Journal of Academia Vol. 10, Issue 2 (2022) 1 – 9

FORECASTING THE COVID-19 MORTALITY RATE WORLDWIDE: A COMPARISON OF UNIVARIATE MODELS

Elya Sara Syuhada Azhar, Noreha Mohamed Yusof*, Nur Salsabila Che Sulaiman, Siti Nuraqina Rosli

Center of Statistical and Decision Science Studies
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA (UiTM), Cawangan Negeri Sembilan, Kampus Seremban,
Persiaran Seremban Tiga/1, Seremban 3, 70300 Seremban,
Negeri Sembilan, Malaysia.

Corresponding author: noreh144@uitm.edu.my

Abstract

Coronavirus disease (COVID-19) is an infectious disease caused by the SARS-CoV-2 virus. COVID-19 disease was initially discovered in Wuhan, China and now spread throughout the countries. Most people infected with the virus will develop mild to moderate respiratory problems and recover without the need for special treatment. However, some people will become severely ill and require medical treatment and can cause death. COVID-19 mortality rates nationwide are increasing day by day and growing concerns. On 13 December 2021, 5,325,079 deaths worldwide were recorded. Thus, this study is regarding the mortality rate of COVID-19 using univariate forecasting techniques. The data was retrieved from GitHub Our World in Data. Holt's method was selected as the best univariate model in order to forecast the mortality rate. Holt's method shows the lowest error measures. The predicted value of the mortality rate for COVID-19 is decreasing between 1 November 2021 to 31 January 2022. The decreasing predicted value might be due to the vaccinated programs done worldwide. A further study should be done to measured the factors related to the improved spread of COVID-19.

Keywords: forecasting, COVID-19, mortality rate, mortality rate worldwide, univariate model

Article History:- Received: 21 July 2022; Revised: 6 September 2022; Accepted: 7 September 2022; Published: 31 October 2022

Introduction

This study is regarding the mortality rate of COVID-19 worldwide. What is COVID-19? COVID-19 is a new virus that is called Coronavirus disease 2019. COVID-19 also known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It was officially named by the International Committee on Taxonomy of Viruses based on phylogenetic analysis (Liu et al., 2020). According to Liu et al. (2020), COVID-19 is the fifth pandemic after the 1918 flu pandemic that could affect all people around the world. COVID-19 or formerly called as 2019-nCoV was first reported and subsequent outbreak of respiratory illness caused in Wuhan City, Hubei Province, China since late December 2019. On **1** December 2019, it was the earliest date of symptom onset (Liu et al., 2020). According to Liang et al. (2020) in Scientific Reports, more than 8.7 million people have been infected and more than 460 thousand have died worldwide, as of 20 June 2020.

Figure 1 shows the total deaths for all country until 25 October 2021 that was retrieved from World Health Organization (2020) Coronavirus (COVID-19) Dashboard. On 25 October 2021, the total deaths worldwide reported is 35,945 deaths compared to 18 October 2021 which is 49,558 deaths. While, the highest deaths reported on COVID-19 is on 18 January 2021 which is 102,780 deaths since the pandemic outbreak.

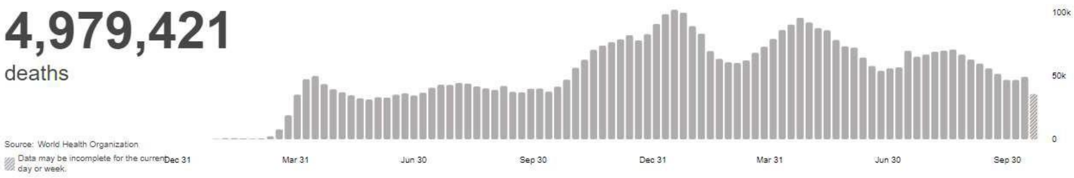

Figure 1. Worldwide COVID-19 Total Deaths as of 25 October 2021

Based on the previous COVID-19 deaths reported for all country, forecasting has become an essential for effective governmental decision making especially to decrease the number of mortality due to this pandemic since forecasting can identify the behaviour patterns and trends to predict the future scenarios.

Therefore, forecasting models become very useful in order to predict the future of COVID-19 daily deaths specially to support health systems with new strategic decision making, planning, and health policy formulation that help in the fight against COVID-19. Data from GitHub Our World in Data was used to forecast the COVID-19 mortality rate for daily three months ahead.

Literature Review

There are various publications on the use of forecasting models to predict the future of COVID-19 pandemics. The exponential smoothing technique is widely used for forecasting and it is one of the popular forecasting methods for short term periods. According to Khan et al. (2021), Bayesian Dynamic Linear Model (BDLM) was used to forecast the daily new infections, deaths and recovery cases regarding COVID-19. They found the maximum number of new infections is 4,031 per day with 95% prediction interval (3,319-4,743). Death forecast shows that the maximum number of deaths with 95% prediction intervals are 81 and (67-93), respectively.

Meanwhile, Al-Turaiki et al. (2021) used time-series forecasting on confirmed and recovered COVID-19 cases and deaths for Saudi Arabia at the country level and confirmed cases for Riyadh at the city level. Based on this research, the researchers found in the last period which is from 9 March 2021 until 5 April 2021, the Exponential Smoothing (ES), Holt's, Holt-Winters and Autoregressive Integrated Moving Average (ARIMA) models achieved better performance than other models. They included that Holt's, Holt-Winters and ARIMA achieved good performance compared to other forecasting models. Furthermore, the researcher said that the Single Exponential Smoothing (SES) model yields the worst accuracy for predicting deaths in Saudi Arabia. It is because a good model to use on any time series data will depend on the factors such as simplicity, accuracy and stability (De Gooijer & Hyndman, 2006). The accuracy of these forecasts is important since the more accurate the forecast, the more confidence users have in the forecast and the forecasting process (Klimberg et al. 2005).

Additionally, a research by Harini (2020) used Double Exponential Smoothing (DES) method to forecast the COVID-19 cases in Indonesia where the researcher explained that DES is one of the methods that can be used to optimize the estimation of the ARIMA model with smoothing parameters α. Data for this research were taken from the tabulation of the Indonesia National Disaster Management Agency BNPB) from 2 March until 7 April 2020. The finding shows that the pattern of COVID-19 data distribution behavior in Indonesia shows an exponential distribution pattern where the addition of positive cases of COVID-19 increases significantly from every day and it's also followed by a pattern of distribution of the number of people who recovered and died. The researcher analyzed and confirmed that the DES is the best to predict the future COVID-19 cases in Indonesia.

Therefore, in this research, the researchers decided to use a univariate modelling techniques that involve Naïve model, and ES such as SES, DES and Holt’s method to forecast the worldwide mortality rate of COVID-19.

Methods

Study Area

The data was retrieved from GitHub Our World in Data. The data was chosen from 1 July 2021 until 31 October 2021. It was used to forecast the mortality rate of COVID-19 worldwide for daily three months ahead starting 1 November 2021 until 31 January 2022. Four methods of Univariate Techniques are used to examine the mortality rate of COVID-19 worldwide. The models used are Naïve model, Single Exponential Smoothing, Double Exponential Smoothing and Holt's method.

Naïve Model

Naive forecasting models are based exclusively on historical observation of sales or other variables. Naive model strongly believes that what happens today will happen again tomorrow or any other time in the future. It performs well in series which exhibits a slow change in the fluctuations. Sudden change in the current data would severely affect the accuracy of the forecast values. In this study, the forecast number of new deaths of COVID-19 from 24 January 2020 until 31 October 2021 were used. Let the number of new deaths, *Yi* and the corresponding values at time t=l,2,3,4,5,..., t to be $Y_1, Y_2, Y_3, Y_4, ..., Y_t$

Hence, mathematically the model can be represented as

$$F_{t+m} = y_t \tag{1}$$

Single Exponential Smoothing

Single exponential Smoothing (SES) is the simplest method in the exponential smoothing technique. SES is also called Simple Exponential Smoothing. This method is suitable for short-term forecasting where commonly used for a one month period (Nazim & Afthanorhan, 2014). The method assumes the data rise and fall around the statistical mean value and has no clear trend or seasonal growing pattern in the historical data. The equation for Simple Exponential Smoothing method is

$$F_{t+1} = \alpha Y_t + (1-\alpha)F_t \tag{2}$$

where

F_{t+1} = the Single Exponentially Smoothed value in time oft+1

α = constant value between O and 1, alpha

Y_t = the actual value of period *t*

F_t = the forecast value of period *t*

Double Exponential Smoothing

This method is used when the data shows a linear trend characteristic. Double Exponential Smoothing uses two parameters, to update the components at each period which is level and trend (Nazim & Afthanorhan, 2014). The level is a smoothed estimate value at the end of each period (Sidqi & Sumitra, 2019) . There are four main formulas use in Double Exponential Smoothing method. The first two formula are

$$S_t = \alpha Y_t + (1-\alpha)S_{t-1} \tag{3}$$

$$S_t^{'} = \alpha S_t + (1-\alpha)S_{t-1}^{'} \tag{4}$$

where

S_t =the exponentially smoothed value of Y_t at time *t*

$S_t^{'}$ =the double exponentially smoothed value of Y_t at time t

The next formula is the parameter formula that use to calculate the final forecast value.

$$a_t = 2S_t - S_t^{'} \quad (5)$$

$$b_t = \left(\frac{\alpha}{1-\alpha}\right)\left(S_t - S_t^{'}\right) \quad (6)$$

$$F_{T+m} = a_T + b_T xm \quad (7)$$

where

F_{T+m} = the forecast value at time t form=1,2,3,...

Holt's Method

There are three separated equations in Holt's model that work together to generate a final forecast. The first equation is a basic smoothing equation that directly adjusts the last smoothed value for last period's trend where it can be calculated as follows:

$$S_t = \alpha Y_t + (1-\alpha)(S_{t-1} + T_{t-1}) \quad (8)$$

In this method, the trend itself is updated over time through the second equation and can be calculated as follows:

$$T_t = \beta(S_t - S_{t-1}) + (1-\beta)T_{t-1} \quad (9)$$

Where the trend is expressed as the difference between the last two smoothed values. Lastly, the third equation is used to produce the final forecast as follows:

$$F_{t+m} = S_t + T_t xm \quad (10)$$

Measurement Errors

There are several types of forecasting performance measures that are used to evaluate the amount of the error. One of the popular forecasting performance measures of the size of the error is the mean square error (MSE). The value of MSE can be calculated by using the formula of the sum of the squares of the forecast errors as follows:

$$\text{MSE} = \frac{\sum_{t}^{n} e_t^2}{n} \quad (11)$$

which $e_t = Y_t - \hat{Y}_t \quad (12)$

Where the value of MSE can measure the amount of dispersion of the errors. For MSE, the model becomes the better and more accurate forecasting model if the other than that, the mean absolute percentage error (MAPE) is one of the forecasting performances measures that is widely used to evaluate the forecasting model which does attempt to consider the effect of the magnitude of the actual values. The value of MAPE can be calculated as the average of the absolute values of percentage errors

as follows:

$$\text{MAPE} = \frac{\sum_{m=1}^{n} \left| \left(\frac{e_{t+m}}{y_{t+m}} \right) * 100 \right|}{n} \quad (13)$$

$$\text{which } e_t = Y_t - \hat{Y}_t \quad (14)$$

The same goes for MSE, if the model has the lowest value of MAPE, then the forecasting model becomes more accurate.

Result and Discussion

The results of analysis are shown in this part regarding the time series model and the forecast values. Figure 2 shows the trend of COVID-19 worldwide daily mortality rate reported from 1 July 2021 to 31 October 2021. From this figure, it shows that there is fluctuated trend in the COVID-19 mortality rate worldwide since this pandemic occurred. Based on the linear trend line it indicates that the mortality rate is decreasing from 0.00012% in 1 July 2021 to 6.673x10-^{5}% in 31 October 2021. The highest mortality rate is on 1 September 2021 with a value of 0.000167% while the lowest value is in 3 October 2021 with 6.071xl0-^{5}%.

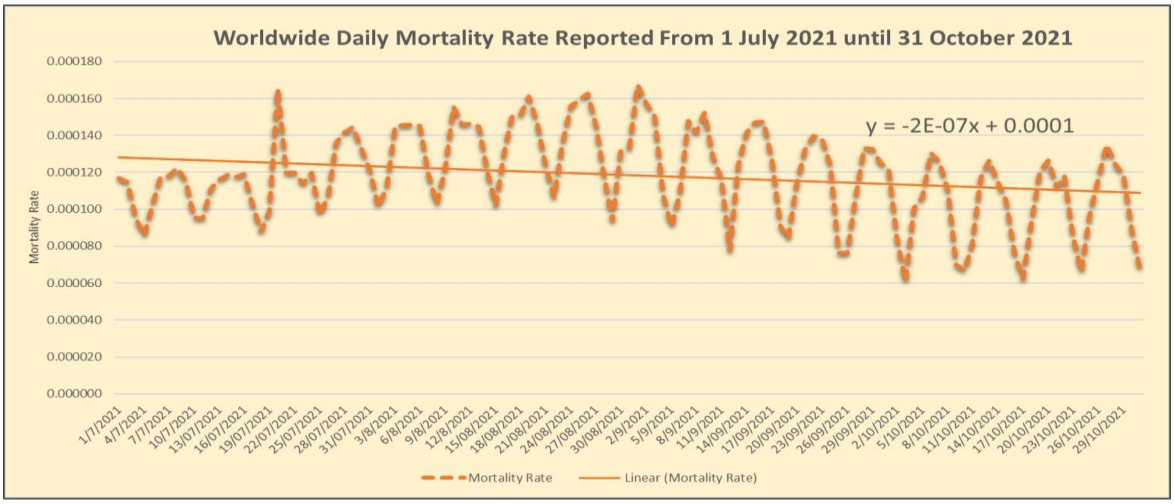

Figure 2. The Worldwide Daily Mortality Rate

Forecasting Analysis

The analysis model using the four univariate techniques are discussed in this part. The discussion is followed:

The fitted and actual values of Naive Model starting from **1** July 2021 until 31 October 2021 using Naive Model is illustrated in Figure 3. The MSE estimation is 4.4011x10-^{10} and MSE evaluation is 4.9751x10^{-10}. The value of MAPE estimation and evaluation part for Naive Model is 13.5373 and 21.0455 respectively.

Journal of Academia Vol. 10, Issue 2 (2022) 1 – 9

- Naïve Model

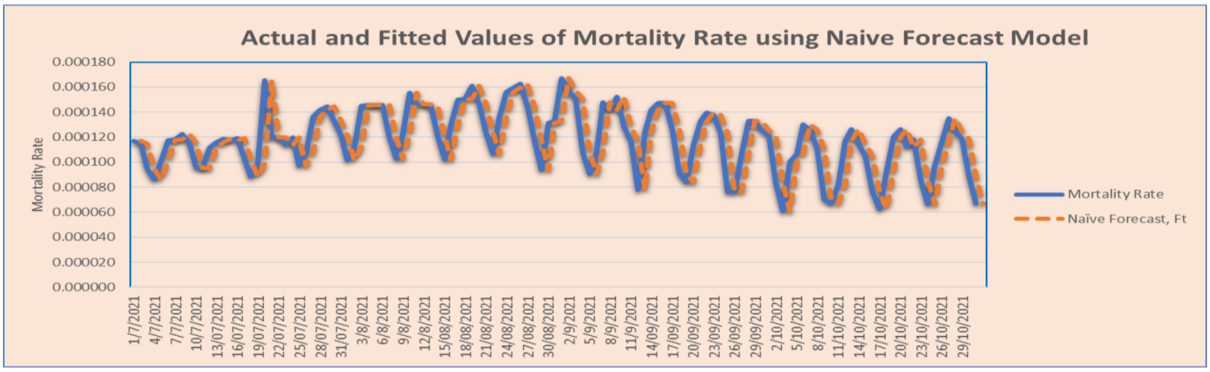

Figure 3. The Actual and Fitted Values of Mortality Rate using Naïve Model

Figure 4 shows the fitted and actual value of COVID-19 mortality rate using Single Exponential Smoothing. The two types of the lines show a similar pattern on graft. The MSE estimation is 4.6159x10-^{10}, and MSE evaluation is 5.6701x10-^{10}. Meanwhile, the value of MAPE estimation is 14.4619 and MAPE evaluation is 23.2613.

- Single Exponential Smoothing

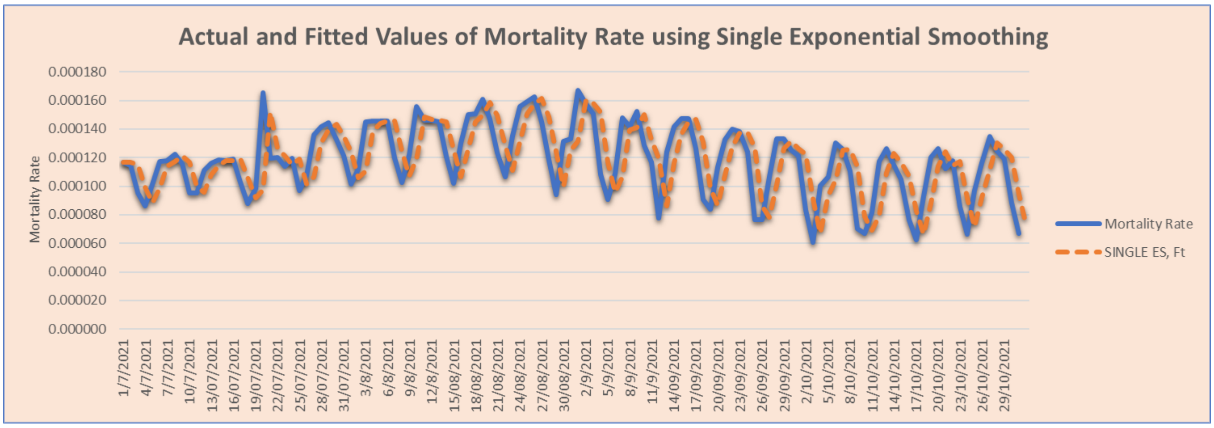

Figure 4. The Actual and Fitted Values of Mortality Rate using Single Exponential Smoothing

The fitted and actual value of COVID-19 mortality rates using Double Exponential Smoothing is shows in Figure 5. The MSE estimation 6.5529x10-^{10} and MSE evaluation is 7.0288x10-^{10}, where the value of MSE evaluation for double has the highest value from other method for MSE evaluation. While, the value of MAPE estimation and evaluation is 16.6381 and 22.7956.

- Double Exponential Smoothing

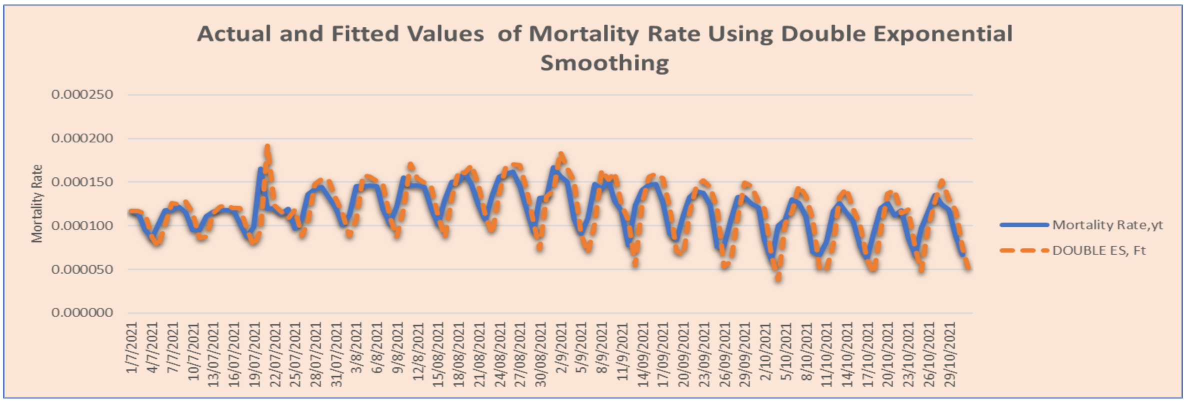

Figure 5. The Actual and Fitted Values of Mortality Rate using Double Exponential Smoothing

The fitted and actual values of COVID-19 Mortality Rate using Holt's method in Figure 6 shows the similarity of the trend from the previous method. The MSE estimation is 4.7922×10^{-10} and MSE evaluation is 5.4091×10^{-10} , while the value of MAPE estimation is 14.1853 and MAPE evaluation is 21.4672.

- Holt’s Method

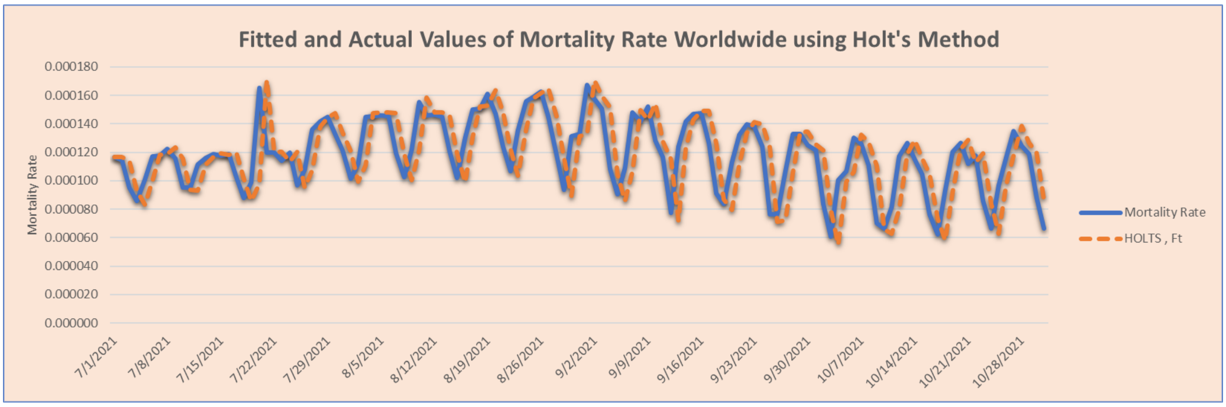

Figure 6. The Actual and Fitted Values of Mortality Rate using Holt’s Method

Error Measures

Table 1 shows the analysis for four different methods of forecasting. This table gives the summarization of error measures for four different methods where Naive Model represent as the benchmark model. The result shows MSE estimation for Naive Model is slightly lower than using Holt's method, Single Exponential Smoothing and Double Exponential Smoothing and also vice versa for MSE evaluation. The result also shows that Holt's method is the best model for forecasting purposes since Holt's method has the smallest value of MSE and MAPE evaluation compared to Single and Double Exponential Smoothing which are 5.4091×10^{-10} and 21.4672 respectively. Hence, this method is used to forecast the daily three months ahead which is from 1 November 2021 until 31 January 2022.

Table 1. Error Measures for Different Methods

Method/ Error Measures	MSE		MAPE	
	Estimation	Evaluation	Estimation	Evaluation
Naive Model	4.4011×10^{-10}	4.9751×10^{-10}	13.5373	21.0455
Single Exponential Smoothing	4.6159×10^{-10}	5.6701×10^{-10}	14.4619	23.2613
Double Exponential Smoothing	6.5529×10^{-10}	7.0288×10^{-10}	16.6381	22.7956
Holt's Method	4.7922×10^{-10}	5.4091×10^{-10}	14.1853	21.4672

Forecasting Values from 1 November 2021 until 31 January 2022

Since the Holt's method is the best model to determine the mortality rate, the forecasting values for mortality rate have been generated. Figure 7 shows the plot of all the forecasting values.

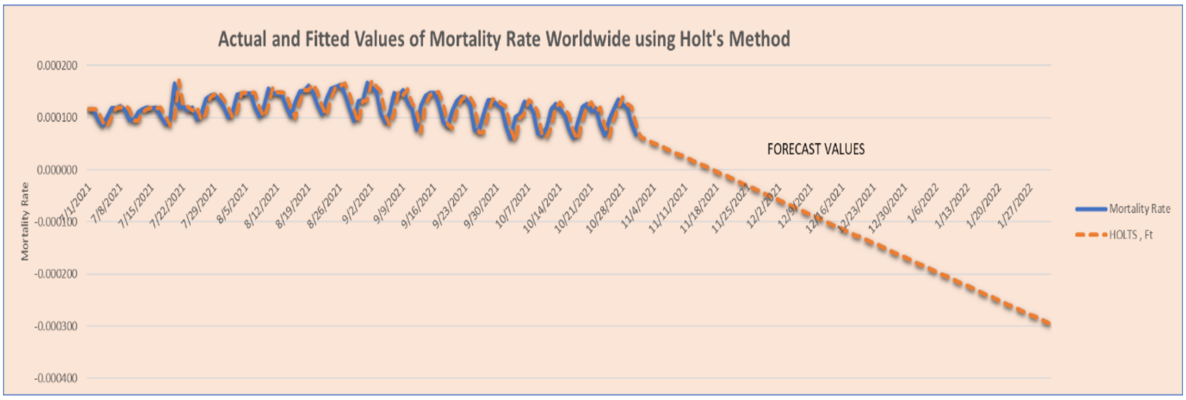

Figure 7. Forecasting values for three months ahead using the best model

This plot of estimated forecast values from 1 November 2021 until 31 January 2022 shows a downward trend pattern of mortality rate. The predicted value of the mortality rate for COVID-19 is decreasing from 0.00063% to -0.000294%. The improved predicted values of the mortality rate of COVID-19 might be because of the vaccinations programs worldwide. In fact, this is one of the indicators of the awareness among people on the importance of COVID-19's vaccine.

Conclusion

Based on the forecasting on the COVID-19 mortality rate, the results show that Holt's method is the best model to be used for forecasting since Holt's method has the lowest value of MSE and MAPE compared to other techniques of forecasting. The findings are supported by Al-Turaiki et al. (2021), where the researchers conclude that Holt's, Holt-Winters and ARIMA achieved good performance compared to other forecasting models. The forecast values indicate that nearly 0% mortality rate of worldwide will die due to the COVID- 19 for daily worldwide population. This is a good indicator of the effectiveness of the vaccination programmes and standard operation procedure of handling COVID-19 worldwide. The forecasting model in predicting the COVID-19 mortality rate enable to support the government especially health systems. With new strategic decision making, planning, and health policy formulation, it will help in the fight against COVID-19. This will enable us to live with our normal life again rather than just stay at home.

Acknowledgement

The authors would like to express their gratitude to the staffs of Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Negeri Sembilan Branch, Seremban Campus, Negeri Sembilan, Malaysia for their supportive effort towards researches and publications.

Author Contribution

Every author is working as a team to come out with the analysis and report writing.

Conflict of Interest

No conflict of interest towards anyone or any organizations.

References

Al-Turaiki, I., Almutlaq, F., Alrasheed, H., & Alballa, N. (2021). Empirical evaluation of alternative time-series models for COVID-19 forecasting in Saudi Arabia. *International Journal of Environmental Research and Public Health*, *18*(16), 8660.

De Gooijer, J. & Hyndman, R. (2006). 25 Years of Time Series Forecasting. *International Journal of Forecasting*, *22*, 443-473.

GitHub, Inc. (2021). Data on excess mortality during the COVID-19 pandemic by Our World in Data. Retrieved from covid-19-data/public/data/excess_mortality at master · owid/covid-19-data · GitHub [Access online 14 September 2021].

Harini, S. (2020). Identification COVID-19 Cases in Indonesia with The Double Exponential Smoothing Method. *Jurnal Matematika MANTIK, 6*(1), 66-75.

Khan F., Ali S., Saeed A., Kumar R., & Khan A.W. (2021). Forecasting daily new infections, deaths and recovery cases due to COVID-19 in Pakistan by using Bayesian Dynamic Linear Models. *PLOS ONE 16*(6): e0253367.

Klimberg, R. K., Lawrence, K., & Lawrence, S. (2005). Forecasting sales of comparable units with data envelopment analysis (dea). *Advances in Business and Management Forecasting, 4,* 201-214.

Liang, L. L., Tseng, C.H., Ho, H.J., & Wu, C.Y. (2020). Covid-19 mortality is negatively associated with test number and government effectiveness. *Scientific Reports, 10*(1), 1-7.

Liu, Y.C., Kuo, R.L., & Shih, S.R. (2020). COVID-19: The first documented coronavirus pandemic in history. *Biomedical Journal*, *43*, 328-333.

Nazim, A. & Afthanorhan, A. (2014). A Comparison between Single Exponential Smoothing (SES), Double Exponential Smoothing (DES), Holts (Brown) and Adaptive Response Rate Exponential Smoothing (ARRES) Techniques in Forecasting Malaysia Population. *Global Journal of Mathematical Analysis*, *2,* 276-280.

Sidqi, F., & Sumitra, I. (2019). Forecasting product selling using single exponential smoothing and double exponential smoothing methods. In *Iop conference series: Materials Science and Engineering* (*Vol. 662*, p. 032031).

World Health Organization (2020). WHO Coronavirus (COVID-19) Dashboard. Retrieved from WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data [Access online 15 September 2021].