International Journal of Trends in Mathematics Education Research
Vol. 5, No. 2 (2022), pp. 191-198
ISSN 2621-8488 (online)
DOI: https://doi.org/10.33122/ijtmer.v5i2.138

SAINTIS publishing

Research Article

Forecasting COVID-19 Cases for Top-3 Countries of Southeast Asian Nation

Hedi[1]*, Anie Lusiani[2], Anny Suryani[3] & Agus Binarto[4]

[1] Department of Energy Conversion Engineering, Politeknik Negeri Bandung, Indonesia
[2] Department of Mechanical Engineering, Politeknik Negeri Bandung, Indonesia
[3] Department of Accounting, Politeknik Negeri Bandung, Indonesia
[4] Department of Electrical Electronic Engineering, Politeknik Negeri Bandung, Indonesia
Corresponding Author: hedi@polban.ac.id | Phone Number [+6285220043920]

Received: 19 April 2022 *Revised: 27 May 2022* *Accepted: 20 June 2022* *Available online: 30 June 2022*

ABSTRACT

Several countries continue controlling the spread of the corona virus to decrease the number of new COVID-19 cases. Currently some Southeast Asian countries require an estimate of the num-ber of daily new COVID-19 cases of in the future in order to reopen or consider lifting strict pre-vention policies. This study applies ARIMA and SARIMA forecasting models to predict the de-cline in the number of new cases in three Southeast Asian countries. The first modelling is carried out using the ARIMA model with optimized model parameters based on the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) analysis. Then, the Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percent Error (MAPE) are evaluated is applied a criterion to select the best model. The best ARIMA and SARIMA models are selected manually and they are used to predict the number of new cases in three Southeast Asian coun-tries. It is expected that the number of new cases in these countries will experience a significant decline in the next month from September 2021. The prediction of SARIMA model indicates a better result than the ARIMA model which confirms the existence of a season in COVID-19 data.

Keywords: COVID-19; ARIMA; SARIMA; Prediction

1. INTRODUCTION

The world is currently facing serious problems covering the social, economic, and financial spheres because more than two-year-old COVId-19 pandemic has not yet to come to the end (Mehta, 2021). COVID-19 has killed more than 4,500,000 people worldwide. As there is no specific drug or vaccine, people are advised to prevent and fight COVID-19 naturally, such as using personal protective equipment, having social-distancing, and lockdowns (Arun Kumar et al., 2021). It is not clearly known how this epidemic are spreading, and how effective is the prevention and control measures should be taken (Moreau, 2020). The increase in the number of new cases depends on the number of crowds (Zhang et al., 2020). The five largest occupational groups as the spreading centres of COVID-19 are health care, transport, sales, cleaning and public safety (Lan et al., 2020).

This pandemic will experience a significant decline if all countries implement the importance of social distancing, wear protective equipment, and avoid crowds. Control of the spread of COVID-19 carried out by Indonesia (IDN) has succeeded in reducing the number of new cases, see Figure 1. While the Philippines (PHL) and Malaysia (MYS) have not been able to control this pandemic. Forecasting the number of new-cases a month later in three Southeast Asian countries is urgently needed to loosen strict policies, so that the negative impact of the spread of these cases can be minimized.

Time series modelling is very helpful in predicting the number of new cases in the short and long term. Various time series models are used to predict the number of future cases and predict the spread of infectious diseases in the near future (Yonar, 2020). Several researchers applied the ARIMA model to predict the short-term spread of COVID-19 (Benvenuto et al., 2020) and forecasting of COVID-19 confirmed cases in different countries (Dehesh et.al., 2020). Recently, Mohan have proposed a hybrid ARIMA and Prophet Model to predict daily confirmed and cumulative confirmed cases of COVID-19 in India (Mohan et al., 2022).

In finances and economics, ARIMA has been widely used in forecasting time series data on the Rupiah currency (Oenara & Oetama, 2021), study about extensive process of building stock price predictive model (Ariyo et al., 2014), and predict the demand (Fattah et al., 2018), gold price (Guha & Bandyopadhyay, 2016), and agricultural price (Jadhav et al., 2017). ARIMA can be combine with the Convolutional Neural Network (CNN) and the Long Short Term Memory (LSTM) network to forecast the carbon futures price (Ji et al., 2019). On the other hand, SARIMA Model also can be used in finances and economics. Gikungu have studied about forecasting inflation rate in Kenya (Gikungu et al., 2015) and Divisekara predict the red lentils commodity market price (Divisekara et al., 2021) using SARIMA Model.

Furthermore, several researchers have compared ARIMA and SARIMA Models. The results show that the SARIMA Model is better than the ARIMA Model (Rivero et al., 2020) and (Chen & Yu, 2020). In 2021, we have studied to estimate the number of COVID-19 new cases in Indonesia applying two mathematical models ARIMA and SARIMA (Suryani & Binarto, 2021). Now, we study about applications of ARIMA and SARIMA forecasting Models to predict the decline in the number of new cases in three Southeast Asian countries.

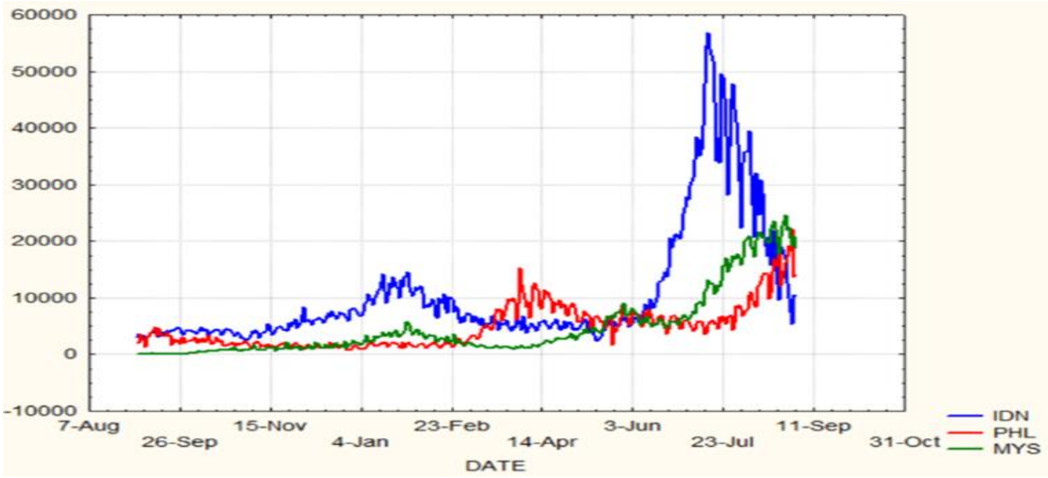


Figure 1. Daily New Cases COVID-19

In this study, the prediction this pandemic will be applied to determine for three countries with the highest number of confirmed cases in Southeast Asia. This study uses the number of new COVID-19 cases reported from September 8, 2020 to September 1, 2021, and the data are obtained from the Johns Hopkins coronavirus data centre. Furthermore, the forecast for this pandemic in the following month from September 2021 using the ARIMA and SARIMA Models.

2. RESEARCH METHOD

2.1 ARIMA Model

If the time series yt,, t = 1, 2, 3, …T, where T is the number of observations, and yt, is stationary after transformation with a difference of d times, this time series can be predicted using ARIMA(p, d, q). Then p and q are determined based on the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots for ARIMA the number of models to be assessed is a permutation of the p and q parameters. The general form of this equation is

$$\phi_p(B)(1-B)^d(Y_t-\mu)=\theta_q(B)\varepsilon_t \quad (1)$$

where:

$\phi_p(B)=1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p;$

$\theta_q(B)=1+\theta_1 B+\theta_2 B^2+\cdots+\theta_q B^q;$

t = 1,2,....T, with T = number observed; B = backshift operator; d = differenced level; μ = mean; and ε_t= residual.

2.2 SARIMA Model

The SARIMA Model is an ARIMA Model development with seasonal additions (P, D, Q)m and m the number of time steps corresponding to one seasonal period. Parameter D was determined based on the seasonal data transformation through differencing D times. Meanwhile, P and Q were determined based on the ACF and PACF plots, the general form of the equation is

$$\phi_p(B)(1-B)^d\Phi_P(B^m)(Y_t-\mu)=\theta_q(B)\Theta_Q(B^m)\varepsilon_t \quad (2)$$

$$\Phi_P(B^m)=1-\Phi_1 B^m-\Phi_2 B^{2m}-\cdots-\Phi_p B^{Pm}$$

$$\Theta_q(B^m)=1+\Theta_1 B^m+\Theta_2 B^{2m}+\cdots+\Theta_Q B^{Qm}$$

2.3 Identify Model and Evaluation Model

Identification Model begins with making ACF and PACF plots. From this plot, the probability of parameters p and q in the ARIMA model and p, q and P and Q in the SARIMA model are obtained. These parameters are selected based on the AIC and BIC values, the suitable model is the one with the smallest AIC and BIC values. Mathematically AIC is expressed in the equation $AIC = -2\log(L(\hat{\theta})) + 2k$ (3)
where $\log(L(\hat{\theta}))$ states the likelihood function and k is the number of model parameters. Another model criterion is BIC, this criterion considers the number of observations, the equation is $BIC = -2\log(L(\hat{\theta})) + k\log(T)$ (4)
The evaluation of the two models in this study was carried out by calculating MAE, RMSE, and MAPE. The equation for the three errors is as follows

$$MAE = \frac{1}{T}\sum_{i=1}^{T}|y_i - \hat{y}_i| \quad (5)$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(y_i - \hat{y}_i)^2} \quad (6)$$

$$MAPE = \frac{1}{T}\sum_{i=1}^{T}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (7)$$

where, $\hat{y}_i$ is predicted value, y_i is actual value.

2.4 Methods

This study uses daily data from September 8, 2020 to September 1, 2021. The first step is to test the stationarity of the data by applying the Augmented Dickey-Fuller (ADF) test. For non-stationary data, transformation is carried out by applying differencing data. The next step is to identify stationary data by making ACF and PACF plots. This plot determines the possible parameters of the ARIMA model. These parameters are selected based on the results of the smallest AIC and BIC calculations see Figure 2. Based on the significant ACF and PACF plots at lag m, the SARIMA model was developed with seasonal AR order P and seasonal MA order Q. The parameters were selected based on the smallest AIC and BIC values. Furthermore, the selection of the best model is determined with the smallest value calculated from MAE, RMSE, and MAPE. Based on the best model, the forecast for this pandemic is determined the following month starting from September 2021.

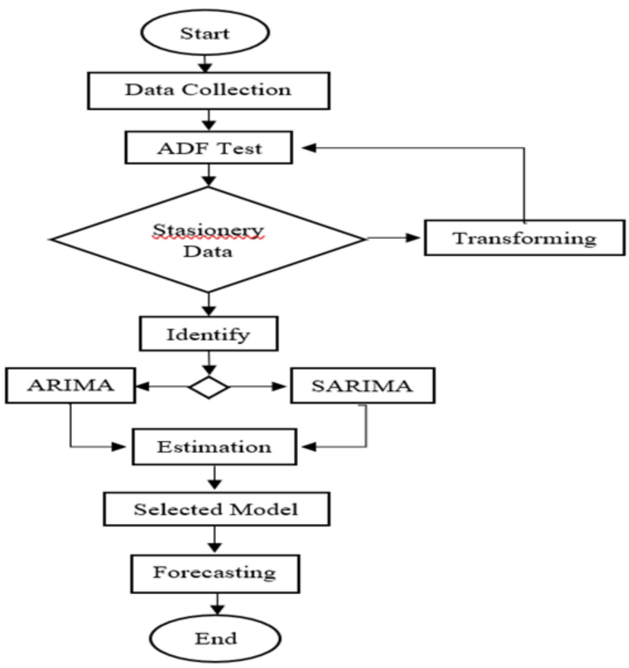


Figure 2. The process of data analysis using ARIMA and SARIMA Models

3. RESULTS AND DISCUSSION

The time series data which states daily data on new COVID-19 cases from the three most Southeast Asian countries, starting on September 8, 2020 to September 1, 2021 (number of observations of 366 days) is shown in **Figure 1**. The first step is a stationary test at each time data series yt with ADF test. Based on the test on each data, the results are not stationary. Furthermore, each time series data set is transformed through first differencing 〖Dy_t. Based on the results of the ADF test on the transformed data, the t-statistic value is below the critical value (see **Table 1**), which means that the three time series data are stationary. Thus, the three data prediction models are ARIMA (p, 1, q).

The next step is to identify a model to predict the possible values of p and q through the ACF plot and the PACF. Figure 2 shows the plot of the three time series data. The time series parameters p and q for top-3 Countries of Southeast Asian Nation are estimated to be 1, 2, 3 and 4. Based on the ACF and PACF plots, a high correlation found at the 7th lag which indicating the existence of a seasonal every 7 days, making it possible to apply the SARIMA model with m = 7.

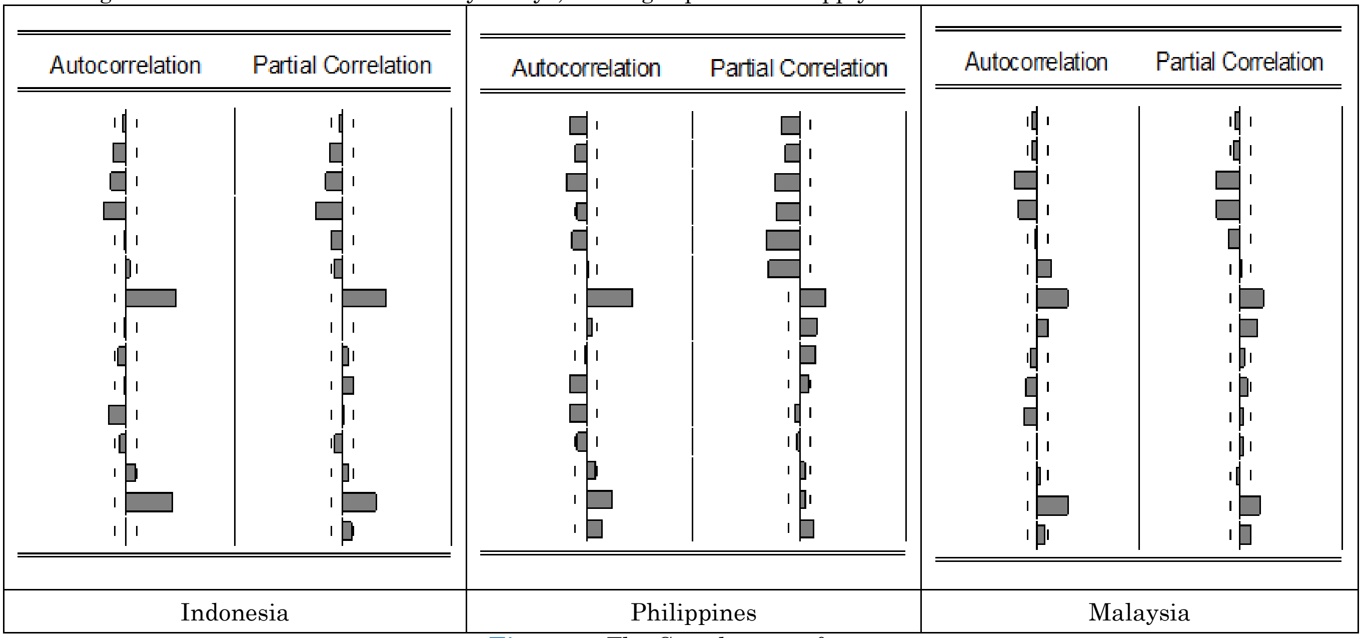


Figure 3. The Correlogram of Dy_t

The SARIMA approach is applied to Dy_t stationary data by entering a 7-day seasons, permutations of parameters p, q, P and Q, with p = i, q = j, P = k and Q = l where i = 1, 2, 3, 4; j = 1, 2 , 3, 4; k = 1, 2 and l = 1, 2, the number of possibilities is 64, from these 64 possibilities the AIC and BIC values are determined. The results of the number of new cases in Indonesia are indicated in pairs of p = 3, q = 4, P = 2 and Q =1the Philippines of p = 2, q = 2, P = 2 and Q = 1 and Malaysia of p = 4, q = 4, P = 1 and Q = 1 see Table 3. Based on the percent error values that are not much different in tables 1 and 2, the ARIMA and SARIMA models can explain the number of new COVID-19 cases for 3 major Southeast Asian countries, namely Indonesia, the Philippines, and Malaysia. The actual value vs the predicted model value is shown in the **Figure 3**. The form of the prediction model is the same as the actual value of the three countries.

Table 2. The Selected ARIMA Models

Country	ARIMA	AIC	BIC	RMSE	MAE	MAPE
Indonesia	(4, 1, 4)	18.311594	18.418659	2254.189	1271.623	11.48939
Philippines	(2, 1, 3)	16.675406	16.750351	989.8211	623.0198	15.59873
Malaysia	(2, 1, 3)	16.750351	15861035	634.2175	413.0340	32.82712

Table 3. The Selected SARIMA Models

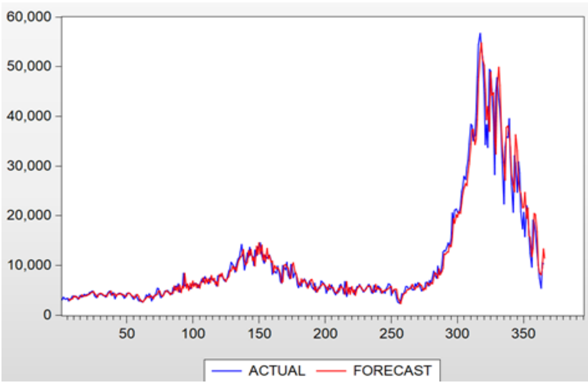
Country	SARIMA	AIC	BIC	RMSE	MAE	MAPE
Indonesia	(3, 1, 4) (2,0,1)7	18.095522	18.223999	2071.460	1197.358	11.33709
Philippines	(2, 1, 2) (2,0,1)7	16.494383	16.590742	902.0344	563.6439	14.12966
Malaysia	(4, 1, 4) (1,0,1)7	15.727864	15.856341	611.4148	406.1535	16.47829

The forecast for the number of new cases in the Philippines a month later will increase with the ARIMA Model estimated will be ≈18,464 while with the SARIMA Model it is estimated will be ≈ 22,818. The forecast for the number of new cases in Malaysia a month later will be slightly the decrease with the ARIMA Model is estimated will be ≈ 22,282 while with the SARIMA Model it is estimated will be ≈19,277.

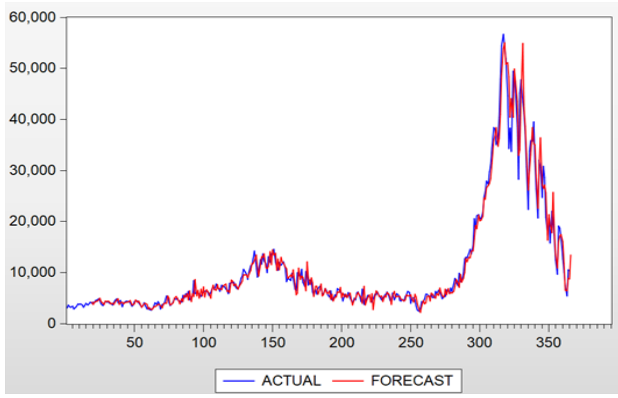
Based on the comparison of the ARIMA and SARIMA prediction Models in Tables 1 and 2, the SARIMA (3, 1, 4)(2,0,1)7 Model is better at predicting the number of new cases in Indonesia with the calculation results of RMSE = 2071.460, MAE = 1197.358, and MAPE = 11.33709. Furthermore, the SARIMA (2, 1, 2)(2,0,1)7 Model is also better at predicting. The Philippines show the results of RMSE = 902.0344, MAE = 563.6439, and MAPE = 14.12966. In Malaysia, the SARIMA (4, 1, 4), (1,0,1)7 Model is also better than the ARIMA Model with the values of RMSE = 611.4148, MAE = 406.1535, and MAPE = 16.47829.

Indonesia

ARIMA (4, 1, 4)

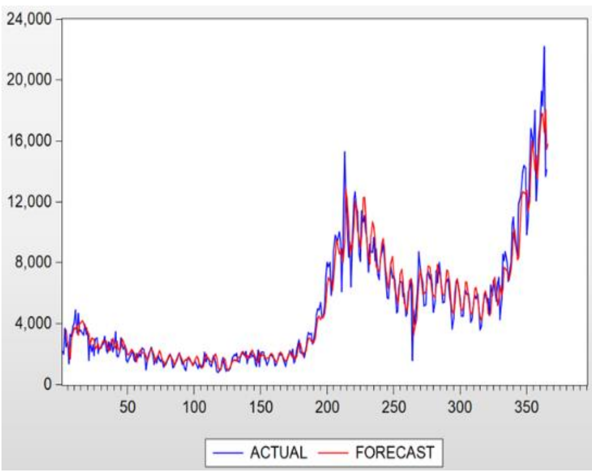


SARIMA (3, 1, 4)(2, 0, 1)7

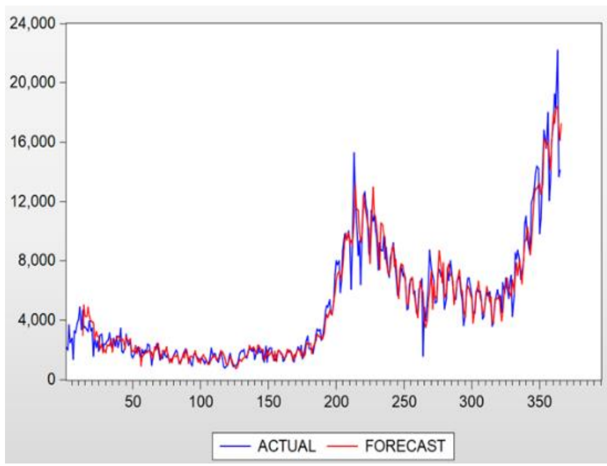


Philippines

ARIMA (2, 1, 3)

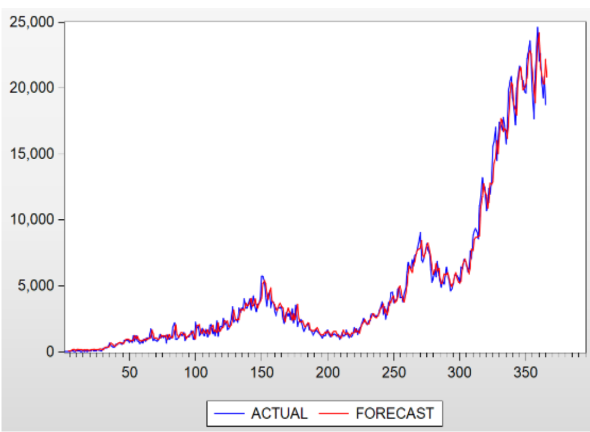


SARIMA (2, 1, 2)(2, 0, 1) 7

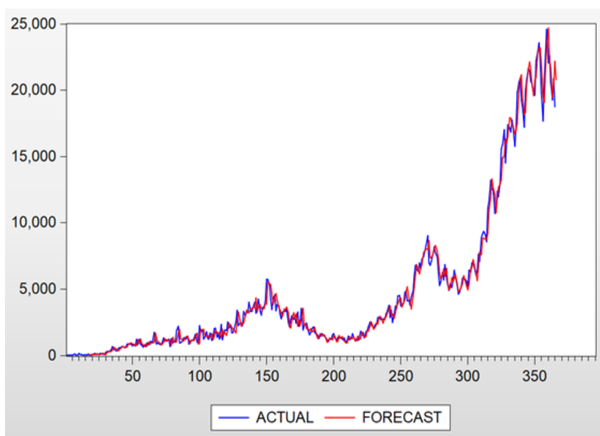


Malaysia

ARIMA (2, 1, 3)

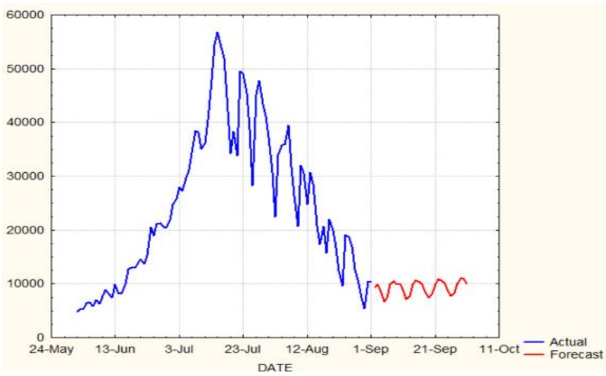


SARIMA (2, 1, 2)(2, 0, 1) 7

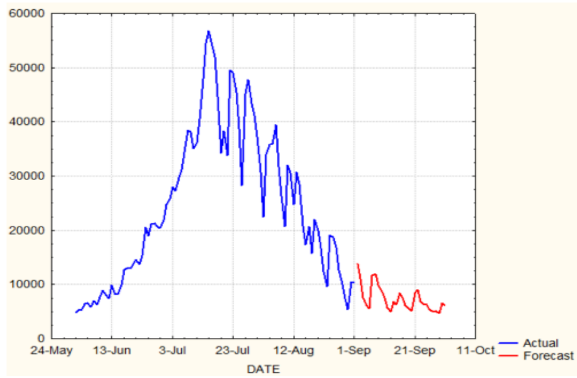


Indonesia

ARIMA (4, 1, 4)

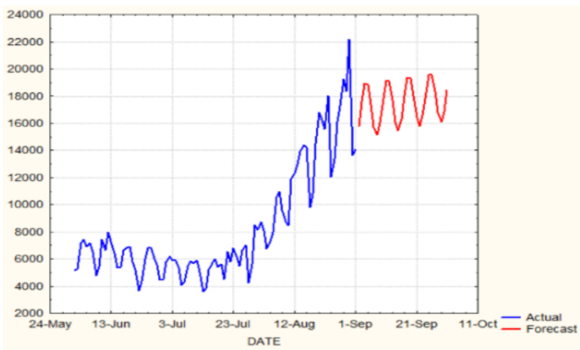


SARIMA (3, 1, 4)(2, 0, 1)7

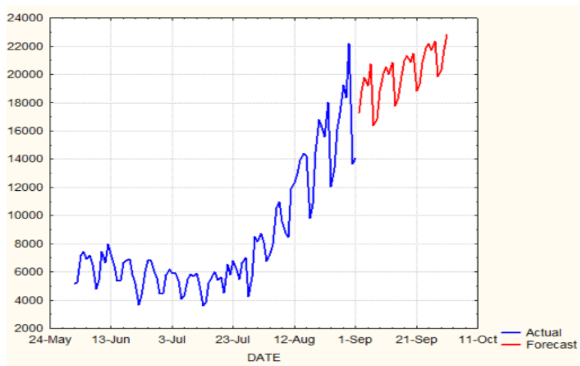


Philippines

ARIMA (2, 1, 3)

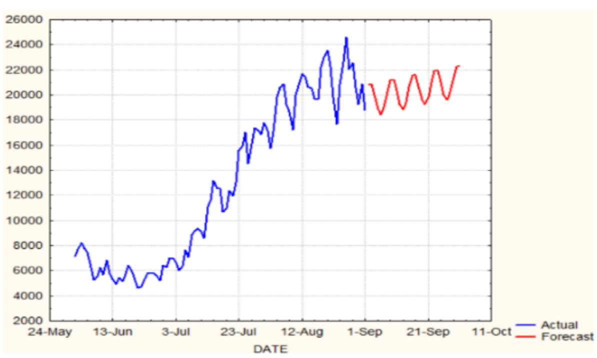


SARIMA (2, 1, 2)(2, 0, 1) 7

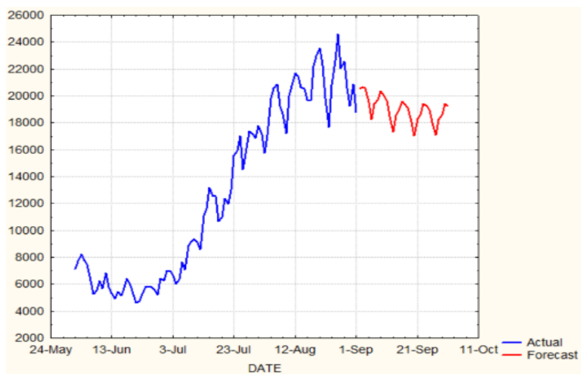


Malaysia

ARIMA (2, 1, 3)



SARIMA (4, 1, 4)(1, 0, 1)7



4. CONCLUSION

In this study, the prediction results of the two models are almost the same, the forecasting trend of the two models is not different. However, based on the calculation of the Criteria Model and the Error Criteria Model, the SARIMA Model is the most suitable for predicting the number of new cases in Indonesia, the Philippines, and Malaysia. This shows that there is a seasonal factor in new cases in the three countries. Based on the forecast results a month later, starting September 2, 2021, there will be a decrease in the number of new cases in Indonesia and the forecast results will fluctuate by around 6,000. This means that the Indonesian gov-ernment has succeeded in reducing the number of new cases. Therefore, they can be reopen or consider lifting the strict pre-vention policies. Meanwhile in the Philippines, based on the forecast results a month later from September 2, 2021, there was an increase and the forecast results fluctuated around 21,000. Furthermore, the forecast for the number of new cases in Ma-laysia and based on the forecast results a month later from September 2, 2021 decrease slightly and the forecast results fluctuated around 19,000.

ACKNOWLEDGEMENTS

The authors would like to thank all those who have helped in the completion of this paper to publish in journal.

AUTHOR'S CONTRIBUTIONS

The authors discussed the results and contributed to from the start to final manuscript.

CONFLICT OF INTEREST

There are no conflicts of interest declared by the authors.

REFERENCES

Ariyo, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Stock price prediction using the ARIMA model. *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, 106–112.

ArunKumar, K. E., Kalaga, D. V., Sai Kumar, C. M., Chilkoor, G., Kawaji, M., & Brenza, T. M. (2021). Forecasting the dynamics of cumulative COVID-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Auto-Regressive Integrated Moving Averag. *Applied Soft Computing*, *103* (February), 107161. https://doi.org/10.1016/j.asoc.2021.107161

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*, *29*, 105340. https://doi.org/10.1016/j.dib.2020.105340

Chen, X., & Yu, B. (2020). First two months of the 2019 Coronavirus Disease (COVID-19) epidemic in China: real-time surveillance and evaluation with a second derivative model. *Global Health Research and Policy*, *5*(1). https://doi.org/10.1186/s41256-020-00137-4

Forecasting of COVID-19 Confirmed Cases in Different Countries with ARIMA Models, medRxiv 1 (2020). https://doi.org/10.1101/2020.03.13.20035345

Divisekara, R. W., Jayasinghe, G., & Kumari, K. (2021). Forecasting the red lentils commodity market price using SARIMA models. *SN Business & Economics*, *1*(1), 1–13.

Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management*, *10*, 1847979018808673.

Gikungu, S. W., Waititu, A. G., & Kihoro, J. M. (2015). Forecasting inflation rate in Kenya using SARIMA model. *American Journal of Theoretical and Applied Statistics*, *4*(1), 15–18.

Guha, B., & Bandyopadhyay, G. (2016). Gold price forecasting using ARIMA model. *Journal of Advanced Management Science*, *4*(2).

Jadhav, V., CHINNAPPA, R. B. V, & Gaddi, G. M. (2017). *Application of ARIMA model for forecasting agricultural prices*.

Ji, L., Zou, Y., He, K., & Zhu, B. (2019). Carbon futures price forecasting based with ARIMA-CNN-LSTM model. *Procedia Computer Science*, *162*, 33–38.

Lan, F. Y., Wei, C. F., Hsu, Y. T., Christiani, D. C., & Kales, S. N. (2020). Work-related COVID-19 transmission in six Asian countries/areas: A follow-up study. *PLoS ONE*, *15*(5), 1–11. https://doi.org/10.1371/journal.pone.0233588

Mehta, P. (2021). The Asian region and COVID-19: approaches and gaps in controlling the virus. *Asian Education and Development Studies*, *10*(2), 185–197. https://doi.org/10.1108/AEDS-05-2020-0090

Mohan, S., Solanki, A. K., Taluja, H. K., & Singh, A. (2022). Predicting the impact of the third wave of COVID-19 in India using hybrid statistical machine learning models: A time series forecasting and sentiment analysis approach. *Computers in Biology and Medicine*, *144*, 105354.

Moreau, V. H. (2020). Forecast predictions for the COVID-19 pandemic in Brazil by statistical modeling using the Weibull distribution for daily new cases and deaths. *Brazilian Journal of Microbiology*, *51*(3), 1109–1115. https://doi.org/10.1007/s42770-020-00331-z

Oenara, J. H., & Oetama, R. S. (2021). Forecasting Of Rupiah Exchange Rate in ASEAN throughout the COVID-19 Pandemic using ARIMA. *2021 6th International Conference on New Media Studies (CONMEDIA)*, 34–39.

Rivero, A. J. L., Alayón, C. A. M., Ferro, R., de la Iglesia, D. H., & Secades, V. A. (2020). Network traffic modeling in a wi-fi system with intelligent soil moisture sensors (Wsn) using iot applications for potato crops using arima and sarima time series. *Applied Sciences (Switzerland)*, *10*(21), 1–19. https://doi.org/10.3390/app10217702

Suryani, A., & Binarto, A. (2021). Forecasting the Number of New Cases of COVID-19 in Indonesia Using the ARIMA and SARIMA Prediction Models. *2nd International Seminar of Science and Applied Technology (ISSAT 2021)*, 63–68.

Yonar, H. (2020). Modeling and Forecasting for the number of cases of the COVID-19 pandemic with the Curve Estimation Models, the Box-Jenkins and Exponential Smoothing Methods. *Eurasian Journal of Medicine and Oncology*, *4*(2), 160–165. https://doi.org/10.14744/ejmo.2020.28273

Zhang, S., Diao, M. Y., Yu, W., Pei, L., Lin, Z., & Chen, D. (2020). Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis. *International Journal of Infectious Diseases*, *93*, 201–204. https://doi.org/10.1016/j.ijid.2020.02.033