SEMARAK ILMU PUBLISHING

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage: https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

# Modelling and Forecasting the COVID-19 Mortality Rates in Malaysia by using ARIMA Model

Siti Rohani Mohd Nor[1,*], Nurul Syuhada Samsudin[1], Muhammad Asri Manap[1], Siti Mariam Norrulashikin[1]

[1] Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 Johobr Bahru, Johor, Malaysia

**ARTICLE INFO**

***Article history:***
Received 1 September 2023
Received in revised form 1 March 2024
Accepted 11 April 2024
Available online 12 May 2024

***Keywords:***
COVID-19; Mortality Rates; Forecast; ARIMA

**ABSTRACT**

Over the last year, the COVID-19 epidemic has afflicted over 150 million individuals and killed over three million people globally. Various forecasting models attempted to estimate the temporal course of the COVID-19 pandemic during this time period in order to determine effectiveness of the government action in facing COVID-19 outbreak. In this study, Autoregressive Integrated Moving Average (ARIMA) models were used in order to forecast the COVID-19 mortality rates data in Malaysia. The accuracy of the ARIMA models is then evaluated by using Mean Absolute Error (MAE) and Root Mean Square Absolute Error (RMSE). The forecasting model with the lowest error is picked as the best. In this study, ARIMA (1,1,3) outperformed the ARIMA (1,1,2) and ARIMA (1,1,4) models since it has the lowest MAE and RMSE values. However, as compared to ARIMA (1,1,4), the study found that ARIMA (1,1,3) model is not adequate in terms of model fitting due to the errors were not normally distributed. Hence, ARIMA (1,1,4) model was chosen to make prediction of COVID-19 mortality rates. Accordingly, the findings through this study can be used as a preliminary study to predict the COVID-19 mortality rates and other future pandemic cases to mitigate risk of increasing cases.

## 1. Introduction

The World Health Organization confirmed Coronavirus disease 2019 (COVID-19), a novel pneumonia disease originating in Wuhan, on January 12, 2020, before it became an epidemic in all countries [1]. The first case outside of China was reported in Thailand, prompting a stringent screening process at all Malaysian airports [2]. Until April 14, 2020, Malaysia reported two waves of COVID-19 cases, with the first wave ending successfully within two months. The second wave began early in March, with worrying circumstances. Following the Sabah state election in September 2020, there were a number of outbreaks at Top Glove facilities in late 2020 that led to a third wave of COVID-19 infections in the country.

A severe economic impact has been caused by the pandemic, which has devalued the currency in Malaysia and shrank its GDP [2]. The pandemic also had widespread effects on Malaysian society.

* *Corresponding author.*
*E-mail address: sitirohani@utm.my*

https://doi.org/10.37934/araset.45.1.215223

In the early months of 2020, the pandemic was associated with an unrelated political crisis, hampering the government and society's early response. Repeated COVID-19 outbreaks and emergency measures exacerbated political instability throughout 2020 and 2021.

The Malaysian government began enforcing a Movement Control Order on March 18, 2020, to break the chain of COVID-19 [3]. A hashtag #stayhome was prominently featured in the media. Non-governmental organizations and prisoners have produced personal protective equipment for frontline workers. A variety of organizations have organized events to raise money to provide necessities primarily to hospitals. The Ministry of Health was granted collaborations with healthcare providers, as well as more laboratories to enhance its capabilities. The government had to do several things to prevent COVID-19 from spreading. However, the problem is how to determine whether or not those actions are effective. Therefore, there must be a way we can use to predict whether the government action in decreasing the COVID-19 mortality rates is effective or not.

Forecasting is the process of using historical data to create informed estimates for predicting future trends based on the direction of those trends. In this study, time series analysis will be used to forecast COVID-19 mortality rates in Malaysia. Time series analysis is a particular method of analysing a sequence of data points collected over time. As opposed to recording data points intermittently or randomly over a set period, analysts record data points at consistent intervals in time series analysis. Auto regressive integrated moving average (ARIMA) model is one of the most time series models that has been conducted on past pandemic infections and diseases, such as to predict the incidence of Hepatitis A Virus (HAV) [4]. Other than that, ARIMA model was applied to the number of confirmed cases per day for severe acute respiratory syndrome (SARS) [5]. ARIMA was also used to predict the incidence of Hemorrhagic Fever with Renal Syndrome (HFRS) in China based on data from 2009 to 2011 [6]. The ARIMA model was used to predict the Hantavirus Pulmonary Syndrome (HPS) [7].

Furthermore, several studies have also been conducted to estimate the spread of COVID-19 by using the ARIMA model. ARIMA and Richard's model were used to forecast population impacts of COVID-19 in India [8]. Denmark, Norway, and Sweden could predict the number of patients with COVID-19 in a short period of time based on simple time series [9]. In Italy, Spain, and France, the number of COVID-19 cases were predicted based on the ARIMA model [10]. The number of COVID-19 confirmed cases in Saudi Arabia for the next four weeks was also forecasted by using ARIMA model [11]. In European countries, ARIMA was used to forecast the rate of infection over the next seven days [12]. Since ARIMA model has been used by all previous cases involving disease and COVID-19, this study aims to predict the COVID-19 mortality rates by using ARIMA models.

## 2. Methodology

### *2.1 ARIMA Model*

COVID-19 data time-series are modelled by Box-Jenkins type stochastic processes, which is by using ARIMA model. ARIMA model is a time series model that is used to forecast the future time series of data. For ARIMA, the letter "I" is used to differentiate the series when they are not stationary. In the meantime, stationary time series may be modelled using any of three class processes, namely autoregressive (AR), moving-average (MA), and mixed autoregressive and moving-average (ARMA).

An autoregressive model of order p which indicate as AR ($p$) is expressed as:

$$y_t = \mu + \varphi_1 \gamma_{t-1} + \varphi_2 \gamma_{t-2} + \cdots + \varphi_p \gamma_{t-p} + u_t \quad (1)$$

The Moving Average of order q which denoted as MA $(q)$ is expressed as:

$$y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q} \quad (2)$$

where $u_t$ is a white noise disturbance term with:

$E(u_t) = 0$ and $Var(u_t) = \sigma^2$

Assumed that the error term, $\varepsilon_t$ is uncorrelated with anything, with zero mean and constant variance,$\sigma^2$. The process is called White Noise (WN) process with denotation:

$$\varepsilon_{t \sim WN(0,\sigma^2{}_t)} \quad (3)$$

ARIMA $(p, q)$ model is formed by combining AR $(p)$ and MA $(q)$ models as shown below:

$$y_t = \mu + \varphi_1 \gamma_{t-1} + \varphi_2 \gamma_{t-2} + \cdots + \varphi_p \gamma_{t-p} + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q} \quad (4)$$

or in sigma notation:

$$y_t = c + \sum_{i=1}^{p} \varphi_i \gamma_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (5)$$

where $y_t$ is the COVID-19, $C$ is constant term, $\varphi_t$ is the parameter of AR of order $p$, $\theta_j$ is the parameter of MA of order $q$ and $t$ is the error at time $t$. The order of $p$ and $q$ must be positive integers.

ARIMA involved 5 steps modelling. First, to begin processing ARIMA$(p, d, q)$, it is necessary to determine if the time series is stationary or non- stationary. Second, a stationary time series is required for ARIMA to work. Third, time series that are not stationary need to undergo appropriate degrees of differencing until they reach stationary before they can be used for ARIMA. Next, taking the autocorrelation function (ACF) and partial autocorrelation function (PACF) of a time series as input. Lastly, autoregressive order $p$ and moving average $q$ can be determined.

### *2.2 Performance of ARIMA Models*

This study computes the difference between the forecasted value and out-of-sample data, then calculates Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to measure the accuracy of ARIMA models in forecasting.

The Mean Absolute Error (MAE) measures the average size of errors without considering their direction. The formula is as follows:

$$MAE = \frac{\sum_{t=1}^{n} (y_t - \hat{y}_t)}{n} \quad (6)$$

Root Mean Square Error (RMSE) is the error measurement of finding out the differences between the forecasted value and observations. The formula is given as:

$$RMSE = \sqrt{\frac{(y_t - \hat{y}_t)^2}{n}} \quad (7)$$

where $y_t$ is the original value according to time t, $\widehat{y_t}$ is the prediction value according to time *t* and *n* is the number of forecasts.

As the accuracy of the forecasting model is measured by these two errors, the model that has the lowest number of MAE, and RMSE is the best model.

## 3. Results

### *3.1 Data Descriptive*

The COVID-19 mortality rates data of Malaysia were obtained from Ministry of Health Malaysia's website. The data cover from 17th March 2020 until 31st December 2022 with a total of 1020 observations. Afterwards, the data are separated into two subsets, one is in-sample data and the other out-sample data. In-sample data are range from 17th March 2020 until 11st November 2022 with total of 970 observations and it will be used to forecast the next 50 observations. The trend of COVID-19 mortality rates is recorded in Figure 1. Figure 1 illustrates that the COVID-19 mortality rates of Malaysia were at the spike around the year 2021, and the trends are decreasing as the year increases to 2022. Next section will discuss whether the trends of mortality rates of COVID-19 will continue to increase or not based on the past historical data with ARIMA model.

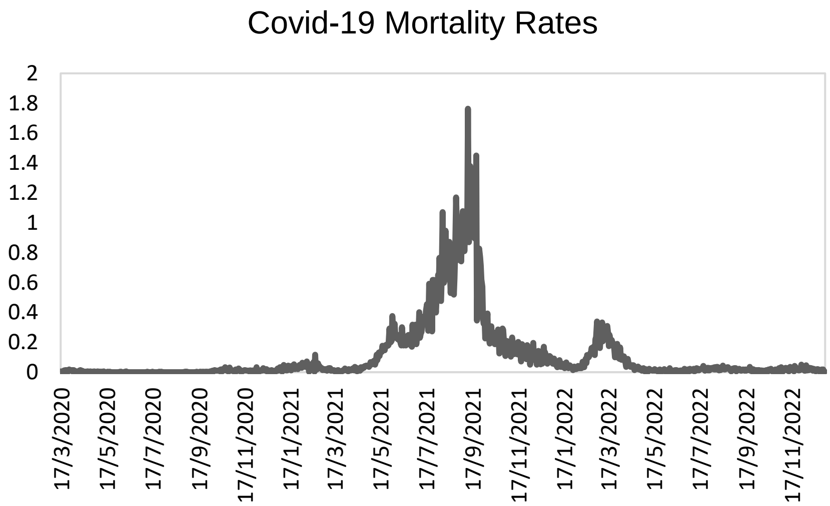



**Fig. 1.** Time series of COVID-19 mortality rates of Malaysia (Mac 2020-November 2021)

### *3.2 Arima Model*

ARIMA model is chosen based on the lag of ACF and PACF. The suggested ARIMA models after the first differencing are ARIMA (0,1,0), ARIMA (0,1,1), ARIMA (0,1,2), ARIMA (0,1,3), ARIMA (0,1,4), ARIMA (1,1,0), ARIMA (1,1,1), ARIMA (1,1,2), ARIMA (1,1,3), and ARIMA (1,1,4). Figure 2 shows ACF and PACF after first differences.

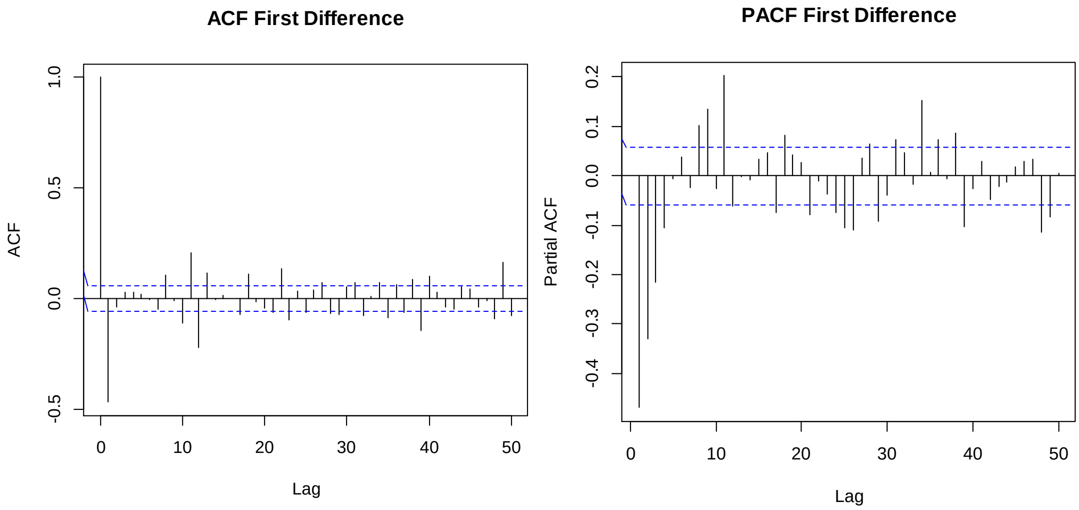



**Fig. 2.** ACF and PACF after first differences

Next, the three best models are chosen based on lowest AIC value, as given in Table 1 below. The three best models that are chosen: ARIMA (1,1,2), ARIMA (1,1,3) and ARIMA (1,1,4).

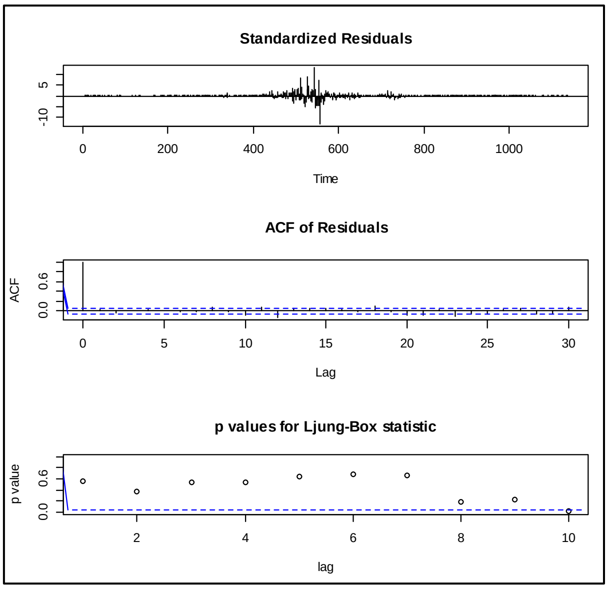
**Table 1**
AIC values of ARIMA

| Model | AIC |
|---|---|
| ARIMA (0,1,0) | -2197.21 |
| ARIMA (0,1,1) | -2578.93 |
| ARIMA(0,1,2) | -2585.32 |
| ARIMA(0,1,3) | -2604.04 |
| ARIMA (0,1,4) | -2610.60 |
| ARIMA (1,1,0) | -2436.54 |
| ARIMA (1,1,1) | -2582.60 |
| ARIMA(1,1,2) | -2621.80 |
| ARIMA(1,1,3) | -2620.54 |
| ARIMA (1,1,4) | -2619.75 |

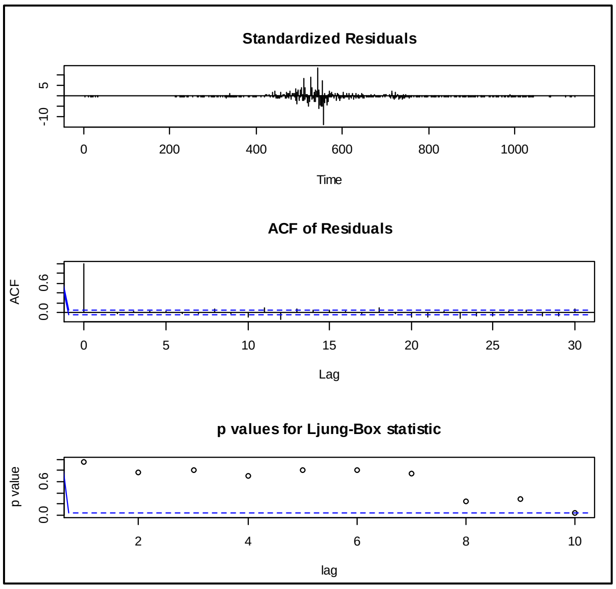
Since the model identification has been obtained in previous, diagnostic checking is then applied to test the adequacy of ARIMA (1,1,2), ARIMA (1,1,3) and ARIMA (1,1,4) in fitting the data. Ljung-Box test is used to check whether the residuals of a time series model are normally distributed. Hypothesis of Ljung-Box test is defined as follows:

$H_o$: The residuals are normally distributed.
$H_1$: The residuals are not normally distributed.

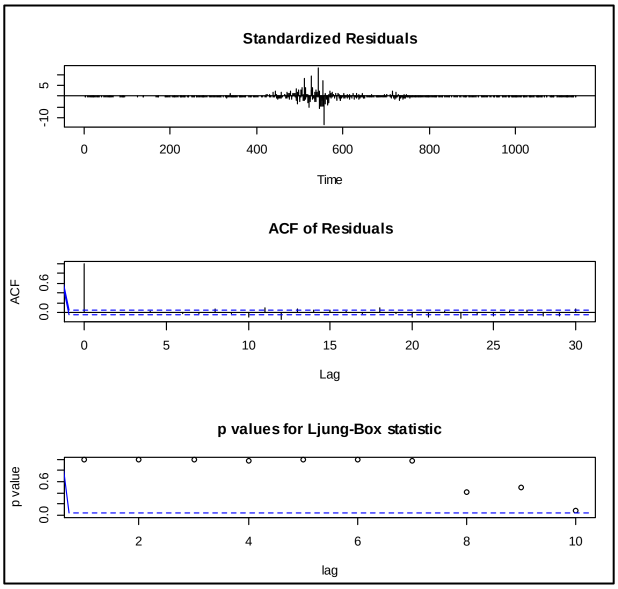
According to Figure 3**,** only ARIMA (1,1,4) is adequate in term of the Ljung-Box test since the *p*-value for all lags is above the significance level of 1%.




a) Residual diagnostic of ARIMA (1,1,2)




b) Residual diagnostic of ARIMA (1,1,3)




c) Residual diagnostic of ARIMA (1,1,4)

**Fig. 3.** Residual diagnostic of ARIMA models (a) Residual diagnostic of ARIMA (1,1,2) (b) Residual diagnostic of ARIMA (1,1,3) and (c) Residual diagnostic of ARIMA (1,1,4)

### *3.2 Performance Between Models*

To measure the performances of modelling, we need to use accuracy criteria. There are a few types of accuracy measures are considered in this study, which are Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). Table 2 tabulates the values of MAE and RMSE for ARIMA models for the in-sample and out-sample data. According to in-sample part in Table 2, ARIMA (1,1,3) outperformed the others since it has the lowest measurement errors. However, according to Figure 3, ARIMA (1,1,3) model is inadequate, thus ARIMA (1,1,4) is selected to forecast the COVID-19 mortality rates.

**Table 2**
Measurement errors of ARIMA models

| | In-sample | | Out-sample | |
|---|---|---|---|---|
| Model | MAE | RMSE | MAE | RMSE |
| ARIMA (1,1,2) | 0.0234 | 0.0622 | 0.0112 | 0.0132 |
| ARIMA (1,1,3) | 0.0233 | 0.0622 | 0.0111 | 0.0131 |
| ARIMA (1,1,4) | 0.0234 | 0.0622 | 0.0111 | 0.0132 |

Figure 4 shows that in the future, the Malaysia's COVID-19 mortality rates is predicted to decreases. This coincides with the initiative taken by the government since the emergence of COVID-19 case, by implementing Movement Control Order (MCO) on March 18, 2020. Although MCO has downturn the economy, but Malaysia manages to record a low number of COVID-19 cases from first and second waves with the first waves ending successfully within less than 2 months [13]. In order to support the Ministry of Health's and the National immunization Programs' goal of developing effective immunization to stop the spread of infectious disease, temporary hospitals were also created. National Vaccination Programs that focused on older people and high-risk population during the early of vaccination campaign has shown a decline number in ICU admission among these population [14]. Currently, the trends from all over the world have shown a decreasing number of COVID-19 cases [15,16]. Therefore, it is not surprising if forecasting result show a decrease in number of COVID-19 cases in the future. With the decreasing rate of COVID-19 cases in the future, government may consider potential actions and strategies to restore economic stability and growth. This can be in terms of tourism activites [17], biodiversity conversation [18], and education [19,20].

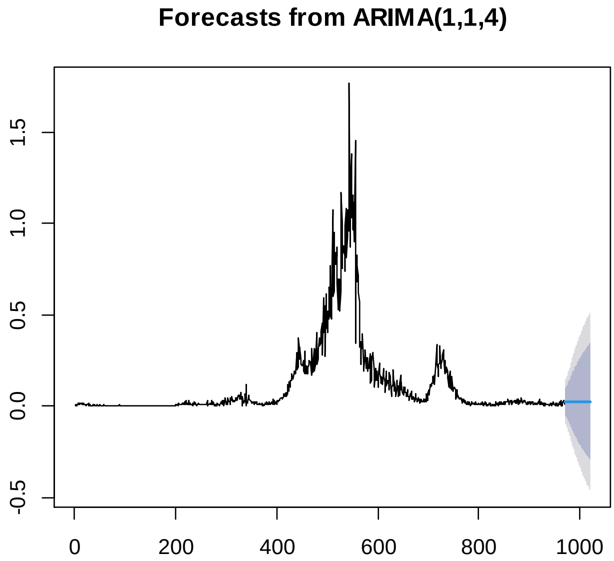



**Fig. 4.** Forecasting performance of ARIMA (1,1,4)

## 4. Conclusions

In this study, ARIMA models are used to predict the COVID-19 mortality rates for Malaysia from the year of 2020 to 2022.There were there ARIMA models that were chosen which were ARIMA (1,1,2), ARIMA (1,1,3) and ARIMA (1,1,4). Based on the residual diagnostic, ARIMA (1,1,4) is selected as the best fit ARIMA model to forecast as compared to others since the residuals of the ARIMA (1,1,4) are normally distributed and independent. This study can be used as a preliminary study to predict the COVID-19 mortality rates and other future pandemic cases to mitigate risk of increasing cases.

**Acknowledgement**
The authors would like to acknowledge the funder by the Ministry of Higher Education, Malaysia under Fundamental Research Grant (FRGS), with Registration Proposal Number: FRGS/1/2020/STG06/UTM/02/6 with vot R.J130000.7854.5F370 and Universiti Teknologi Malaysia (UTM) for UTM Fundamental Research (UTMFR) with vot no Q.J130000.2554.21H65.

**References**
[1] Le Page, Michael. "First known covid case was Wuhan market trader after all." *New Scientist (1971)* 252, no. 3362 (2021): 18. https://doi.org/10.1016/S0262-4079(21)02107-2
[2] Hashim, Jamal Hisham, Mohammad Adam Adman, Zailina Hashim, Mohd Firdaus Mohd Radi, and Soo Chen Kwan. "COVID-19 epidemic in Malaysia: epidemic progression, challenges, and response." *Frontiers in public health* 9 (2021): 560592. https://doi.org/10.3389/fpubh.2021.560592
[3] Nur Hidayah Che Ahmat, Syafiqah Rahamat and Susan Arendt, “COVID-19: Socio-Economic Consequences on the Hospitality Industry in Malaysia,” in Global Academy of Training and Research Enterprise 2021, *Global Conference on Business and Social Sciences Proceeding* 12 (2021): 52-52. https://doi.org/10.35609/gcbssproceeding.2021.12(52)
[4] Guan, Peng, De-Sheng Huang, and Bao-Sen Zhou. "Forecasting model for the incidence of hepatitis A based on artificial neural network." *World journal of gastroenterology: WJG* 10, no. 24 (2004): 3579. https://doi.org/10.3748/wjg.v10.i24.3579
[5] Earnest, Arul, Mark I. Chen, Donald Ng, and Leo Yee Sin. "Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore." *BMC Health Services Research* 5, no. 1 (2005): 1-8. https://doi.org/10.1186/1472-6963-5-36
[6] Liu, Qiyong, Xiaodong Liu, Baofa Jiang, and Weizhong Yang. "Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model." *BMC infectious diseases* 11, no. 1 (2011): 1-7. https://doi.org/10.1186/1471-2334-11-218
[7] Nsoesie, Elaine O., Richard J. Beckman, Sara Shashaani, Kalyani S. Nagaraj, and Madhav V. Marathe. "A simulation optimization approach to epidemic forecasting." *PloS one* 8, no. 6 (2013): e67164. https://doi.org/10.1371/journal.pone.0067164
[8] Kumar, Pavan, Ram Kumar Singh, Chintan Nanda, Himangshu Kalita, Shashikanta Patairiya, Yagya Datt Sharma, Meenu Rani, and Akshaya Srikanth Bhagavathula. "Forecasting COVID-19 impact in India using pandemic waves Nonlinear Growth Models." *MedRxiv* (2020): 2020-03. https://doi.org/10.1101/2020.03.30.20047803
[9] Petropoulos, Fotios, Spyros Makridakis, and Neophytos Stylianou. "COVID-19: Forecasting confirmed cases and deaths with a simple time series model." *International journal of forecasting* 38, no. 2 (2022): 439-452. https://doi.org/10.1016/j.ijforecast.2020.11.010
[10] Ceylan, Zeynep. "Estimation of COVID-19 prevalence in Italy, Spain, and France." *Science of The Total Environment* 729 (2020): 138817. https://doi.org/10.1016/j.scitotenv.2020.138817
[11] Alzahrani, Saleh I., Ibrahim A. Aljamaan, and Ebrahim A. Al-Fakih. "Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions." *Journal of infection and public health* 13, no. 7 (2020): 914-919. https://doi.org/10.1016/j.jiph.2020.06.001
[12] Kufel, Tadeusz. "ARIMA-based forecasting of the dynamics of confirmed Covid-19 cases for selected European countries." *Equilibrium. Quarterly Journal of Economics and Economic Policy* 15, no. 2 (2020): 181-204. https://doi.org/10.24136/eq.2020.009
[13] Shah, Ain Umaira Md, Syafiqah Nur Azrie Safri, Rathedevi Thevadas, Nor Kamariah Noordin, Azmawani Abd Rahman, Zamberi Sekawi, Aini Ideris, and Mohamed Thariq Hameed Sultan. "COVID-19 outbreak in Malaysia: Actions taken by the Malaysian government." *International Journal of Infectious Diseases* 97 (2020): 108-116. https://doi.org/10.1016/j.ijid.2020.05.093
[14] Suah, Jing Lian, Masliyana Husin, Peter Seah Keng Tok, Boon Hwa Tng, Thevesh Thevananthan, Ee Vien Low, Maheshwara Rao Appannan et al. "Waning COVID-19 vaccine effectiveness for BNT162b2 and CoronaVac in Malaysia: an observational study." *International Journal of Infectious Diseases* 119 (2022): 69-76. https://doi.org/10.1016/j.ijid.2022.03.028

[15] Sarker, Rapty, A. S. M. Roknuzzaman, Nazmunnahar, Mohammad Shahriar, Md Jamal Hossain, and Md Rabiul Islam. "The WHO has declared the end of pandemic phase of COVID-19: Way to come back in the normal life." Health Science Reports 6, no. 9 (2023): e1544. https://doi.org/10.1002/hsr2.1544

[16] Sulistyowati, Dita, Hanny Handiyani, Moh Heri Kurniawan, and Rahma Dwi Syukrini. "Exploring the aftermath: Work demands, workplace violence, and job satisfaction among nurses in a public hospital in the post-COVID-19 era." *Belitung Nursing Journal* 9, no. 5 (2023). https://doi.org/10.33546/bnj.2731

[17] Khan, Md Ashraful Azam, and Haslinda Hashim. "The effect of Covid-19 on tourism and hospitality industry in Malaysia, resurgence in the post-pandemic Era: A conceptual criterion." *International Journal of Tourism & Hospitality Review* 7, no. 2 (2020): 54-62. https://doi.org/10.18510/ijthr.2020.726

[18] Sandbrook, Chris, Erik Gómez-Baggethun, and William M. Adams. "Biodiversity conservation in a post-COVID-19 economy." *Oryx* 56, no. 2 (2022): 277-283. https://doi.org/10.1017/S0030605320001039

[19] Abd Rahman, Haliza, Zarina Mohd Khalid, Noraslinda Mohamed Ismail, Nur Arina Bazilah Kamisan, Siti Mariam Norrulashikin, Siti Rohani Mohd Nor, Ani Shabri, and Muhammad Fauzee Hamdan. "Statistical Analysis on Students' Evaluation and Students' Final Exam Marks in Undergraduate Mathematical Courses at Universiti Teknologi Malaysia." *International Journal of Advanced Research in Future Ready Learning and Education* 27, no. 1 (2022): 1-8.

[20] Nor, Siti Rohani Mohd, Adina Najwa Kamarudin, and Nurul Aini Jaafar. "Comparison on the Student's Performances during Physical and Online Learning in Financial Mathematics Course." *International Journal of Advanced Research in Future Ready Learning and Education* 28, no. 1 (2022): 1-8.