\ 0\ 2)^{12}$ that are:

Table 5. Parameters Estimation

Type	Coef.	SE Coef.	T-Value	P-Value
AR 1	0.999820	0.000318	3139.18	0.000
SAR 12	0.9151	0.0830	11.03	0.000
SAR 24	-0.9284	0.0757	-12.27	0.000
MA 1	0.639980	0.000965	766.45	0.000
MA 2	0.2841	0.0529	5.37	0.000
SMA 12	0.535	0.150	3.57	0.001
SMA 24	-0.761	0.148	-5.15	0.000

Data source: Minitab software Output, results of ARIMA $(1\ 0\ 2)(2\ 0\ 2)^{12}$ *model of the train passenger data in Sumatra*

Based on Table 5, all results from parameter estimation get $p - value < 0.05$ values that can be interpreted that all parameters can be used for forecasting.

Table 6. Residual of Sum Square

DF	SS	MS
114	168600	1478.95

Data source: Minitab software Output, results of ARIMA $(1\ 0\ 2)(2\ 0\ 2)^{12}$ *model of the train passenger data in Sumatra*

Based on Table 6, the model $\text{ARIMA}(1\ 0\ 2)(2\ 0\ 2)^{12}$ has the smallest mean square of other models so it can be said that the model $\text{ARIMA}(1\ 0\ 2)(2\ 0\ 2)^{12}$ is the best model.

3.2.3 Diagnostic Checking

The result of Ljung-Box Chi Square statistic to check the white noise model $\text{ARIMA}(1\ 0\ 2)(2\ 0\ 2)^{12}$ is in Table 7.

Table 7. Ljung-Box Chi Square Statistic

Lag	12	24	36	48
Chi-Square	10.81	22.53	30.29	44.89
DF	5	17	29	41
$\boldsymbol{P-Value}$	0.055	0.165	0.399	0.312

Data source: Minitab software Output, results of ARIMA $(1\ 0\ 2)(2\ 0\ 2)^{12}$ *model of the train passenger data in Sumatra*

Based on Table 7 we can be seen that all $p - value\ > 0.05$ **(Failed to Reject)** $\boldsymbol{H_0}$ which means that the model has $\text{ARIMA}(1\ 0\ 2)(2\ 0\ 2)^{12}$ white noise. Next is to see if the residuals from the model have been normally distributed.

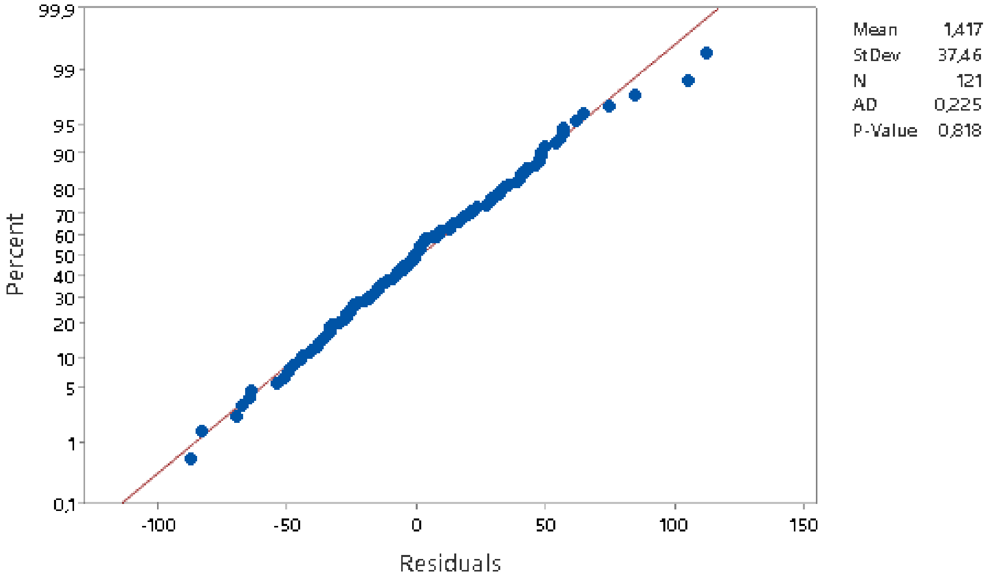


Figure 4. Residual Normality Plot Training Data of the Train Passenger in Sumatra

It can be seen from **Figure 4** that the p-value of the residual is more than 0.05 (failed to reject) H_0 so the residual model has been normally distributed and the normality assumption of the ARIMA model is confirmed. In this case the best model is $\text{ARIMA}(1\ 0\ 2)(2\ 0\ 2)^{12}$ in modeling residual regression.

3.2.4 ARIMAX Modeling

The following is the model used to $\text{ARIMA}(1\ 0\ 2)(2\ 0\ 2)^{12}$ forecast the future period:

$$\phi_1(B)\Phi_2(B^{12})z_t = \theta_2(B)\Theta_2(B^{12})a_t$$

The value of $\phi_1 = 0.9998.\ \theta_1 = 0.7399.\ \theta_2 = 0.2841.\ \Phi_1 = 0.9151.\ \Phi_2 = -0.9284.\ \Theta_1 = 0.535$ and $\Theta_2 = -0.761$ .So that the final model used for forecasting becomes with the following equation $\text{RegARIMA}(1\ 0\ 2)(2\ 0\ 2)^{12}$:

$$\begin{aligned} Yt = {} & 154.9 + 139.9\, d_{1,t} + 48.5\, d_{2,t} + 3.6\, d_{3,t} + 7.3\, d_{4,t} + 63.0\, d_{5,t} + 4.696\, d_{6,t} + 26.4\, d_{7,t} \\ & - 152.4\, d_{8,t} + 0.9998 z_{t-1} + 0.9151 z_{t-12} - 0.9284 z_{t-24} - 0.9149 z_{t-13} \\ & - 0.535\, a_{t-12} + 0.761 a_{t-24} - 0.7399 a_{t-1} - 0.2841 a_{t-2} + a_t \end{aligned}$$

Next is to forecast the number of passengers in Sumatra for the next 8 months and then compare it with testing data from February 2023 to September 2023 which is as follows:

Table 8. Forecasting 8 Month in 2023

Month	Number of Train Passengers	
	Original Data	Forecasting Result
February	515	597.174
March	571	577.102
April	594	603.696
May	648	621.988
June	658	590.960
July	576	612.084
August	581	556.678
September	451	614.010

Data source: Excel software Output, forecasting result of model $RegARIMA(1\ 0\ 2)(2\ 0\ 2)^{12}$

Table 8 presents the forecasting results for the number of rail passengers in Sumatra for the next eight months, from February 2023 to September 2023. The forecasting results are compared with the actual out-of-sample data obtained from February 2023 to September 2023. **Table 8** summarizes the comparison between the original data and the forecasting results for each month. The original data represents the actual number of rail passengers during the specified months, while the forecasting results show the estimated number of passengers obtained from our forecasting model. From **Table 8**, it can be seen that the forecasting results generally show a fairly good agreement with the original data for most months. In particular, the forecasting model predicts higher ridership for February, April, May, July, and September while predicting lower ridership for March, June, and August compared to the original data. Differences between the forecasting results and the original data can be due to such things as seasonal variations, long-term trends, or other external factors. Despite such variations, overall, the forecasting model shows its effectiveness in providing a reasonable forecast for the number of rail passengers in Sumatra for the specified period.

The graphs of the original data. training data and forecast data are as follows:

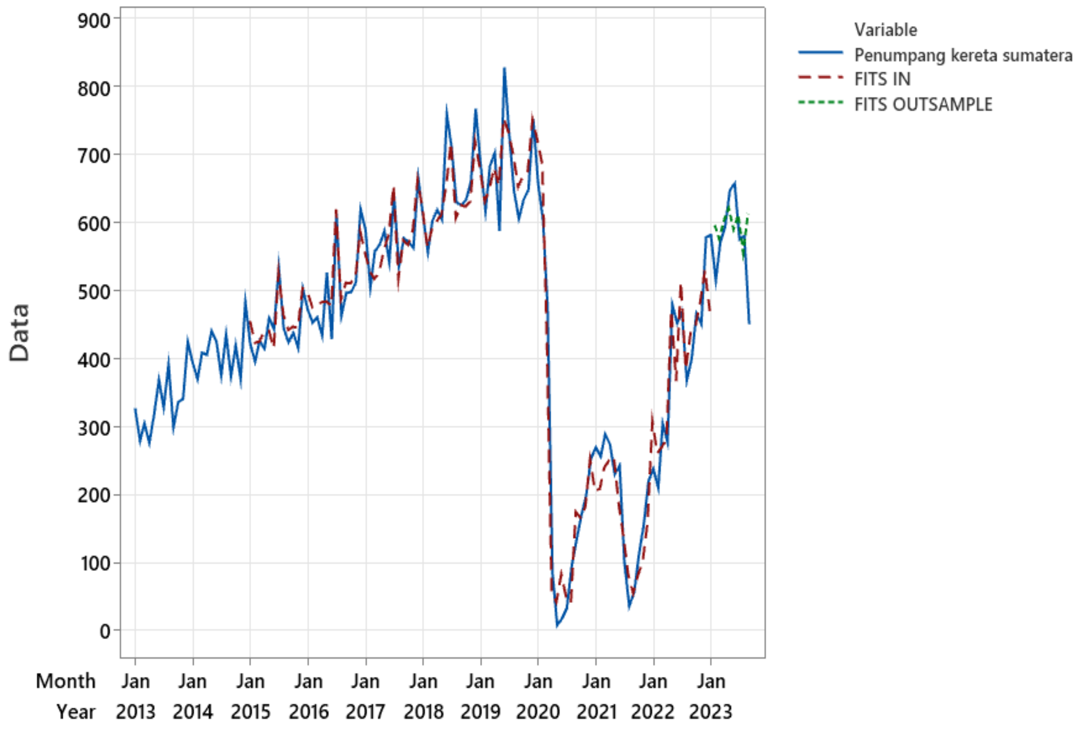


Figure 5. Time Series Data and Forecasting Plot of the Train Passenger in Sumatra

Figure 5 displays the actual data and forecasting results for both the training and testing data of the number of train passengers on the island of Sumatra. The red dotted line is the result of forecasting the number of train passengers in Sumatra on the training data, while the green dotted line is the result of forecasting the number of train passengers on the island of Sumatra on the testing data. In general, the results of forecasting the number of train passengers on the island of Sumatra displayed in Figure 5 show that the pattern of forecasting values tends to follow the pattern of previous data. This indicates the effect of calendar variation on the forecasting results of the number of railroad passengers. It can be seen that in January 2020, the number of passengers experienced a drastic decline; this was due to the Coronavirus disaster that hit various countries, including the island of Sumatra in Indonesia. Then, sometime later, the number of passengers increased again after the coronavirus pandemic began to get better.

Next is to look at the average error or MAPE (Absolute Mean Percentage Error) training and testing as follows:

Table 9. MAPE

Training	Testing
18.12%	9.93 %

Data source: Excel software Output, forecasting result model

A lower MAPE value indicates better accuracy, it can be seen that the training MAPE is 18.12% suggests there is some room for improvement in the models fit to the training data. This happens because the model has not been able to capture events when COVID-19 occurs. But this is good enough because the MAPE test has a smaller average error of 9.93%, which is less than 10% that indicates better accuracy. From the model $RegARIMA(1\ 0\ 2)(2\ 0\ 2)^{12}$, we can see that the seasonal components $(2\ 0\ 2)^{12}$helps capture yearly recurring events like Eid al-Fitr by accounting for patterns that repeat every 12 months. The regression component also allows the model to include the impact of external events like COVID-19. So, from this model, we can represent that on Eid al-Fitr, there is an increase in the number of passengers of up to 49 people every Eid al-Fitr, and during the COVID-19 condition, the data has decreased to 140 people.

4. CONCLUSIONS

To predict the number of Sumatra Island train passengers, ARIMA modeling with time series regression (calendar variations) can capture the uniqueness of existing data. The model obtained is in the good category with MAPE training of 18.12% and testing of 9.93%. A model with a good category means that the model is still acceptable and has a low error. This shows that the prediction of the number of train passengers is not much different from the actual number of passengers. In the case of forecasting the model of the number of Sumatra train passengers from January 2013 to September 2023, the best model is $\text{RegARIMA}(1\ 0\ 2)(2\ 0\ 2)^{12}$ because it can overcome certain patterns such as the increased passengers during Eid Al-Fitr and the sharp decreased in the number of train passengers due to Covid-19. In this model, certain patterns can be overcome by adding exogenous variables in the form of dummies by looking at unique patterns in the time span of the data. For further research, the Support Vector Regression (SVR) algorithm can be used. The SVR algorithm has significant advantages over ARIMA, which is commonly used in time series models. ARIMA often cannot handle outlier data and model nonlinear time series.

ACKNOWLEDGMENT

This research was supported by the Directorate General of Science and Technology Resources and Higher Education (Kemenristekdikti), through the scholarship "Master Education Program Toward Doctor for Excellent Scholars (PMDSU)" which funded this research.

REFERENCES

[1] H. A. Karim. S. H Lis Lesmini. D. A Sunarta... & M. Bus. *Manajemen transportasi*. Surabaya: Cendikia Mulia Mandiri. 2023.

[2] R. Ravico. & B Susetyo. “Sejarah Pembangunan Jalur Kereta Api Sebagai Alat Transportasi Di Sumatera Selatan Tahun 1914-1933.” *Agastya J. Sej. Dan Pembelajarannya*. vol. 11. no. 1. pp. 68–82. 2021.

[3] I. S. Nurjanah. D. Ruhiat. and D. Andiani. “Implementasi Model Autoregressive Integrated Moving Average (Arima) Untuk Peramalan Jumlah Penumpang Kereta Api Di Pulau Sumatera.” *TEOREMA Teor. dan Ris. Mat.*. vol. 3. no. 2. p. 145. 2018. doi: 10.25157/teorema.v3i2.1421.

[4] N. N. D. Hayati and S. Martha. “Prediksi Data Jumlah Penumpang Kereta Dengan Efek Variasi Kalender Pada Model Sarimax.” *Bimaster Bul. Ilm. Mat. Stat. …*. vol. 10. no. 4. pp. 379–388. 2021. [Online]. Available: https://jurnal.untan.ac.id/index.php/jbmstr/article/view/49536%0Ahttps://jurnal.untan.ac.id/index.php/jbmstr/article/download/49536/75676590652

[5] A. R. Nisa. T. Tarno. and A. Rusgiyono. “Peramalan Harga Cabai Merah Menggunakan Model Variasi Kalender Regarima Dengan Moving Holiday Effect (Studi Kasus: Harga Cabai Merah Periode Januari 2012 Sampai Dengan Desember 2019 Di Provinsi Jawa Barat).” *J. Gaussian*. vol. 9. no. 2. pp. 170–181. 2020. doi: 10.14710/j.gauss.v9i2.27819.

[6] S. N. Intan. E. Zukhronah. and S. Wibowo. “Peramalan Banyaknya Pengunjung Pantai Glagah Menggunakan Metode Autoregressive Integrated Moving Average Exogenous (ARIMAX) dengan Efek Variasi Kalender.” *Indones. J. Appl. Stat.*. vol. 1. no. 2. p. 70. 2019. doi: 10.13057/ijas.v1i2.26298.

[7] Indonesia. *Government Regulation Number 33 of 2021 on The Organization of The Railway Sector*. Law Number. indonesia. 2021.

[8] W. W. S. Wei. *Time Series Analysis Univariate and Multivariate Methods*. 2nd ed. New York: Pearson Education. 2006.

[9] R. J. Makridakis. S.. Wheelwright. S. C.. dan Hyndman. *Forecasting Methods and Applications*. New Jersey: John Wiley & Sons. 2008.

[10] J. E. Hanke and D. Wichern. *Business Forecasting*. 9th ed. New Jersey: Pearson Education. 2014.

[11] J. D. Cryer and K.-S. Chan. “Time Series Analysis - Front Pages.” *Time Time Ser. Anal. with Appl. R*. 2008.

[12] G. T. Wilson. “Time Series Analysis: Forecasting and Control. 5th Edition. by George E. P. Box. Gwilym M. Jenkins. Gregory C. Reinsel and Greta M. Ljung. 2015. Published by John Wiley and Sons Inc.. Hoboken. New Jersey. pp. 712. ISBN: 978-1-118-67502-1.” *J. Time Ser. Anal.*. vol. 37. no. 5. pp. 709–711. 2016. doi: 10.1111/jtsa.12194.

[13] D. Rosadi. *Analisis Ekonometrika dan Runtun Waktu Terapan dengan R*. Yogyakarta: Andi. 2011.

[14] M. Cools. E. Moons. and G. Wets. “Investigating the variability in daily traffic counts through use of ARIMAX and SARIMAX models.” *Transp. Res. Rec.*. no. 2136. pp. 57–66. 2009. doi: 10.3141/2136-07.

[15] W. W. Daniel. *Statistika Nonparametrik Terapan*. Jakarta: PT. Gramedia. 1989.

[16] A. Al-Khowarizmi. O. S. Sitompul. S. Suherman. and E. B. Nababan. “Measuring the Accuracy of Simple Evolving Connectionist System with Varying Distance Formulas.” *J. Phys. Conf. Ser.*. vol. 930. no. 1. 2017. doi: 10.1088/1742-6596/930/1/012004.

[17] BPS. “Jumlah Penumpang Kereta Api (ribu orang). 2006-2023.” *September 2023*. 2023. Jumlah Penumpang Kereta Api - Tabel Statistik - Badan Pusat Statistik Indonesia (bps.go.id) (accessed :Nov. 15. 2023).

[18] M S Akbar et al. GSTAR-SUR Modeling With Calendar Variations And Intervention To Forecast Outflow Of Currencies In Java Indonesia. *Journal of Physics.: Conference. Series. 974 012060*. 2018. doi :10.1088/1742-6596/974/1/012060.