PAPER • OPEN ACCESS

A comparative forecasting model of COVID-19 case in Indonesia

A comparative forecasting model of COVID-19 case in Indonesia

T W Septiarini[1*], M R Taufik[2], and T A E Prasetya[3,4]

[1] Department of Islamic Economics, Faculty of Economics and Management, University of Darussalam Gontor, Ponorogo, Indonesia
[2] Department of Occupational Safety and Health, Faculty of Health Science, University of Darussalam Gontor, Ponorogo, Indonesia
[3] Occupational Safety and Health Department, Faculty of Vocational Studies, University of Airlangga, Surabaya, Indonesia
[4] Research Methodology Program, Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Thailand

*Corresponding author :triwijayantiseptiarini@unida.gontor.ac.id

Abstract. COVID-19 had been a disaster in Indonesia. Moreover, it is needed a study to analyze the trend of this case. The objectives of this study were (1) to propose the model for predicting COVID-19 using exponential smoothing, autoregressive integrated moving average (ARIMA), neural network, and fuzzy time series and (2) to compare the performance for each model by using RMSE as evaluation tool. In this study, the splitting data is implemented by 3:1 ratio on train and test data set. The results show that the neural network has the smallest error, 772.46 for RMSE. It means that neural network perform better than other forecasting model. Once, the characteristic data had big impact to building forecasting model whether in classical or modern model.

1. Introduction

COVID-19 is the abbreviation of the coronavirus disease of 2019 [1]. It has been spreading initially in Wuhan, China. Critical acute respiratory syndrome coronavirus 2 (SARS-CoV2) cause COVID-19. WHO stated COVID-19 as global pandemic which reached the confirmation cases 200,000 patients with more than 8000 deaths across over 160 countries on this March [2]. According to Setiati and Azwar [3] in the end of March 2020, Indonesia had been widely impacted by COVID-19 since case fatality rate being 8.9%. As of June 14 2020 in Indonesia, there were 37,420 confirmed cases and 2,091 the deaths. According to Laing [4], COVID-19 as global pandemic has not only affected infections and deaths, but it has also caused chaos in the global economy on the Great Depression level. COVID-19 has the power to abolish personal life, finance, market, industries and whole economies. The survey stated that GDP of Indonesia had been loss on 0.2%. There are 1.9 million of Indonesian people who have lost their jobs due to pandemic [5].

The forecasting COVID-19 model had been conducted by numerous researchers. Logistic model and machine learning is utilized as tool to forecast the trend of COVID-19 pandemic in Brazil, Russia, India, Peru and Indonesia [6]. The Boltzmann function had been implemented to forecast the rate of COVID-19 case in Brazil [7]. The COVID-19 pandemic in Egypt had been predicted by applying

nonlinear autoregressive artificial neural networks [8]. The measurement of death impact of COVID-19 in Jakarta, Indonesia had been analysed by using ARIMA [9]. Yet, it had been faced the study gap to find the comparative study of forecasting COVID-19 case in Indonesia.

In order to fulfill the study gap, this paper will discuss about forecasting model by using four methods. Exponential smoothing and ARIMA are as statistical model. Thus, fuzzy time series and neural network are as artificial model. The forecasting result from statistical and artificial intelligent will be compared by using RMSE as evaluation tools.

2. Methodology

2.1. Data management
The Novel Corona Virus (COVID-19) data is obtained from https://ourworldindata.org/coronavirus-source-data. There are 80 daily dataset from March, 27 2020 until June, 14 2020. Figure 1 capture trend of COVID-19. In this study, the data set will be grouped into two group namely train and test dataset. Based on Septiarini and Musikasuwan [10], the optimal ratio of dataset is 3:1, for training and testing, respectively. Creating stationary data is through differencing process. Stationary data is a primary assumption in classical forecasting model (exponential smoothing, moving average, ARIMA, etc). It has constant statistical properties so can be predicted easily [11]. The ACF plot will rapidly close to zero in stationary. However, the researcher was struggling to reach in stationary level even so there is no stationary exactly. As showed in Figure 2, Indonesia's ACF graph presents the stationary data after differencing process. The ACF depicted several lag appear from the horizontal line. Figure 3 captured time series plot after differencing process.
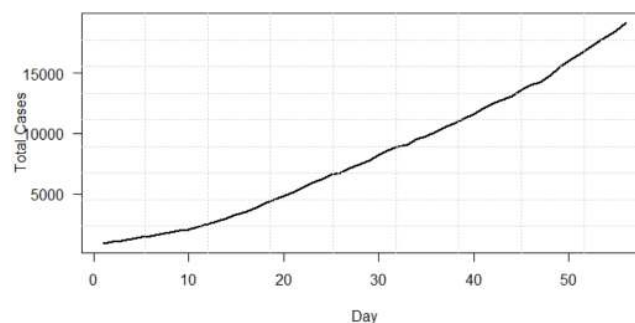

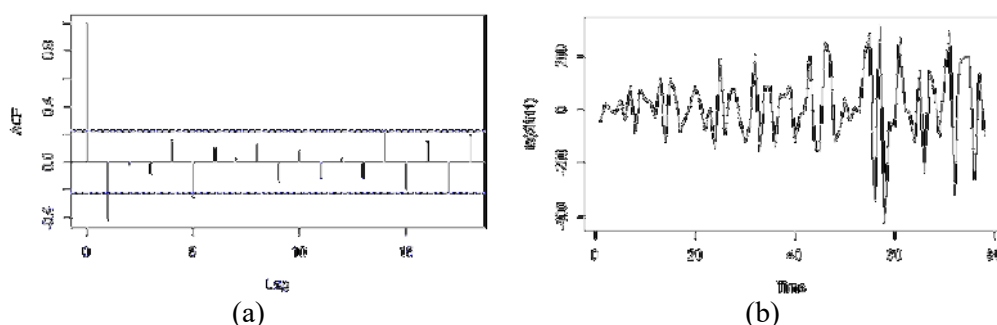**Figure 1.** The plot of time series of Indonesia confirmation case

| (a) | (b) |

Figure 2 (a) The ACF plot after differencing (b) Time series plot after differencing transformation

2.2. Fuzzy Singh's model
The fuzzy time series procedure [12] to forecast COVID-19 confirmation case in Indonesia is explained as follow:
1. The universal set was divided into same length
2. Conduct the fuzzy sets
3. Fuzzification and construct the fuzzy rules

4. Forecasting
5. Diagnostic checking

    After forecasting step, it will be set back whether the residuals correlated by using ACF plot and whether they are normally distributed with mean zero and constant variance by using time plot and histogram of the residuals as shown in Figure 3.
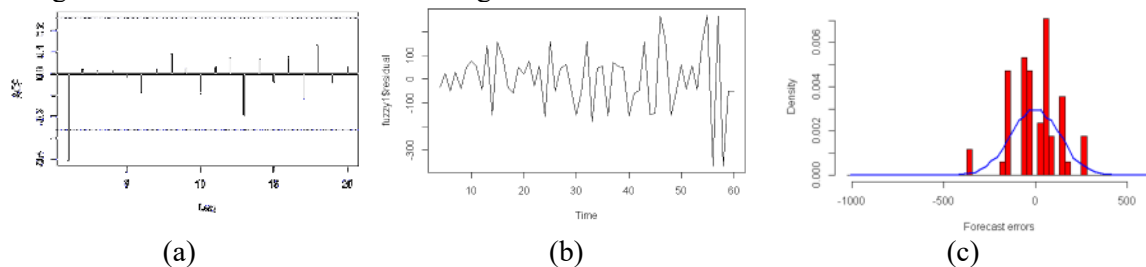

| (a) | (b) | (c) |

Figure 3 (a) ACF plot, (b) Time series plot, (c) Histogram of residuals for fuzzy time series

2.3. Autoregressive Integrated Moving Average (ARIMA)

According to Box and Jenkins [13], there are three elementary steps to propose ARIMA model as follow:

1. Identification

    Observation of the autocorrelation function and partial autocorrelation function plot will be set up in this step.

2. Parameter estimation

    The R program will execute to estimate the parameter for each time series. And the parameter estimation applied ARIMA (0,0,1).

3. Diagnostic checking

    The diagnostic checking deal for 20 forecasted values. As shown on Figure 4, residuals did not appear to be associated, and the residuals satisfy normally distribution with constant variance and zero mean, moreover the constructed ARIMA can be said as a predictive forecasting model.
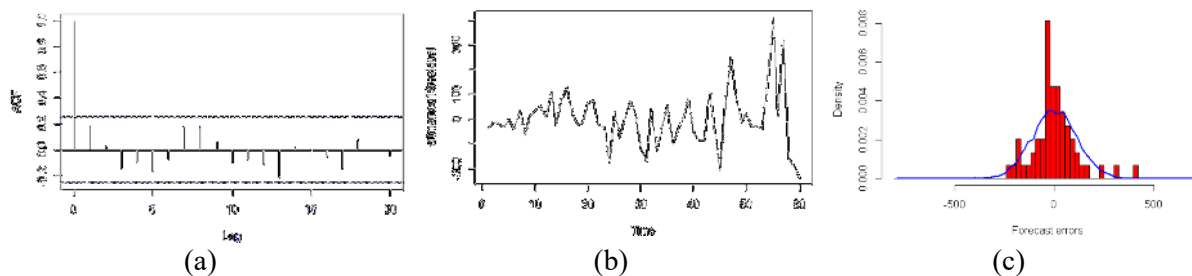

| (a) | (b) | (c) |

Figure 4 (a) ACF plot, (b) time series plot, and (c) histogram of residual for ARIMA

2.4. Exponential smoothing

Simple exponential smoothing is a conventional predicting method which is simple process as its name. Simple exponential smoothing is required to approximate a parameter (α) for training data 75%. The estimating process of simple exponential smoothing model had been executed by using R program which considered the RMSE. In Figure 5, it depicts whether the residuals correlated by using ACF plot and whether they are normally distributed with constant variance and mean zero by using time series graph and the forecast errors.
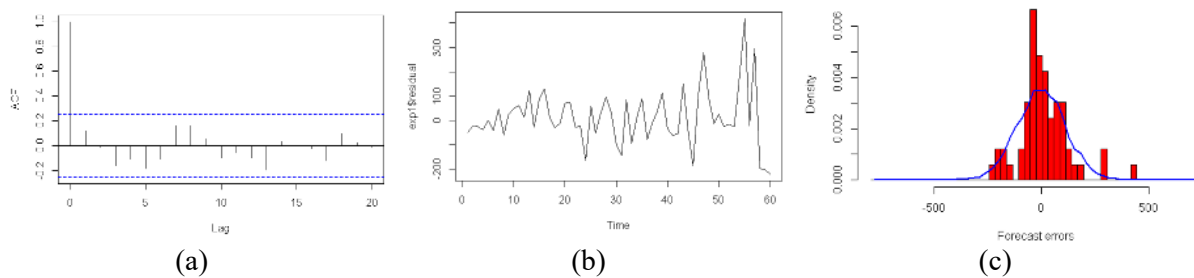
(a)   (b)   (c)

Figure 5 (a) ACF plot, (b) time series plot, and (c) histogram of residual for exponential smoothing

2.5. Neural network
Neural network is kind of machine learning model. There are eight prodcedures in constructing a neural network for predicting aims [14], as follow :
1. Variable selection
2. Data collection
3. Data preprocessing
4. Training, testing, and validation sets
5. Neural network parameter Evaluation criteria
6. Neural network training
7. Implementation

R studio program is applied to execute neural network in this study. In Figure 6, the residuals did not appear to be correlated, and the residuals satisfy normally distributed with constant variance and zero mean, moreover the constructed neural network can be said as a predictive forecasting model.
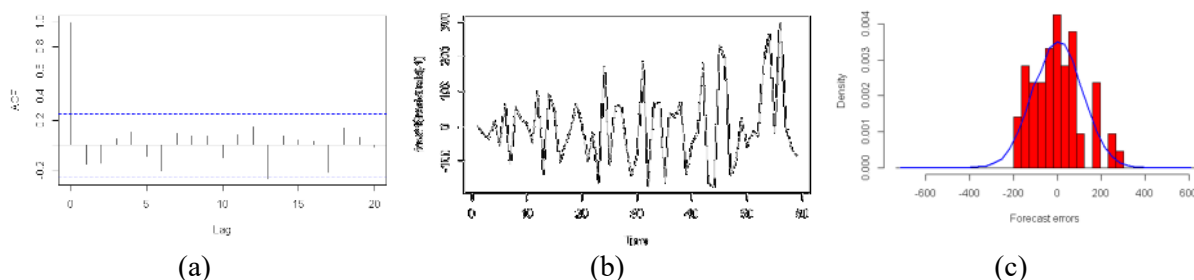

(a)   (b)   (c)

Figure 6 (a) ACF plot, (b) time series plot, and (c) histogram of residual for neural network

3. Result and Discussion

The implementation of four models has been evaluated in this study to forecast COVID-19 confirmation case in Indonesia. The forecasting results are from May, 26 2020 until June, 14 2020. After that the forecasting result will be evaluated with testing data set by using RMSE. Figure 7 captures the results of time series plot. Table 1 depicts the RMSE values for test data set of all models. The smallest RMSE is neural network. It means that neural network has the most powerful model to forecast confirmation case of COVID-19 in Indonesia. According to Musikasuwan and Septiarini [15], for optimizing fuzzy time series model, weighted fuzzy time series can be applied to forecast COVID-19 confirmation case. Again, in order to extend the research about COVID-19, it is needed to add more data since the trend of confirmation case is still going up.

Table 1 The RMSE value for test dataset

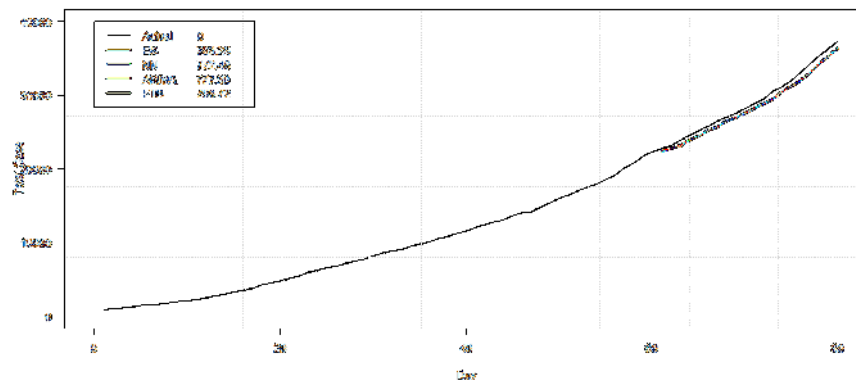|	ES	NN	ARIMA	FTS
RMSE	783.2513	772.4659	777.59	793.721

Figure 7 The plot of forecasting result

4. Conclusions

The objective of this study are to propose the model for predicting COVID-19 applying exponential smoothing, ARIMA, neural network, and fuzzy time series and to compare the performance for each model by using RMSE as an evaluation tool. In this study, the dataset had been classified into two groups which 75% and 25%, train and test dataset, respectively. The result of this study stated that neural network has outperformed to forecast confirmation case of COVID-19 in Indonesia.

References

[1]   Islam R, Bashawir A, Ghani A, Mahyudin E and Manickam N 2017 *Int. J. Econ. Financ.* **7** 355
[2]   Moser G, Rumler F and Scharler J 2007 *Econ. Model.* **24** 470–80
[3]   Jiranyakul K and Opiela TP 2010 *J Asian Econ* **21** 105–12
[4]   Montgomery D C, Jennings C L and Kulahci M 2015 *Introduction to Time Series Analysis and Forecasting* (Canada: John Wiley & Sons. Inc Hoboken New Jersey) p 445
[5]   Grigoryan H 2015 *Database Syst. J.* **6** 14
[6]   Wang P, Zheng X, Li J and Zhu B 2020 *Chaos Solit. Fractals* **139** 110058
[7]   Melo G C D, Araújo N R A D and Araújo K C G M D 2020 *Cad. Saúde Pública* **36** 105720
[8]   Saba A I and Elsheikh A H 2020 *Process Saf. Environ. Prot.* **141** 1
[9]   Fadly F and Sari E 2020 *Unnes J. Public Health* **9**(2) 108
[10] Septiarini T W and Musikasuwan S 2018 *J. Phys.: Conf. Ser.* **1097** 12085
[11] Zhang G P 2003 *Neurocomputing* **50** 159–75
[12] Singh S R 2008 *Math. Comput. Simul.* **79** 539–54
[13] Box G E P and Jenkins G M 1989 *Time Series Analysis: Forecasting and Control* (San Fransisco : CA Holden Day) p 575
[14] Kaastra I and Boyd M 1996 *Neurocomputing* **10** 215-36
[15] Musikasuwan S and Septiarini T W 2020 *Int. J. Innov. Sci. Res. Technol.* **5**(3) 1349-56