The Forecasting of a Traffic Accident in the Pandemic of Covid-19

Hasmar Halim[a], Basyar Bustam[a], Zubair Saing[b]

[a] Associate Professor of Civil Engineering, State Polytechnic Of Ujung Pandang, Makassar.
[b]Professor of Civil Engineering, Muhammadiyah University of North Maluku, Ternate

Article History: *Do not touch during review process(xxxx)*

Abstract: Speaking of the increase in traffic accidents, transportation safety plays an in important role. This study aims to analyze the number of traffic accidents in Makassar City. On the other hand, the implementation of Large-Scale Social Restrictions (LSSB) to prevent the spread of Covid-19 has an impact on transportation movements in Makassar City. According to this, the traffic accident data used is accident data for 2016 to 2020, which the data for 2020 is data on traffic accidents affected by the LSSB policy. From the results of data analysis, it is known that the accident rate trend tends to follow seasonal patterns so that the model used is the SARIMA Model. Sarima's model with a period (1,1,1) $(1,1,1)^6$ is assumed to be the best model with a MAPE value of 81.6%. Based on the parameters generated by this model, it shows that the number of accidents in 2021 has decreased significantly. This is due to the government policy against the spread of the Covid-19 virus, which is implementing the PSBB. This model cannot be utilized in forecasting for a long period. This is because a long period can cause large estimates that fluctuate in value and even have a tendency for the accident rate to be negative.

Keywords: Traffic Accidents, Forecasting, SARIMA, MAPE

1. Introduction

At the beginning of 2020, the world was shocked by the outbreak of a new virus, namely Corona virus 2019 or COVID-19. It is known that this virus originated in Wuhan, China, which was discovered at the end of December 2019. Until now, the death rate has reached nearly 2 million deaths and it is confirmed that 65 countries have contracted this virus. The latest data, dated January 11, 2021, shows that the number of people infected with Covid-19 is 22,410,069 peopleas stated by**WHO**, (2021).Data reported by**Kompas Cyber Media**, (2021) in Indonesia, 818,386 people have been infected with the number of deaths reaching 23,947 people, while in Makassar City the death rate due to this virus has reached 394 people out of 18,250 suffered peopleas stated by**Makassar Government**, (2021). Of course, Covid-19 has a very significant impact on people's lives both economically, educationally and socially. The Indonesian government policy regarding the handling of this disease implements a policy called Large-Scale Social Restrictions (LSSR). **Minister of Health of Republic of Indonesia**, (2020), This policy aims to prevent the possible spread of Covid-19. The application of large-scale social restrictions includes a. school and work holidays; b. restrictions on religious activities; c. restrictions on activities in public places or facilities; d. restrictions on social and cultural activities; e. restrictions on modes of transportation; f. restrictions on other activities, especially those related to defense and security aspects. This policy restriction has an impact on transportation, which is the decrease of transportation movement. On the other hand, the rapid development of transportation has indirectly increased the risk of road accidents. Accidental and unplanned road traffic accidents involving vehicles with or without other road users result in death, injury and material loss. The world's high number of accidents has made it the most important issue in the last two decades, even WHO predicts that every year traffic accidents cause an average death rate of 1.24 million people per year with an increase in the number of accidents by 77%as stated by**Inada et al**., (2020).

To predict the growth rate of the accident rate in the future, it is necessary to make forecasting. Forecasting is one of the most significant elements in decision making because the effectiveness of a decision generally depends on several factors that cannot be seen when making decisions. Forecasting functions to identify or predict the number of accidents and to lessen uncertainty about something that will possibly occur in the future. With this forecast, actual steps can be determined to prevent future accidents to happen. In addition, one method that can be used for forecasting analysis is statistical methods.

2.Review Of Literature

Eze, Asogwa, & Okonkwo(2018)conducted a study on the modelling of accident cases on four main roads to the main city from the Enugu State of Nigeria using the SARIMA Model. Among the most powerful approaches to Time-series data analysis uses the Autoregressive Integrated Moving Average (ARIMA) model. In this paper, build a SARIMA model for accident cases from January 2007 to December 2015 with a total of 108 data points. The model obtained in this paper is used to estimate the monthly cases of accidents on each of the roads for the coming year 2016. The expected results will help Government and Federal road safety commissions figure out how to maintain order on the highway to reduce the cases of road traffic accidents along the road

Manikandan M, Prasad R, Mishra, Konduru, et al,(2018)have predicted the Road Traffic Accident mortality rate (RTA) in India using the seasonal regression automatic moving average (SARIMA) model. This study concludes that there is an increasing trend in the estimated number of road traffic accidents and also shows the seasonal RTA deaths with more accidents during the months of April and May each year. It is recommended that policymakers and transport authorities pay more attention to road traffic accidents and plan some effective interventions to reduce the death burden of RTA**.**

Another study conducted in Kerala, India ran by **Sunny**, et al., (2018)in his research discusses the problem of predicting road accidents using time series analysis in all districts in Kerala. Time series analysis is useful for finding trends in road accidents which allows the prediction of future patterns. In the current MS, we used time-series traffic accident data in Kerala, India over the period January 1999 - December 2016 to understand patterns in the data and to develop suitable models to predict future patterns that would enable the authorities to take preventive measures. We subset the data up to December 2013 as training data for model selection and other data is used for model assessment. The two models discussed here are “Holt-Winters (HW) exponential smoothing” and “Seasonal ARIMA (SARIMA)”. Both models will provide approximate values in the confidence interval of the test data

suggested that traffic accidents are the leading cause of death in developing countries. Deaths from traffic accidents are assessed through an estimated time series of three years. This study is to use trend assessment to predict deaths due to traffic accidents from January 2013 to December 2015 in Kermanshah province, Iran. This study concludes that there is an observed downward trend in accidental deaths. The highest and lowest deaths were seen each year in the spring and fall, respectively. The SARIMA $(0,0,0) \times (1,1,1)^{12}$ model was identified as the most suitable model for the data. The prediction value for traffic accidents shows a decreasing trend in deaths in the coming years.

Al-Hasani, Khan, Al-Reesi, & Al-Maniri(2019)conducted this research by evaluating the optimal time series model for determining the sequence of parameters in road traffic accident data compared to manual processes. Time series traffic accident data were collected over eighteen years from secondary sources, and statistical time series analysis was performed. The decomposition of time series, stationarity and seasonality is examined to identify a suitable model for road traffic accidents, while optimal data analysis is carried out by comparing the two results. AIC, BIC and other error values are used to select the best model and model diagnostic tools are applied to confirm statistical assumptions. This study concludes that the SARIMA $(0,1,2)\ (1,0,2)^{12}$ and SARIMA $(0,1,2)\ (0,0,2)^{12}$ models produce the best models manually and automatically. The diagnostic process shows that the SARIMA model $(0,1,2)\ (1,0,2)^{12}$ is better than the optimal model. Therefore, the modeller who prefers to use the optimal function as a time series model selection tool must consider the accuracy of the model. It would be preferable for the rater to compare various models and select the most suitable one.

Merabet & Zeghdoudi, (2020), This research is to study and model a number of accidents due to road accidents in the Skikda area (northeastern Algeria) according to the Box-Jenkins method using the EViews software using serial data from January 2001 to December 2016. In addition, the Kalman filter method is also provided. For this purpose, the Kalman filter method is used for the purpose of short-term prediction and parametric identification. On the other hand, a comparative study was provided to compare the two methods with the following criteria: Mean absolute percentage error (MAPE), root mean square percentage error (RMSPE) and U Theils statistics. This application uses Eviews 5.0 and SPSS 26 software.

3.Objectives Of The Study

- To analyze the predicted number of accidents in the future by including the number of accidents during the Covid-19 pandemic.
- applying the SARIMA forecasting method to traffic accident data in Makassar City

4. Methodology Of The Study

Collecting accident data is done a series of time for 60 months from 2016 to 2020, including accident data for 2020, which we know as the year of the Covid-19 Pandemic. This data source contains the number of traffic accidents related to pedestrians, motorized vehicles, non-motorized vehicles, the number of accidents and the number of accident victims, deaths, serious and minor injuries, as well as the characteristics of the victims involved.

This method is used in analyzing data using a time series approach model, often referred to as seasonal ARIMA or SARIMA. SARIMA stands for Seasonal Auto-Regressive Integrated Moving Average Model. The average model, which consists of a moving average (MA) and an autoregressive model (AR).

5.Analysis And Interpretation

The data obtained from the traffic accident unit were 60 months from January 2016 to December 2020 as illustrated in Figure 1 (a). Figure (1a) illustrates that the data distribution of the number of traffic accidents does not appear to be stationary because these points do not fluctuate around the constant mean and variance and have a probability value of 0.9359 which means > 0.05 as in Table 1. Because the data is not stationary, it is necessary to do differences. The results are as in Figure (1b) below:

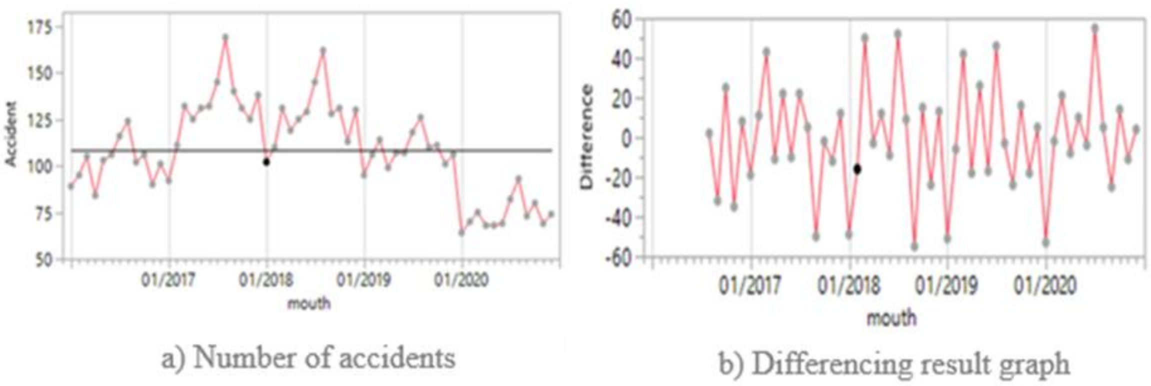
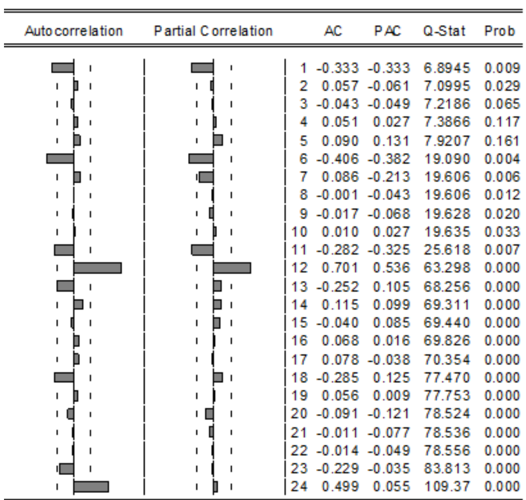


Figure 1 Number of Traffic Accidents in Makassar City Before and After Differencing

The differencing results produce a graphic pattern such as in Figure 1 (b) where it appears that the data has fluctuated around a constant mean and variance so that the data on the number of traffic accidents is said to be stationary. This can also be seen in the results of the ADF test to see the stationary of the data on the mean and variance. The results of the ADF test using Eviews 10 software are presented in Table 1.Based on the p-value of Table 1, it can be seen that the p-value is < 0.05, so it can be concluded that the data on the number of traffic accidents is stationary in terms of mean and variance, and then it is formed ACF and PACF plots. ACF and PACF plots are presented in Figure 2.

Table 1 Stationary test using ADF

Augmented Dickey-Fuller test statistic		Before differencing		After differencing	
		t-Statistic	Prob.*	t-Statistic	Prob.*
		-0.160125	0.9359	-10.59626	0.0000
Test critical values:	1% level	-3.588509		-3.548208	
	5% level	-2.929734		-2.912631	
	10% level	-2.603064		-2.594027	



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.333	-0.333	6.8945	0.009
		2	0.057	-0.061	7.0995	0.029
		3	-0.043	-0.049	7.2186	0.065
		4	0.051	0.027	7.3866	0.117
		5	0.090	0.131	7.9207	0.161
		6	-0.406	-0.382	19.090	0.004
		7	0.086	-0.213	19.606	0.006
		8	-0.001	-0.043	19.606	0.012
		9	-0.017	-0.068	19.628	0.020
		10	0.010	0.027	19.635	0.033
		11	-0.282	-0.325	25.618	0.007
		12	0.701	0.536	63.298	0.000
		13	-0.252	0.105	68.256	0.000
		14	0.115	0.099	69.311	0.000
		15	-0.040	0.085	69.440	0.000
		16	0.068	0.016	69.826	0.000
		17	0.078	-0.038	70.354	0.000
		18	-0.285	0.125	77.470	0.000
		19	0.056	0.009	77.753	0.000
		20	-0.091	-0.121	78.524	0.000
		21	-0.011	-0.077	78.536	0.000
		22	-0.014	-0.049	78.556	0.000
		23	-0.229	-0.035	83.813	0.000
		24	0.499	0.055	109.37	0.000

Figure 2 Plot of ACF and PACF Number of Traffic Accidents in Makassar City

Based on the ACF and PACF plots in Figure 2, it can be seen that the ACF and PACF plots are cut off after lag 1 and there is a seasonal pattern in the 6th period so that it can be assumed that there are several possible SARIMA models, namely SARIMA (1,1,0) $(1,1,1)^6$, SARIMA (0,1,1) $(1,1,1)^6$ and SARIMA (1,1,1) $(1,1,1)^6$.

5.1Estimated Parameters and Best Model Selection

Estimated model parameters SARIMA (1,1,0) $(1,1,1)^6$, SARIMA (0,1,1) $(1,1,1)^6$ and SARIMA (1,1,1) $(1,1,1)^6$ and selecting the best model using the values of Rsquare, RSquare Adj, MAPE and MAE is presented in Table 2.

Table 2 SARIMA Model Estimation

	Model	DF	Variance	AIC	SBC	RSquare	Weights	MAPE	MAE
	Seasonal ARIMA (1, 1, 1)(1, 1, 1)6	48	61,849189	392,76948	402,62094	0,816	0,413565	7,213366	7,539039
	Seasonal ARIMA (0, 1, 1)(1, 1, 1)6	49	68,391544	393,38034	401,26151	0,820	0,304717	7,223064	7,547770
	Seasonal ARIMA (1, 1, 0)(1, 1, 1)6	49	68,428253	393,53729	401,41846	0,818	0,281719	7,324847	7,654990

From Table 2 it is known that the weight value of 0.4135 for the SARIMA (1,1,1) (1,1,1)6 model is the largest value of the three models used to predict the accident rate. Thus it can be concluded that the best model of this accident pattern is to use the SARIMA (1,1,1) $(1,1,1)^6$ model. From the table, it is also known that the SARIMA model (1,1,1) $(1,1,1)^6$ has a coefficient of determination (R-square) of 0.816, which means that this forecasting model has a match of 81.6% to the actual model. The parameters of the SARIMA model (1,1,1) $(1,1,1)^6$ are as shown in Table 3 below.

Table 3 Estimated Parameters

Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob
AR1,1	1	1	0,7353	0,1297	5,67	0,0001
AR2,6	2	6	-0,8872	0,0533	-16,64	0,0001
MA1,1	1	1	1,0000	0,0772	12,96	0,0001
MA2,6	2	6	0,4720	0,1903	2,48	0,0167
Intercept	1	0	-0,7302	0,1053	-6,93	0,0001

5.2 Diagnostic Checking

Diagnostic checking using the Ljung-Box test is presented in Table 4. From this table it is known that the Ljung-Box value of each lag has a value smaller than the Chi-square value and the value □ > 0.05, thus it can be concluded that the model SARIMA (1,1,1) $(1,1,1)^6$ is the best model and can be used to forecast traffic accident rates in Makassar City.

Table 4 Estimated Parameters

Lag	Df	AutoCorr	Ljung-Box Q	χ^2	ρ-Value
6	4	0,0440	16,5130	16,9190	0,0513
12	10	-0,0073	19,9172	18,3070	0,0687
18	16	-0,0289	22,8796	26,2962	0,1953
24	22	0,1452	28,0022	33,9244	0,2599

6. Finding Of The Study

Based on the verification results and the MAPE value on the SARIMA model (1,1,1) (1,1,1)6 which shows that the SARIMA model (1,1,1) (1,1,1)6 can be applied to the number of accidents data Makassar City traffic. The results of forecasting the number of traffic accidents in Makassar City are presented as in Figure 3.

The process of forecasting the number of traffic accidents in Makassar City for the period January 2016 to December 2020 using the SARIMA (1,1,1) (1,1,1) 6 model is considered good, because the forecast results obtained are close to the original data. This supports previous research that uses the same analysis method but with different model namely the SARIMA Model (1,0,0) (2,1,0) 12, this study concludes that SARIMA can be used to forecast the death rate due to traffic accidents **Yadollahi & Gholamzadeh**, (2019), **Igissinov et al**., (2020) and **Yousefzadeh-Chabok et al**., (2016). The same thing was done by **Zolala et al**., (2016)**, Ihueze & Onwurah**, (2018) and**Shao et al**., (2019)the results of their research using the SARIMA model (0.0,0) $(1,1,1)^{12}$ is worked to identify a reduction in the number of deaths due to traffic accidents.

Based on Figure 3, it can be seen that the number of traffic accidents in Makassar City during the next year is expected to decrease. This decrease in the number of traffic accidents is likely due to the policy related to the Covid-19 pandemic, namely the Large-Scale Social Restriction (LSSR) policy. These restrictions include school and work vacations, restrictions on religious activities, restrictions on activities in public places or facilities, restrictions on socio-cultural activities, restrictions on transportation modes, and restrictions on other activities specifically related to defense and security aspects. This has an impact on the lack of movement of traffic on the road which moderates the number of accidents. Although this model is statistically compatible with traffic accident data, in this particular case the 2020 data can be said to be missing data due to a negative trend in 2021. So this model cannot be used in forecasting for a long period. . This is because a long period can cause large estimates that fluctuate in value and even have a tendency for the accident rate to be negative.

Figure 3. Prediction of the Number of Traffic Accidents in Makassar City using the SARIMA Model

. 7. Conclusion

This study uses the SARIMA method to predict the number of road traffic accidents. By using statistical tools, namely statistical software JMP Pro 13, the analysis obtained is to determine the precise model for forecasting. The model is selected based on several criteria, namely the value of Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) and the weight of each model being tested. The greatest weight value is produced by the SARIMA model (1,1,1) (1,1,1)6 with a weight of 0.413565 and from the diagnostic examination with the Ljung-Box test, it is known that the Ljung-Box Q value has a value that is smaller than the value chi quadrat. Hence, this model is the best model to be used in predicting accident rates. Forecasting results show that the number of traffic accidents tends to decrease. Although this model is statistically compatible with traffic accident data, in general, this data has anomalies from previous data due to policies related to the Covid-19 pandemic, namely the Large-Scale Social Restrictions (LSSR) policy.

References

1. Al-Hasani, G., Khan, A., Al Reesi, H., & Al-Maniri, A. (2019). Diagnostic Time Series Models For Road Traffic Accidents Data. *International Journal of Applied Statistics and Econometrics*, 2, 19–26.
2. Eze, C. M., Asogwa, O. C., & Okonkwo, C. R. (2018). On the Modelling of Road Traffic Crashes: A case of SARIMA Models. *GSTF Journal of Mathematics, Statistics and Operations Research*, 5, 15–35.
3. Halim, Hasmar, Mustari, Ismail, Mursalim, & Saing, Zubair. (2020). Overtake Behavior Model of Motorcycle on Accident Risk. *International Journal of Advanced Science and Technology*, 29(5), 12556–12567.

4. Igissinov, N., Aubakirova, A., Orazova, G., et.al Akpolatova, G., Urazova, S., Tarzhanova, D., Zhantureyeva, A., & Kuandykov, Y. (2020). Prediction Mortality Rate Due to the Road-Traffic Accidents in Kazakhstan. *Iranian Journal of Public Health*, 49(1), 68–76.
5. Ihueze, C. C., & Onwurah, U. O. (2018). Road Traffic Accidents Prediction Modelling: An analysis of Anambra State, Nigeria. *Accident Analysis & Prevention*, 112, 21–29.
6. Inada, H., Li, Q., Bachani, A., & Hyder, A. A. (2020). Forecasting global road traffic injury mortality for 2030. *Injury Prevention*, 26(4), 339–343.
7. Kompas Cyber Media. (2021). *Today's Latest News, Covid-19 Data in Indonesia*. KOMPAS.com. https://www.kompas.com/[Accessed 12 January 2021].
8. Makassar Government. (2021). *Makassar City Covid-19 Prevention Information*. https://infocorona.makassar.go.id/. [Accessed 21 January 2021].
9. Manikandan, M., Vishnu, P. R., Amit, K. M., Rajesh, K. K., & Newtonraj, A. (2018). Forecasting Road Traffic Accident Deaths In India Using Seasonal Autoregressive Integrated Moving Average Model. *International Journal Of Community Medicine And Public Health*, *5*(9), 3962–3968.
10. Merabet, F., & Zeghdoudi, H. (2020). On Modelling Seasonal Arima Series: Application on Road Accidents in Northeast Algeria. *WSEAS Transactions On Systems And Control*, 15, 235–246.
11. Minister of Health of Republic of Indonesia. (2020). *Regulation of the Minister of Health of the Republic of Indonesia Number 9 of 2020. Guidelines for Large-Scale Social Restrictions in the Context of Accelerating the Handling of Corona Virus Disease 2019 (Covid 19)*. Minister of Health of Republic of Indonesia.
12. Shao, X., Boey, L. L., & Luo, Y. (2019). Traffic Accident Time Series Prediction Model Based on Combination of ARIMA and BP and SVM. *Journal of Traffic and Logistics Engineering*, 41–46.
13. Sunny, C. M., Nithya, S., Sinshi, K. S., Vinodini M.D., V., Lakshmi K.G., A., Anjana, S., & Manojkumar, T. K. (2018). Forecasting of Road Accident in Kerala: A Case Study. *2018 International Conference on Data Science and Engineering (ICDSE)*, 1–5.
14. WHO. (2021). WHO *Director-General's opening remarks at the media briefing on COVID-19 – 11 January 2021*. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-11-january-2021. [Accessed 12 January 2021].
15. Yadollahi, M., & Gholamzadeh, S. (2019). Five-Year Forecasting Deaths Caused by Traffic Accidents in Fars Province of Iran. *Bulletin of Emergency & Trauma*, 7(4), 373–380.
16. Yousefzadeh-Chabok, S., Ranjbar-Taklimie, F., Malekpouri, R., & Razzaghi, A. (2016). A Time Series Model for Assessing the Trend and Forecasting the Road Traffic Accident Mortality. *Archives of Trauma Research, In Press*.