FORECASTING NUMBER OF TRAIN PASSENGERS USING TIME SERIES REGRESSION INTEGRATED CALENDAR VARIATION AND COVID 19 INTERVENTION

Mega Silfiani[1)]
Farida Nur Hayati[2)]
[1,2)] Department of Statistics, Institut Teknologi Kalimantan, Balikpapan City, Indonesia
e-mail: megasilfiani@lecturer.itk.ac.id

ABSTRACT

The purpose of this study is to obtain a forecasting model for the number of train passengers using time series regression integrated with variations in the Islamic calendar and the effects of COVID 19. This study uses the number of train passengers in Jabodetabek, Java (Non-Jabodetabek), and Sumatra from January 2006 to December 2022 as the data source. Time series regression with variations of the Islamic calendar and the effects of COVID 19 for Jabodetabek, Java (non-Jabodetabek), and Sumatra has an RMSE value for each category of 7657,821; 2453.827 and 275.901. In general, the number of train passengers for all categories (Jabodetabek, Java, Sumatra) has a seasonality. In Jabodetabek and Sumatra, Eid al-Fitr has a big impact on the number of train passengers. Meanwhile, one month before Eid al-Fitr has a big impact on the number of train passengers in Java (Non Jabodetabek). In addition, the impact of COVID 19 significantly affected the number of train passengers for all categories.

Keywords: calendar variation, COVID 19 intervention, time series regression.

INTRODUCTION

Transportation is a crucial aspect of our everyday existence. Transportation has the ability to influence communities, enhance connections, and ultimately enhance the general quality of life. Through the ongoing enhancement of transportation infrastructure and the implementation of cutting-edge technologies, society can effectively address obstacles, foster sustainable mobility, and establish a more interconnected and affluent global community. Transportation refers to the process of relocating, conveying, or altering the path of an entity from one place to another. The primary element of transportation involves the actual conveyance or displacement of goods (commodities) and passengers to different destinations (Fatimah, 2019). Trains are widely recognized as a major factor in promoting efficient and sustainable transportation among many modes of travel. Trains are a widespread mode of transportation in Indonesia and have a vital function in enabling community mobility.

PT Kereta Api Indonesia (KAI) is an Indonesian train company that offers both passenger and freight train services throughout Indonesia. The corporation states that the COVID-19 pandemic has had a substantial impact on the decrease in train passengers throughout the past three years. In addition to the COVID-19 pandemic, the number of train passengers exhibits fluctuations throughout different months, particularly leading up to holidays and the celebration of Islamic Event such as Eid al-Fitr. Providing data on the fluctuations in the numbers of train passenger impacted by these factors will greatly assist KAI in forecasting future increases in passenger volume. A particular technique involves utilizing time series analysis. The time series analysis technique is an appropriate approach for forecasting the future number of passengers. Andalita and Irhamah (2015) conducted multiple studies on forecasting the number of train passengers in Indonesia. They specifically focused on forecasting the number of passengers for economy class trains in Kertajaya using ARIMA and ANFIS. The study demonstrates that ANFIS yields more precise predictions of passenger numbers for the subsequent 14 time periods in comparison to ARIMA (Andalita and Irhamah, 2015). Utomo and Fanani (2020) applied the Seasonal Autoregressive Integrated Moving Average (SARIMA) methodology to forecast the quantity of train passengers in Indonesia. According to the findings, the SARIMA (1,1,2)(0,1,1)12 model, with a mean squared error (MSE) of 0.0468 and a mean absolute percentage error (MAPE) of 6.26%, is identified as the most accurate forecasting model (Utomo and Fanani, 2020). In addition, Nurjanah, Ruhiat, and Andiani (2018) utilized the ARIMA approach to forecast the quantity of train passengers in Sumatra, producing a Mean Absolute Percentage Error (MAPE) of 5.11%. The three studies that focused on the number of passengers forecasts were done prior to the COVID-19 pandemic. Consequently, studies that investigating COVID 19 influenced the number of passenger trains in Indonesia remain fairly novel.

An effective way for integrating the impacts of COVID-19 into time series analysis is the intervention model. Rianda (2021) conducted multiple investigations utilizing the COVID-19 intervention model to forecast the impact of COVID-19 interventions on the volume of airplane passengers at Soekarno Hatta Airport. The results of these studies yielded a Mean Absolute Percentage Error (MAPE) of 14.07%. Furthermore, Silfiani *et al*. (2022) utilized the COVID-19 intervention model to make predictions about the Gross Regional Domestic Product (GRDP) in Bali Province. The final model they used was SARIMA$(0,1,0)(1,0,0)^4$, with an intervention order of r=1. In addition to COVID-19, another significant aspect to take into account when forecasting the number of train passengers is the holiday season. Due to the Eid al-Fitr the celebration advancing around 10 days annually, the conventional seasonal model is inadequate for capturing the unique attributes of this time series. The calendar variation model is a viable technique to address this issue. Susanti *et al*. (2018), Suhartono, Lee, and Prasetyo (2015), and Anggraeni, Vinarti, and Kurniawati (2015) conducted forecasting using a calendar variation model.

Given the annual variation in Eid al-Fitr celebrations and the major effect of COVID-19 on train numbers of passengers, it is essential to conduct relevant studies to forecast future train passenger. This study presents a time series regression model that combines fluctuations in the Islamic calendar with interventions related to COVID-19. This model will be utilized to analyze case studies on the volume of rail commuters in the Jabodetabek, Java (excluding Jabodetabek), and Sumatra regions. Wulansari *et al*. (2014) conducted multiple studies on forecasting approaches that utilize time series regression. They specifically applied time series regression analysis to Indonesian bank currency netflow data, taking into account calendar variation effects. The purpose of their study was to control for banking liquidity in Indonesia. The findings of this study indicate that the ARIMA model outperforms the ARIMA method and data extrapolation in terms of netflow analysis (Wulansari, *et al*., 2014). Zukhronah *et al*. (2021) conducted a study on a model that incorporates calendar variations in time series regression to forecast the number of visitors to Grojogan Sewu. The SARIMA model fails to accurately describe data with calendar fluctuations, as evidenced by the high RMSE value reported by Zukhronah *et al*. (2021). This project aims to develop a highly precise model that surpasses existing models. It seeks to provide a comprehensive analysis of number of train passengers, enabling accurate forecasts of both increases and decreases in train passengers in Jabotabek, Java (Non-Jabotabek), and Sumatra.

METHOD

In the research methodology section, data and research categories, the time series regression integated Islamic calendar variations and the COVID 19 intervention, and the modeling approach will be described.

Data and Research Variable

The research utilizes secondary data obtained from the Statistics Indonesia (BPS). The research variables in this article consist of data related to the monthly number of train passengers in Jabodetabek, Java (excluding Jabodetabek), and Sumatra, spanning from January 2006 to December 2022. The training data for this study ranges from January 2006 to December 2021, while the testing data covers the period from January to December 2022. The number of train passengers in Jabodetabek, Java (Non-Jabodetabek), and Sumatra, measured in thousand people. The independent factors consist of dummy variables representing seasonal months, Islamic calebrations, and the presence of COVID-19.

Time Series Regression Integrated Islamic Calendar Variation and COVID 19 Intervention

There is one response variable and one or more time-conditioned predictor variables in time series regression (Wijiyanto, Kusrini and Irhamah, 2012). Under the assumption of a linear relationship, time series regression predicts the intended time series (Silfiani, *et al.*, 2021). Response variables are often referred regress, dependant, or explained variables. Predictive variables are addressed by regressors, independent variables, and explanatory factors (Hyndman and Athanasopoulos, 2018). Regression models have the benefit of having the capacity to demonstrate an essential relationship between the response desired variable and the predictor variables. Several important indicators, such as seasonality dummy variables, frequently apply in regression for time series data (Hyndman and Athanasopoulos, 2018).

$$Y_t = \beta_1 d_{1,t} + \cdots + \beta_{S-1} d_{S-1,t} + N_t$$

$$d_{i,t} = \begin{cases} 1, & t = i \\ 0, & t \neq i \end{cases} \tag{1}$$

where $d_{i,t}$ is i-th dummy variable at t. Number of dummy variables for seasonality are $S-1$ where *S* is a total of seasonality. That is because the last category of seasonality, e.g., December if the series of data is monthly, is reached by the intercept and is specified when the dummy variables are all set to zero (Hyndman and Athanasopoulos, 2018).

Implementing the same concept as constructing seasonal indicators with a dummy variable in time series regression, it is also feasible to create variations of the Islamic calendar. Incorporating variations of the Islamic calendar enables the general formula of the time series regression function:

$$Y_t = \beta_0 + \beta_1 d_{1,t} + \cdots + \beta_{S-1} d_{S-1,t} + \beta_{IF-1} d_{IF-1,t} + \beta_{IF} d_{IF,t} + \beta_{IF+1} d_{IF+1,t} + \beta_{IA} d_{IA,t} + N_t \tag{2}$$

When the time series regression model with calendar variations does not satisfy the residual independence assumption or there is an autocorrelation within lags, the seasonal autoregressive integrated moving average (SARIMA) must be adapted. SARIMA is incorporated to the time series regression with calendar variation model, giving in equation (3).

$$Y_t = \beta_0 + \beta_1 d_{1,t} + \cdots + \beta_{S-1} d_{S-1,t} + \beta_{IF-1} d_{IF-1,t} + \beta_{IF} d_{IF,t} + \beta_{IF+1} d_{IF+1,t} + \beta_{IA} d_{IA,t} + \frac{\theta_q(B)\Theta_Q(B^S)a_t}{\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D} \quad (3)$$

When the assumption of normal distributions residuals is failed in SARIMA, outlier detection is carried out. There are two types of outlier that applied in this study, i.e., additive and level shif outlier. The outlier effect is introduced into the model as (4).

$$Y_t = \beta_0 + \beta_1 d_{1,t} + \cdots + \beta_{S-1} d_{S-1,t} + \beta_{IF-1} d_{IF-1,t} + \beta_{IF} d_{IF,t} + \beta_{IF+1} d_{IF+1,t} + \beta_{IA} d_{IA,t} + \frac{\theta_q(B)\Theta_Q(B^S)a_t}{\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D} + \sum_{j=1}^{k} \varpi_j v_j(B) I_j^{(T_j)} \quad (4)$$

where $I_j^{(T_j)}$ is an outlier at period T_j, and $v_j(B)=1$ is for additive outlier, $v_j(B) = \frac{1}{(1-B)}$ is for level shift outlier.

The intervention model is a time-series approach used to assess the impact of external influences (Ismail *et al*., 2009). Suppose, Y_t is a time series process that follows the SARIMA (p,d,q)(P,D,Q)S model with calendar variations; consequently, when there is an effect arising from an intervention at period t denoted X_t, the intervention model can be expressed as equation (6).

$$Y_t = f(X_t) + V_t \quad (5)$$

$$V_t = \beta_0 + \beta_1 d_{1,t} + \cdots + \beta_{S-1} d_{S-1,t} + \beta_{IF-1} d_{IF-1,t} + \beta_{IF} d_{IF,t} + \beta_{IF+1} d_{IF+1,t} + \beta_{IA} d_{IA,t} + \frac{\theta_q(B)\Theta_Q(B^S)a_t}{\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D} + \sum_{j=1}^{k} \varpi_j v_j(B) I_j^{(T_j)} \quad (6)$$

where V_t is noise from SARIMA $(p,d,q)(P,D,Q)^S$ with calendar variation.

There are two types of intervention variable i.e., step and pulse functions. The intervention in this research is handled by step functions. The step function is an intervention variable for long-term activities (Ismail *et al*., 2009). If T is the initial event of the intervention, then the step function intervention variable can be expressed as follows:

$$X_t = \begin{cases} 0, & t < T \\ 1, & t \geq T \end{cases} \tag{7}$$

Thus, the mathematical formula for the intervention model with the step function variable can be represented as (8).

$$Y_t = \frac{\omega_s(B)}{\delta_r(B)} B^b X_t + V_t \tag{8}$$

where b is delay time for the effect of the initial intervention occurs, $\omega_s(B) = \omega_0 - \omega_1 B - \omega_2 B^2 - \cdots - \omega_s B^s$ and $\delta_r(B) = 1 - \delta_1 B - \delta_2 B^2 - \cdots - \delta_r B^r$.

Research Procedures

The procedure for developing an integrated time series regression model with variations in the Islamic calendar and the effects of COVID is as follows:

1. Dividing training data from testing data, then splitting training data into two parts, namely data prior to and following the COVID 19 case in Indonesia (March 2020)
2. Constructing the dummy variable for the period of calendar variation
3. Applying time series regression of calendar changes to eliminate the impacts of calendar variations by eq. (2)
4. Conducting a white noise test on N_t either using the Ljung-Box test or by examining its ACF and PACF graphs. If it fits the white noise assumption, then proceed to step 7. If not, the Box-Jenkins ARIMA technique is utilized to model Nt.
5. The ARIMA model in step 4 is used to employ in the number of train passengers, while simultaneously modeling the dummy variable on calendar fluctuations as input to obtain a model with a stochastic trend.
6. Estimating the parameters using conditional least square estimation and then determining their significance using the t test.
7. Applying the Ljung-Box test to verify the residual of white noise assumption
8. Forecasting training data following the COVID 19 outbreak in Indonesia
9. Computing and presenting the forecast error in Step 8
10. Calculating the order b, s, and r based on the graph in Step 9 to generate an integrated time series regression model of calendar variation and COVID-19 intervention:
11. Predicting data testing and calculating Root Mean Square Error (RMSE) as follows (9) (Wei, 1990).

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} \left(Y_t - \hat{Y}_t\right)^2}{n}} \tag{9}$$

Performing forecast monthly number of train passengers for Jabodetabek, Java (Non Jabodetabek) and Sumatra in 2023.

RESULTS AND DISCUSSION

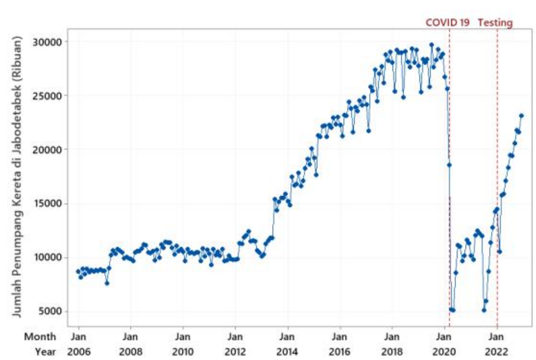
Table 1 presents the characteristics of the data on the number of train passengers in Indonesia. Jabodetabek recorded the highest average number of train passengers from January 2006 to December 2022, with a total of 15,937 thousand passengers. Most of these passengers are commuters. Sumatra had the lowest mean number of trips, with a total of 404 thousand passengers. The peak number of train travelers in both the Jabodetabek and non-Jabodetabek districts was observed one month after the Eid al-Fitr celebration, namely in July 2019. This phenomenon occurs because, following Eid al-Fitr, a significant number of Muslims migrate from rural regions to metropolitan areas. In June 2019, Sumatra had the greatest passenger volume during the Eid al-Fitr month.

Table 1. Descriptive Statistics of Number of Train Passengers in Indonesia

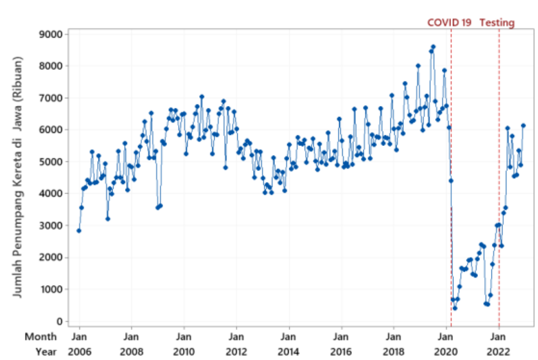
Category	Mean	St.Dev	Minimum	Q1	Median	Q3	Maximum
Jabodetabek	15937	7121	5077	10323	11792	22281	29714
Java	5056	1545	399	4486	5316	6030	8589
Sumatra	404	156.2	8	300.5	393	498.8	829

Figure 1 illustrates a consistent upward trend in the number of train passengers in Jabodetabek, Java, and Sumatra. This is seen in the trajectory of growth observed between January 2006 and December 2020. Nevertheless, this has been altered after the onset of COVID-19, which happened in March 2020. There was a significant decrease in the number of rail passengers across the entire region during that time period. This arises from government policies pertaining to social constraints, such as the Community Activity Restrictions (PPKM) policy. In addition, Figure 1 illustrates a significant surge in the number of rail passengers in all regions both before and after Eid al-Fitr.

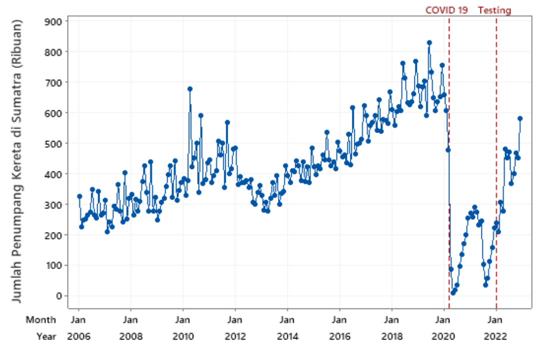
Figure 1 displays a consistent and upward trend in Islamic celebrations, indicating that dummy variables can be generated based on calendar fluctuations for time series regression modeling. Time series regression using calendar variations can detect seasonal patterns in specific months and Islamic holy celebrations. The reason for this is that Islam does not rely on the solar calendar to establish the timing of celebrations. Instead, it follows the Hijri calendar, which is decided by the the lunar position. Consequently, the holiday celebration takes place around 10 days earlier each year.



(a)



(b)



(c)

Figure 1. Time Series Plot Number of Train Passengers (a) Jabodetabek, (b) Java (Non Jabodetabek) and (c) Sumatra from January 2006 to December 2022

Prior to doing time series regression modeling using variations of the Islamic calendar and COVID-19 intervention, it is necessary to partition the data into training and testing sets. The training data serves as the foundation for time series regression modeling, taking into account variations in the Islamic calendar and COVID-19 interventions. On the other hand, the testing data is employed to assess the precision of the model's forecasted outcomes. Furthermore, it is necessary to partition the training data into two distinct segments: the training data prior to the COVID-19 intervention and the training data subsequent to it. This is implemented to mitigate the impact of COVID-19.

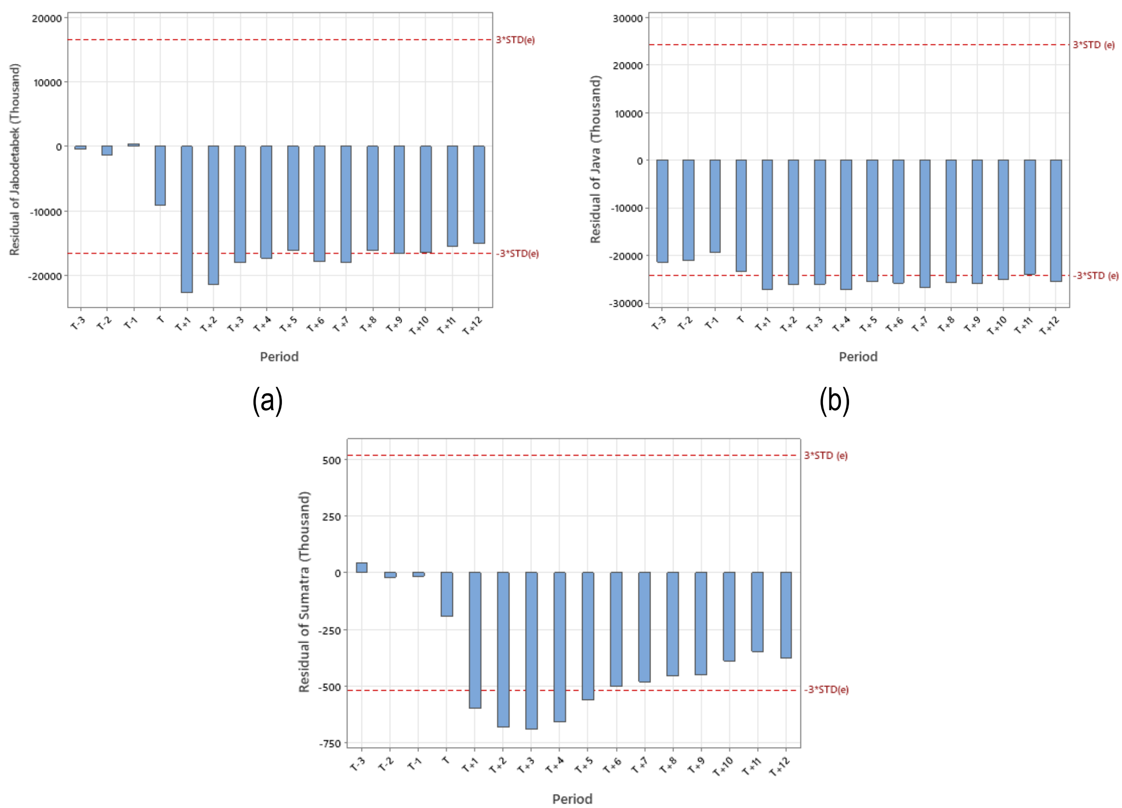
This study assumes the effect of calendar variation occurs a month before Eid al-Fitr, a month on Eid al-Fitr, a month after Eid al-Fitr, and a month on Eid al-Adha. Table 2. displays the results of time series regression modeling integrated the Islamic calendar variations.

Each category number of train passengers has a different number of significant parameters on the dummy variable of seasonality and , as indicated in Table 2. There are three, eleven, and two significant parameters, respectively, in models of time series regression integrated calendar variation of Jabodetabek, Java, and Sumatra. In addition, the ARIMA model distinguishes between each category. Both the Jabodetabek and Sumatra categories adapt to the SARIMA model, particularly SARIMA(1,1,[3])(1,0,0)12 and SARIMA (1,1,[2])(1,0,0)12. Meanwhile, Javanese (Non Jabodetabek) categories adhere to the ARIMA model (2,1,[3]). In addition, the diagnostic analysis indicates that all models satisfied the assumption of white noise residual (independent) at a significance level (α)=5%. In addition, the diagnostic checks for the assumption of normally distributed residuals confirms that only the Jabodetabek is significant at the significance level (α)=1%, whereas the Java (Non Jabodetabek) and Sumatra categories satisfy the normal distribution assumption at the significance level=5%.

After constructing a time series regression model with calendar variations, we are able to predict all training data prior to and following COVID 19. With data used for training predictions, residuals in the training data might be determined. The acquired residuals will subsequent be visualized to assess the influence of COVID-19. The residuals from the time series regression model with calendar variations for each category are presented in Figure 2.

Table 2. Time Series Regression Modeling Integrated The Calendar Variations

Category	Parameter Estimation			White noise		Normality Distribution
	Parameter	Estimated	P-value	To Lag	P-Pvalue	P-value
Jabodetabek	θ_3	-0.238	0.003	6	0.208	0.014
	ϕ_1	-0.339	<.0001	12	0.374	
	Φ_1	0.602	<.0001	18	0.135	
	β_{IF}	-841.299	<.0001	24	0.147	
	β_{IF+1}	536.292	0.011	30	0.258	
	β_2	-1478.400	<.0001			
Java (non Jabodetabek)	θ_3	0.271	0.003	6	0.913	0.125
	ϕ_1	-0.534	<.0001	12	0.720	
	ϕ_2	-0.260	0.005	18	0.962	
	β_{IF}	315.057	0.012	24	0.808	
	β_{IF-1}	-387.513	0.002	30	0.838	
	β_1	-680.904	<.0001			
	β_2	-1297.600	<.0001			
	β_3	-673.699	<.0001			
	β_4	-747.252	<.0001			
	β_5	-427.701	0.001			
	β_8	-695.624	<.0001			
	β_9	-637.382	<.0001			
	β_{10}	-478.289	0.001			
	β_{11}	-848.103	<.0001			
Sumatra	θ_2	0.443	<.0001	6	0.355	0.147
	ϕ_1	-0.752	<.0001	12	0.886	
	Φ_1	0.451	<.0001	18	0.149	
	β_{IF}	63.314	0.000	24	0.224	
	β_2	-57.127	0.007	30	0.413	



(a) (b)

(c)

Figure 2. Residual Plot of Time Series Regression with Calendar Variation on Number of Passengers in (a) Jabodetabek, (b) Java (Non Jabodetabek) and (c) Sumatra Area

Figure 2 demonstrates that, throughout all categories, COVID 19 exhibits an adverse effect on the number of train passengers. This can be determined by the significant number of lags that exceed the lower control limit of three standard deviations for the residual of training data. One month after discovering of the first case of COVID 19 in March, the effect of COVID 19 proceeded to reduce the number of train passengers. The number of train passengers steadily shrank resulting in an abrupt reduction in May 2020. This occurred because Eid al-Fitr was celebrated in that month. The effect of COVID 19 on the number of train passengers then gradually lessened, leading to an increase in train passengers. The growth in the number of train passengers was the consequence of the government's decision to lift social restrictions, which was followed by the implementation of a new normal policy in which all activities are conducted as usual on the requirement that all health procedures are executed.

Figure 2 indicates the impact of COVID 19 occurring in the T+1 period for each category, leading us to determine b=1. Order b in the intervention model is the duration of time required for the effect of a new intervention to come into effect, defined by the initial lag, which is capped at 3 standard deviations. Figure 2(a) and Figure 2(c) suggest that the effects of COVID 19 diminished, then eliminated entirely or that the lag was within the lower control limits. On the assumption of this structure, it could potentially be considered that both categories Jabodetabek and Sumatra each have a S order having the values s=3 and s=4, respectively. The order s displays how long the intervention's effect remains in place until the system recovers to a stable state. Figure 2(c) fails to represent the absence of an intervention's effect, consequently it could be regarded as having an order of r=1. The order r on the intervention has a significant duration of impact before stabilizing.

Table 3. Time Series Regression Modeling Integrated The Islamic Calendar Variations and COVID 19 Intervention

Category	Parameter Estimation			White noise		Normality
	Parameter	**Estimate**	**P-value**	**to Lag**	**P-value**	**P-value**
Jabodetabek	θ_1	0.228	0.001	6	0.118	
	θ_{15}	0.366	<.0001	12	0.686	
	Φ_1	0.991	<.0001	18	0.934	
	Φ_2	-0.446	<.0001	24	0.863	
	β_{IF}	-819.633	<.0001	30	0.965	
	β_2	-1769.700	<.0001	36	0.873	
	ϖ_{LS171}	-9741.600	<.0001			0.415
	ϖ_{LS91}	3719.600	<.0001			
	ϖ_{LS189}	4350.400	<.0001			
	ϖ_{AD122}	1574.000	<.0001			
	ω_0	-13820.900	<.0001			
	ω_2	-3770.900	<.0001			
	ω_3	-4741.000	<.0001			
Java (Non Jabodetabek)	θ_3	0.272	0.003	6	0.424	
	Φ_1	-0.220	0.006	12	0.220	
	ϕ_1	-0.471	<.0001	18	0.461	
	ϕ_2	-0.252	0.006	24	0.349	
	β_{IF-1}	-389.364	0.003	30	0.505	
	β_1	-553.797	0.001	36	0.121	
	β_2	-1275.800	<.0001			
	β_3	-786.557	<.0001			0.085
	β_4	-754.804	<.0001			
	β_5	-437.548	0.004			
	β_8	-793.289	<.0001			
	β_9	-711.089	<.0001			
	β_{10}	-487.248	0.003			
	β_{11}	-790.016	<.0001			
	ϖ_{LS176}	1831.200	0.000			

Category	Parameter Estimation			White noise		Normality
	Parameter	Estimate	P-value	to Lag	P-value	P-value
	ϖ_{AD171}	-1579.000	0.001			
	ϖ_{AD173}	-872.599	0.044			
	ω_0	-4556.200	<.0001			
	δ_1	0.217	0.045			
Sumatra	ϕ_1	-0.537	<.0001	6	0.353	>0.150
	Φ_1	0.564	<.0001	12	0.799	
	β_{IF}	60.579	<.0001	18	0.909	
	β_2	-45.827	0.027	24	0.816	
	β_8	-44.578	0.028	30	0.956	
	ϖ_{AD52}	295.807	<.0001			
	ω_0	-485.104	<.0001			
	ω_4	-188.008	<.0001			

All model parameters are significant at a significance level of 5%, as shown in Table 3. In diagnostic testing, two assumptions of residual white noise and normal distribution have also been satisfied. Figure 3 highlights that the estimated sequence of COVID-19 intervention models is different from the actual models in the Jabodetabek and Sumatra area. This is because some parameters are not significant at the 5% significance level. In contrast, the sequence of intervention models in the Java category is consistent with estimation. Before and after the intervention, the order of the seasonal autoregressive integrated moving average (SARIMA) in the Jabodetabek and Sumatra regions was different. Meanwhile, the autoregressive integrated moving average (ARIMA) model was replaced with the SARIMA model for Java due to seasonality. In addition, each category has additional dummy variables as a result of outlier detection. This approach is done because models built without additional dummy variables from outliers fail to satisfy the assumption of a 5% significance level that residuals are normally distributed. The presence of outliers can violate the assumption of the normal distribution of residuals. Therefore, additional dummy variables need to be implemented in the model to adjust for outliers. This study involved additive and level shift outliers. In addition, any time series regression model featuring calendar variations and COVID-19 interventions can be evaluated for the accuracy of its predictions using the RMSE.

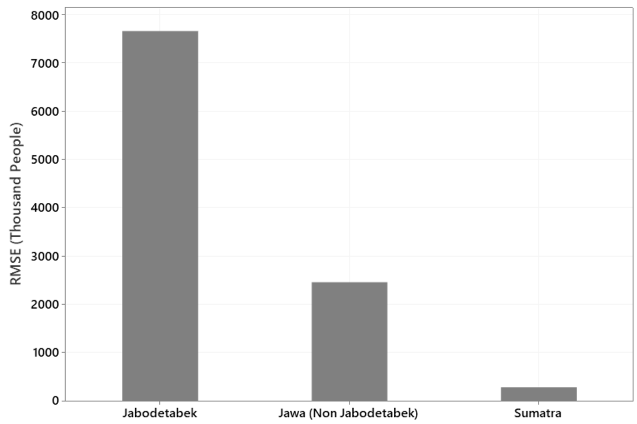
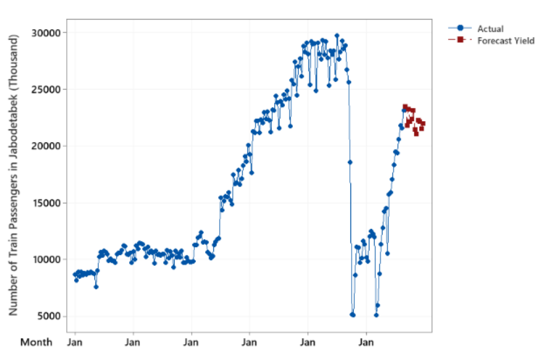


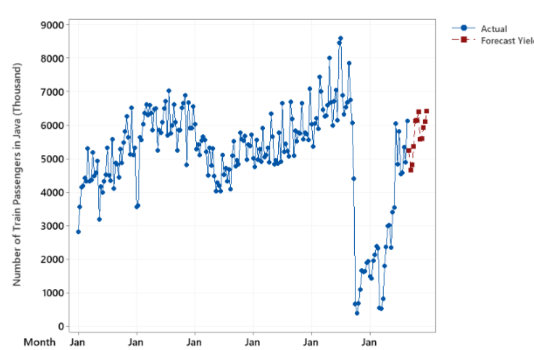
Figure 3. RMSE of Time Series Regression Integrated Calendar Variation and COVID 19 Intervention

Figure 3 illustrates the highest to lowest RMSE of the model number of train passengers using time series regression integrated calendar variations and COVID 19 interventions for Jabodetabek, Java (Non-Jabodetabek), and Sumatra, respectively. The RMSE value indicates the disparity between the actual value and the predicted values around that value. Furthermore, after accomplishing a particular degree of accuracy, it is feasible to forecast the number of monthly train passengers in 2023.

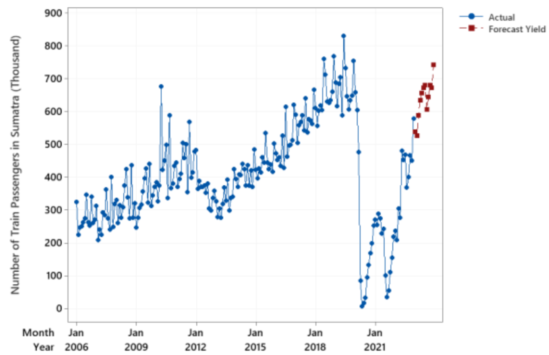
The anticipated statistics for the number of train passengers in Java (non-Jabodetabek) and Sumatra indicates an upward trend (Figure 4 (a) and Figure 4(b)). This is evident from the red chart's movement pattern, which exhibits an upward tendency. This growth is relating to the improving post-pandemic situation in Indonesia, which has enhanced accessibility and activity levels. Meanwhile, number of train passengers in Jabodetabek is expected to have fluctuated pattern. However, it has the same spike as Sumatra in April 2023. It is anticipated that the number of train passengers would rise tremendously because of Eid al-Fitr. Compared to the previous year, number of train passenger in the Jabodetabek and Sumatra areas jumped by 29% and 32%, respectively. In contrast, it is projected that train passengers in the Java will slightly decline by 1 percent compared to the previous year. With a growing number of train passengers, it is expected that KAI will have the capacity to enabled to keep providing safe, comfortable, and well-organized travel while minimizing the risk of contracting the COVID 19 virus.



(a)



(b)



(c)

Figure 4. Time Series Plot of Forecasting the Number of Train Passengers (a) Jabodetabek, (b) Java (Non Jabodetabek) dan (c) Sumatra

CONCLUSION

This study aims to analyze the forecasting model for the number of train passengers by integrating time series regression with Islamic calendar variations and COVID 19 effects. The source of data for this study is the number of train passengers in Jabodetabek, Java (Non-Jabodetabek), and Sumatra from January 2006 through December 2022. Time series regression integrating variations of the Islamic calendar and the effects of COVID 19 for Jabodetabek, Java (non-Jabodetabek), and Sumatra produces RMSE values of 7657,821; 2453.827 and 275.901 for each category, respectively. In general, the number of train passengers is seasonally for all categories (Jabodetabek, Java, and Sumatra). In Jabodetabek and Sumatra, Eid al-Fitr has a major impact on train passengers. One-month prior Eid al-Fitr has a significant effect on the amount of Java train passengers (Non Jabodetabek). Furthermore, COVID 19 had a meaningful impact on the number of train passengers for all categories.

REFERENCE

Andalita, I. & Irhamah. (2015). Peramalan Jumlah Penumpang Kereta Api Kelas Ekonomi Kertajaya Menggunakan ARIMA dan ANFIS. *Jurnal Sains dan Seni ITS*, *4*(2), doi:10.12962/j23373520.v4i2.11624.

Anggraeni, W., Vinarti, R. A. & Kurniawati, Y.D. (2015). Performance Comparisons Between Arima and Arimax Method in Moslem Kids Clothes Demand Forecasting: Case Study. *Procedia Computer Science*, *72*, 630-637. https://doi.org/10.1016/j.procs.2015.12.172

Fatimah, S. (2019). *Pengantar Transportasi*. Ponorogo: Myria Publisher.

Hyndman, R. J. & G. Athanasopoulos. (2018). *Forecasting: Principles and Practice 2nd ed*. Australia: Monash University.

Ismail, Z., Suhartono, Yahaya, A. & Efendi, R. (2009). Intervention Model for Analyzing the Impact of Terrorism to Tourism Industry. *Journal of Mathematics and Statistics*, *5*(4), 322-329. https://doi.org/10.3844/jmssp.2009.322.329

Nurjanah, I. S., Ruhiat, D., Andiani, D. (2018). Implementasi Model Autoregressive Integrated Moving Average (ARIMA) untuk Peramalan Jumlah Penumpang Kereta Api di Pulau Sumatera. *Teorema: Teori dan Riset Matematika*, *3*(2), 145-156. http://dx.doi.org/10.25157/teorema.v3i2.1421

Rianda, F. (2021). Pemodelan Intervensi Untuk Menganalisis dan Meramalkan Jumlah Penumpang

Pesawat di Bandara Soekarno-Hatta Akibat Pandemi Covid-19. *Prosiding Seminar Nasional Official Statistics*, 283-292. https://doi.org/10.34123/semnasoffstat.v2021i1.857

Silfiani, M., Hayati, F.N., Sari, S.P., & Prabowo, A. (2022). COVID-19 Intervention Model for Bali Province GRDP Prediction. *Eigen Mathematics Journal*, *5*(2), 86-91. https://doi.org/10.29303/emj.v5i2.141

Silfiani, M., Hayati, F.N., Nurlaily, D., & Fitria, I. (2021). Household Electrical Load Forecasting: a Hybrid of Linear Models and Radial Basis Function Neural Network. *International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA)*, 253-257. https://doi.org/10.1109/ICAMIMIA54022.2021.9807693

Suhartono, Lee, M. H., & Prasetyo, D. (2015). Two Levels ARIMAX and Regression Models for Forecasting Time Series Data with Calendar Variation Effects. AIP Conference Proceedings 1691, 050026. https://doi.org/10.1063/1.4937108

Susanti, A., Suhartono & Setyadi., H. J., Taruk, M., Hawiludin, Widagdo, P.P. (2018). Forecasting Inflow and Outflow of Money Currency in East Java Using a Hybrid Exponential Smoothing and Calendar Variation Model. *Journal of Physics: Conf*. Series 979, 012096. https://doi.org/10.1088/1742-6596/979/1/012096

Utomo, P. & Fanani, A. (2020). Peramalan Jumlah Penumpang Kereta Api di Indonesia Menggunakan Metode Seasonal Autoregressive Integrated Moving Average (SARIMA). *Jurnal Algebra*, *1*(2), 169–178. https://doi.org/10.29080/algebra.v1i2.6.

Wei, W. (1990). *Time Series Analysis: Univariate and Multivariate Methods*. New York: Addison-Wesley Publishing Company, Inc.

Wijiyanto, A. N., Kusrini, D. E., & Irhamah. (2012). Peramalan Nilai Kontrak Konstruksi PT ‘X’ dengan Menggunakan Pendekatan Regresi time series dan ANFIS. *Jurnal Saing dan Seni ITK*. *1*(1), D201-D206.

Wulansari, R. E. Suryanto, E., Ferawati, K., Andalita, I., & Suhartono. (2014). Uang Kartal Bank Indonesia Sebagai Solusi Kontrol Likuiditas Perbankan di Indonesia. *Statistika*, *14*(2), 59-68. https://doi.org/10.29313/jstat.v14i2.1203

Zukhronah, E., Sulandari, W., Slamet, I., Sugiyanto & Susanto, I. (2021). Model Variasi Kalender pada Regresi time seriesuntuk Peramalan Jumlah Pengunjung Grojokan Sewu. *Indonesian Journal of Applied Statistics*, *4*(2), 67-75.