
Modelling of infant mortality rate in Nigeria using autoregressive moving average and neural network autoregression

Bamidele Mustapha Oseni^{1*} and Olalekan Joseph Igboroodowo¹

¹Department of Statistics, Federal University of Technology, Akure, Nigeria

^{1*}E-mail: bmoseni@futa.edu.ng

¹E-mail: olalekanjoseph20@gmail.com

Abstract

In the last three decades, the infant mortality rate in Nigeria has been improving but with Nigeria set to miss the Sustainable Development Goal (SDG) target on infant mortality, more need to be done. The study made use of data on infant mortality obtained from the World Bank in determining the infant mortality rate in Nigeria. The trend of infant mortality is identified and models were developed to accurately predict the mortality rate in Nigeria. The time series models; Autoregressive Integrated Moving Average (ARIMA), and Neural Network Autoregression (NNAR); are tested and compared. Model evaluation was done using the five (5) criteria of the Mean Squared Error (MSE), the Root Mean Squared Error (RMSE), the Mean Percentage Error (MPE), the Mean Absolute Error (MAE), and the Mean Absolute Percentage Error (MAPE). Though both models produced similar trends and were found to be suitable for prediction, the NNAR was found to be better than the ARIMA model on all five criteria. Analysis indicates a downward trend in the infant mortality rate in Nigeria and the NNAR model predicts a 12% reduction in the infant mortality rate over the next 10 years.

Keywords: Autoregressive Integrated Moving Average, Neural Network Autoregression, Infant mortality, Time series, Data modelling

1. Introduction

Child mortality is a key indicator of a country's socio-economic growth as well as the quality of life of its citizens, particularly mothers. The survivor-ship of children is mostly at risk between the age of 0 and 5, most especially within the first year of birth. In 2019, globally, 70 per cent of deaths among children and youth under 25 years of age occurred among children under 5 years of age (UN IGME, 2020). What is more disturbing is that a good percentage of these deaths are preventable.

Infant mortality refers to the death of children before their first birthday. Child mortality is divided into three; neonatal mortality (0–27 days), infant mortality (1–11 months), and child mortality (12–59 months). A World Health Organization report states that 5.3 million children

*Corresponding author, e-mail: bmoseni@futa.edu.ng

Submission date: 11/ 02/ 2022

Acceptance date: 28/ 07/ 2022

Published date: 31/ 08/ 2022

under age five died in 2018 alone. The risk of a child dying before completing its first years is very high, especially in Africa (52 per 1000 live births), which is 7 times higher than in Europe (7 per 1000 live births) (World Health Organization, 2018).

Although infant mortality is a global crisis, there are great disparities in the chances of child survival within countries and regions of the world. Sub-Saharan Africa remains the region with the highest infant mortality rate in the world, bearing the burden of more than half the total under-five deaths in the world, with Central and Southern Asia accounting for 28% of total deaths (UN IGME, 2020). Regions like Europe and Northern America have very low child mortality rates. At the country level, the disparity is even more pronounced, 49% of all under-five deaths in 2019 occurred in just five countries: Nigeria, India, Pakistan, the Democratic Republic of the Congo, and Ethiopia.

Globally. From an assessed rate of 65 deaths per 1000 live births in 1990 to 29 deaths per 1000 live births in 2018, the infant mortality rate has declined. Annual infant deaths have declined from 8.7 million in 1990 to 4.0 million in 2018. In Nigeria, infant mortality fell rapidly from 166 deaths per 1000 in 1970 to 75 deaths per 1000 in 2018. But even as such progress is commendable, a lot of work still needs to be done, as Nigeria still has one of the highest infant mortality rates in the world.

Infant deaths are responsible for a good percentage of human deaths. These deaths may be due to a variety of factors. Every year, millions of children die avoidable deaths. Over 9 million children under the age of five die mostly from preventable and treatable diseases (WHO, 2015). To adequately plan for child survival programs in Nigeria, a large investment is required. In light of Nigeria's current economic condition, precise estimates of childhood mortality would aid in the efficient use of the country's scarce health resources.

One of the targets of The Sustainable Development Goals (SDGs) is to end preventable deaths of newborns and children under 5 years of age by 2030, with all countries aiming to reduce neonatal mortality to as low as 12 per 1,000 live births and under-5 mortality to as low as 25 per 1,000 live births. While many countries have reached the target, Nigeria, as well as several sub-Saharan African Countries, are on course to miss the target. Findings from the NDHS (2018), place under-5 mortality in Nigeria at 132 deaths per 1,000 live births. This implies that more than 1 in 8 children in Nigeria dies before their 5th birthday.

Almost one million children die in Nigeria more than in any other country in Africa, largely from preventable diseases Ogunjimi (2012). This makes you wonder what can be done to curb this menace. Causes of infant mortality are largely distinct from those operating at older ages because infants are in many ways the most vulnerable to adverse effects of environmental conditions. The contribution of infant mortality to the total loss of years of human life is substantial because it occurs early in life and its level is relatively higher than other age categories. As infants are sensitive not only to conditions in their immediate environment after birth but also to the pre and post-natal health of their mothers. The concern for infant mortality is quite high because infants are the most vulnerable population to adverse effects of environmental conditions, especially in poor societies (UNFPA, 2007). They are sensitive to the conditions in their immediate environment after birth, as well as their mothers' ante and post-natal health, and the quality of access to health support services.

With almost 200 million people, Nigeria is Africa's most populous country. To significantly reduce the infant mortality rate, there is a need for combined effort to help achieve the SDGs targets. This calls researchers to find out the root causes of persistent higher mortality at every stage of a child's life. The government to step up and intensify its effort to save young lives and families to do better in seeking appropriate medical attention quickly.

Infant deaths are a problem that must be faced and tackled as a people. It can even be argued that infant mortality has a direct impact on the rapid growth of some populations, since there is a high mortality rate, families most especially low-income families tend to give birth to several more children to increase their chances of not being childless at old age, hence leading to a population sprout. This work aims to model the infant mortality rate in Nigeria using both the Autoregressive Moving Average and the Neural Network Autoregression. The models obtained are then compared to determine the better model for the infant mortality rate in Nigeria.

There are several methods used in time series analysis, the autoregressive integrated moving average (ARIMA) model is one of the most important and commonly used time series models. Box and Jenkins popularized a technique that incorporates moving average and autoregressive methods Bartholomew et al (1971).

Researchers are split on what models provide the best results in modelling and forecasting time series, Traditional Statistical Methods or Artificial Neural networks. Several papers have been published supporting one or the other. Gorr (1994) suggested that future research should investigate and better define the border between where ANNs and traditional techniques outperform one another. Other researchers that have compared the two models in modelling some specific data are Maleki, Nasser, Aminabad, and Hadi (2018) who compared the models when utilized in forecasting water treatment plant's influent characteristics, and Perone (2021) who compared the two models with Exponential Smoothing (ETS), trend and seasonal components (TBATS) and hybrid models to forecast the second wave of COVID-19 and hospitalizations in Italy. The basic equation for the ARIMA model can be written as (Davidson, 2000; Perone, 2021)

$$\Delta^d y_t = \phi_1 \Delta^d y_{t-1} + \phi_2 \Delta^d y_{t-2} + \dots + \phi_p \Delta^d y_{t-p} + \gamma_1 \varepsilon_{t-1} + \dots + \gamma_q \varepsilon_{t-q} + \varepsilon_t \quad (1)$$

where y_t is the predicted value at time t , Δ^d difference operator, p is the lag order of the AR process, ϕ is the coefficient of each parameters p , q is the order of the MA process, γ is the coefficient of each parameter q , and ε_t denotes the residuals of the errors at time t . The basic form of the neural network autoregression equation is (Hyndman, 2018; Perone, 2021)

$$y_t = f(\mathbf{y}_{t-1}) + \varepsilon_t \quad (2)$$

where y_t is the predicted values at time t and $\mathbf{y}_{t-1} = (y_{t-1}, y_{t-2}, \dots, y_{t-n})'$ is a vector containing the lagged values of the observed data, f is the neural network with n hidden neurons in a single layer and ε_t is the error at time t .

2. Materials and Methods

2.1 Data Source

This study was conducted using data obtained from the World Bank www.worldbank.org and estimated by United Nations Inter-Agency Group for Childhood Mortality Estimation (UN-IGME). The data contains the mortality estimates for infants in Nigeria between 1964 and 2019.

2.2 Data Analysis

Two Time series models; The Autoregressive Integrated Moving Average (ARIMA) and Neural Network Autoregression (NNAR) were applied to forecast annual Infant mortality rates; their accuracies were compared and the best model was used to forecast infant mortality rate in

the years 2021, 2022, 2023 and 2029. The historical mortality data span from 1964 to 2019, giving a total of 56 observations. The models were developed with the dataset partitioned into two to avoid under/over-fitting a model due to faulty dataset training. The trained dataset consisted of about 80% of the time-series data, while the rest of the data was withheld as a test/validation dataset, to evaluate model accuracy. The performances of these approaches were evaluated by considering their error measurement using the Mean Squared Error (MSE), the Root Mean Squared Error (RMSE), the Mean Percentage Error (MPE), the Mean Absolute Error (MAE), and the Mean Absolute Percentage Error (MAPE). All analysis was performed using R development software (version 4.0.5). The criteria can be respectively obtained using

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2 \quad (3)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (4)$$

$$\text{MPE} = \frac{100\%}{n} \sum_{t=1}^n \frac{y_t - \hat{y}_t}{y_t} \quad (5)$$

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n \left(\frac{|y_t - \hat{y}_t|}{\frac{1}{n-1} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \right) \quad (6)$$

$$\text{MAPE} = \frac{100\%}{n} \sum_{t=1}^n \left(\frac{|y_t - \hat{y}_t|}{y_t} \right) \quad (7)$$

where n is the number of observations, y_t is the actual observation value at time t , and \hat{y}_t is the predicted value of the observation. The RMSE, MSE, and the MAE are particularly very important because minimizing the first two measures leads to the predictions of the mean while minimizing the third measure leads to the prediction of the median (Perone, 2021). Both MAPE and MPE are also very important because they are not scale-dependent. MAPE has been the most widely used of the three measures (Goodwin & Lawton, 1999; Kim & Kim, 2016), however, MPE has the added advantage of also being a measure of bias because both positive and negative forecast errors can offset each other.

3. Results and Discussion

The time plot of the infant mortality rate is shown in Figure 1. The figure shows a downward trend indicating that the infant mortality rate in Nigeria has been constantly decreasing over the years from 195 in 1964 to 74 in 2019. In these years, there is an approximately constant horizontal movement between 1981 and 1994 which shows that the infant mortality rate neither increases significantly nor decreases but fluctuates between 122 and 124.

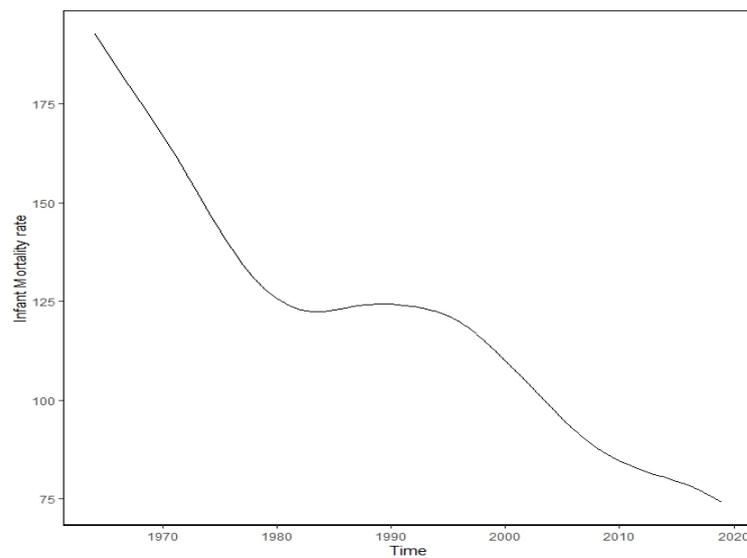


Figure 1 The time series plot of Infant mortality in Nigeria

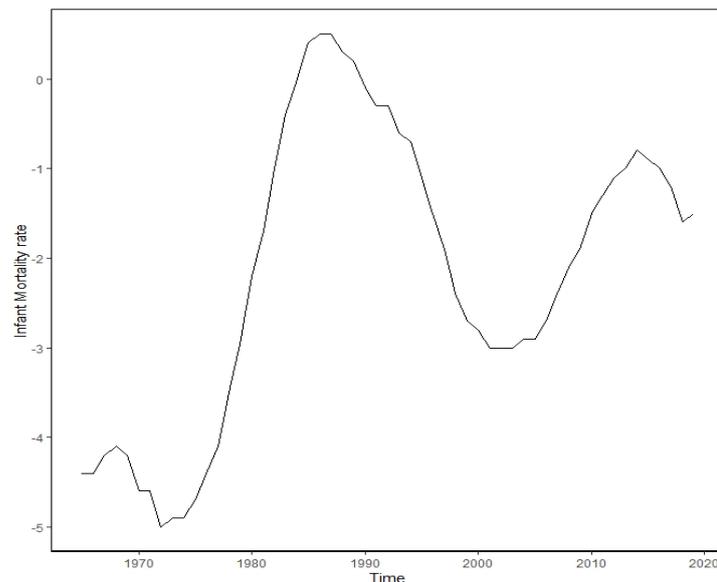


Figure 2 The time series plot after first differencing of Infant mortality dataset in Nigeria

The augmented Dickey-Fuller test is used to investigate the presence of unit root. This is a necessary step before the proper modelling. A 0.05 level of significance is used in testing for the stationary of the series. If the series is stationary, it implies that the data has no unit root. The hypothesis for the test is stated as follows;

Null hypothesis = Series is not stationary (Unit Root present)

Alternative hypothesis = Series is stationary (Unit Root not present)

Table 1 shows the result of the unit root test. From the table, it can be inferred that the series is stationary as the p-value of the Augmented Dickey-Fuller test is obtained as 0.03537 which is less than the level of significance. The plot of the stationary dataset obtained through the first differencing is shown in Figure 2.

Table 1 Unit root test

| ADF Trend | Lag order | p-value |
|-----------|-----------|---------|
| -3.6707 | 2 | 0.03537 |

One of the methods of transforming a stationary time series into non-stationary series is by obtaining the lag. Taking the lag of the data results in stationary series after first differencing. Since a stationary series has been obtained, modelling is conducted using both the ARIMA and NNAR.

3.1 Modelling the data with Autoregressive Integrated Moving Average

The stationary series is modelled using the Autoregressive Integrated Moving Average (ARIMA) technique. The resulting model is evaluated using the Akaike Information Criterion (AIC). The AIC is utilized in the selection of the Autoregressive component as well as the Moving Average component. Also, the “*auto-arima*” functionality of the “*tseries*” package of the R development software is utilized in the model selection. The ARIMA(1,2,2) was selected as the best model based on this criterion. Table 2 shows the estimates of the models. The parameter estimates of the AR1, MA1, and MA2 are obtained respectively as 0.7887, -3.556, and 0.7203 with standard errors of 0.0900, 0.1062, and 0.1647.

Table 2 ARIMA(1,2,2) model estimates for Infant mortality rate in Nigeria

| Models | AR1 | MA1 | MA2 | Sigma | log-likelihood | AIC |
|----------------|--------------------|--------------------|--------------------|---------|----------------|----------|
| ARIMA(1, 2 ,2) | 0.7887 (0.0900) | -3.556 (0.1062) | 0.7203 (0.1647) | 0.02743 | 18.450 | -27.8878 |

Some of the models obtained which are less suited due to higher AIC are ARIMA(1,2,3) with an AIC of -25.50889, ARIMA(2,2,2) with an AIC of -25.49579, ARIMA(4,2,0) with an AIC of -27.72934 and ARIMA(3,2,1) with an AIC of -24.2867.

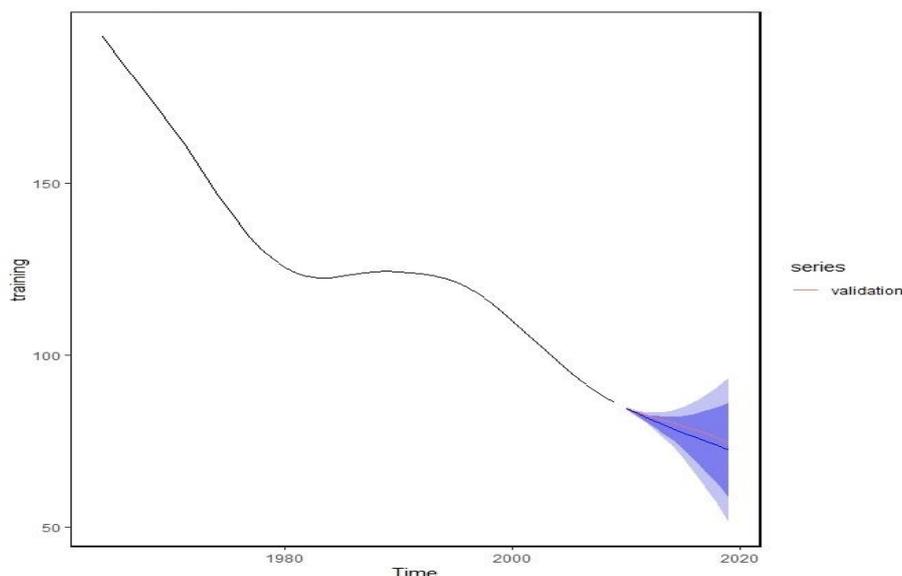
**Figure 3** The forecast from ARIMA (1,2,2).

Figure 3 presents the forecast value obtained from ARIMA(1,2,2). Part of the data which were trained and withheld are data on infant mortality rate between 2010 and 2019. This dataset is used for model validation. The plot in Figure 3 shows that the predicted values from the ARIMA(1,2,2) model are consistently lower than the actual values.

3.2 Modelling the data Neural Network Autoregression

The time series artificial neural network was implemented by the Feed-Forward neural network core algorithm. The model used for the analysis of the dataset on Infant mortality rate in Nigeria is NNAR (1, 1) with an average of 20 networks, each of which is a 1-1-1 network with 4 weights options as linear output units.

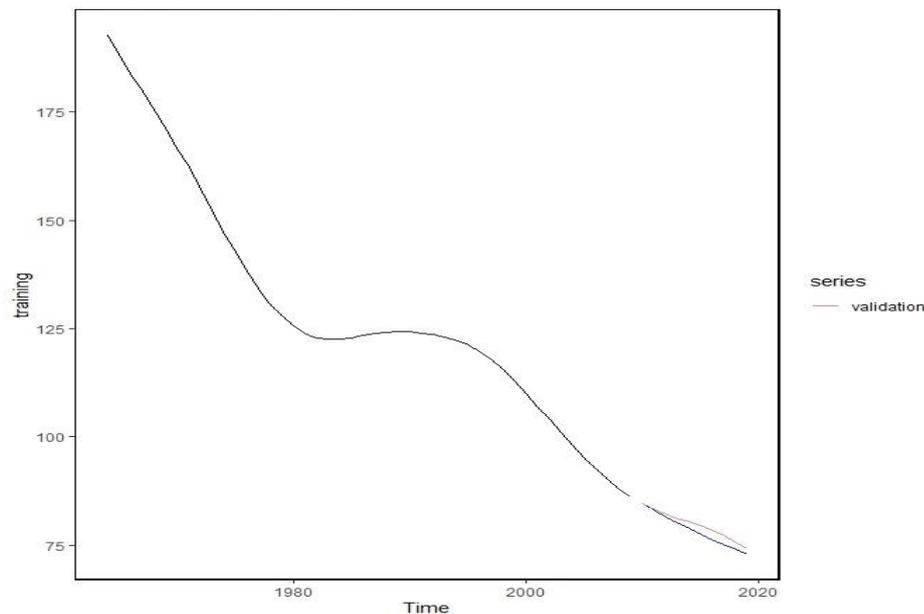


Figure 4 A plot of the predicted values using NNAR(1,1)

Conducting a model validation similar to the validation in Figure 3 carried out for ARIMA(1,2,2) yields the plot in Figure 4. The expected future occurrences according for the years from 2010 to 2019 which were trained and not used in constructing the model show that the predicted values from the model are slightly lower than the actual value and there is a convergence to the actual value as the years further increases.

3.3 Model Comparison and Selection

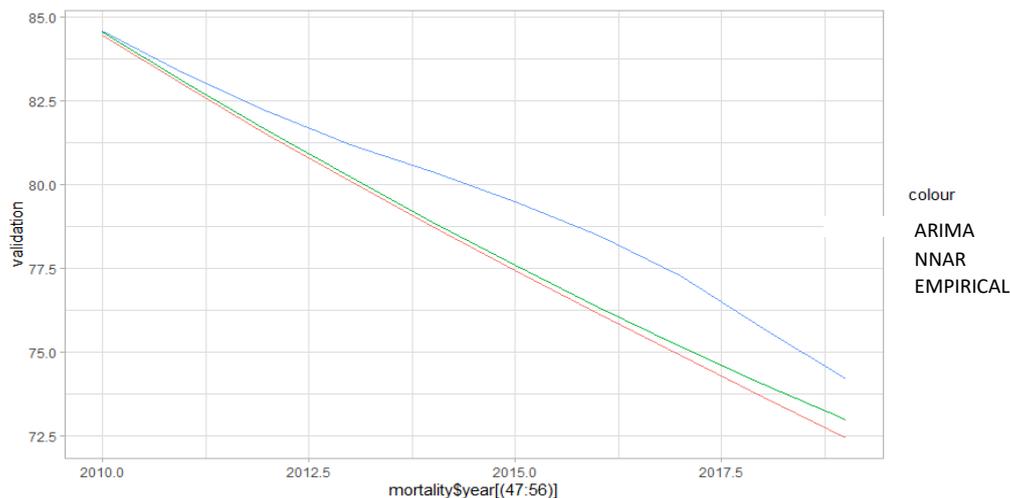
Generally, both models produce values that are slightly different from the actual indicating that both models are good for modelling the data. However, it is always desirable to choose the best model when presented with alternatives.

To compare the performances of these models, five (5) criteria were used. The criteria used are the Mean Squared Error (MSE), the Root Mean Squared Error (RMSE), the Mean Percentage Error (MPE), the Mean Absolute Error (MAE), and the Mean Absolute Percentage Error (MAPE).

Table 3 Model selection criteria

| Criteria | ARIMA(1,2,2) | NNAR(1,1) |
|----------|--------------|-----------|
| MSE | 0.59220 | 0.50433 |
| RMSE | 1.65039 | 1.43391 |
| MPE | 1.85720 | 1.57988 |
| MAE | 1.45286 | 1.23729 |
| MAPE | 1.85720 | 1.57988 |

Table 3 shows the performances of the two models based on the five criteria considered. Based on the value of all five criteria, the NNAR can be considered a better alternative to the ARIMA model when modelling the time series data of infant mortality rate in Nigeria. Figure 5 further shows the validity of the inference. The predicted value of the NNAR is much closer to the actual values through the years considered.

**Figure 5** Plot of the empirical data with the estimated values based on ARIMA and NNAR

Predicting the future years based on the NNAR model shows that the infant mortality rate in Nigeria is at the end of the years 2021, 2022, 2023, and 2029 will be expected to be 72.2986, 71.4252, 70.6021, and 66.6387 which shows that further reduction in the rate is expected in the coming years. This also implies that about a 12% decrease in infant mortality rate can be achieved at the end of 2029.

4. Conclusion

The result of modelling the time series data of infant mortality in Nigeria shows a consistent downward trend. This shows that efforts are being made to reduce the number of children dying before their first birthday. The results obtained after comparing the models indicate that the NNAR model outperformed the traditional time series model (ARIMA) in all performance measures. Moreover, we also signified that Artificial Neural Network provides satisfactory results hence it can be used for forecasting mortality rates in Nigeria with impressive accuracy. The result of the forecast showed a further reduction in the infant mortality rate in Nigeria by about 12%.

5. References

- [1] Bartholomew, D. J., Box, G. E. P., & Jenkins, G. M. (1971), Time Series Analysis Forecasting and Control, Operational Research Quarterly (1970-1977), Vol. 22, No. 2, pp199. <https://doi.org/10.2307/3008255>
- [2] Davidson, J. (2000), Econometric Theory, Hoboken, Wiley Blackwell.
- [3] Goodwin, P., and Lawton, R. (1999), On the asymmetry of the symmetric MAPE, International Journal of Forecasting, Vol. 15 No. 4, pp405-408. [https://doi.org/10.1016/S0169-2070\(99\)00007-2](https://doi.org/10.1016/S0169-2070(99)00007-2)
- [4] Gorr, W. L. (1994), Editorial: Research perspective on neural network forecasting, International Journal of Forecasting, Vol. 10 No. 1, 1–4. [https://doi.org/10.1016/01692070\(94\)90044-2](https://doi.org/10.1016/01692070(94)90044-2)
- [5] Hyndman, R. J. and Athanasopoulos, G. (2018), Forecasting: principles and practice. OTexts: Melbourne. OTexts.com/fpp2
- [6] Kim, S., and Kim, H. (2016), A new metric of absolute percentage error for intermittent demand forecasts, International Journal of Forecasting, Vol. 32 No. 3, pp669-679. <https://doi.org/10.1016/j.ijforecast.2015.12.003>
- [7] Maleki, A., Nasser, S., Aminabad, M. S., & Hadi, M. (2018), Comparison of ARIMA and NNAR model for forecasting water treatment plants influent characteristics, KSCE Journal of Civil Engineering, Vol. 22, pp3233-3245.
- [8] NDHS. (2018), Nigeria Demographic and Health Survey, <http://dhsprogram.com/pubs/pdf/FR359/FR359.pdf>
- [9] Ogunjimi L. O. (2012), Curbing maternal and child mortality: The Nigerian Experience. International Journal of Nursing and Midwifery, 4(3). <https://doi.org/10.5897/ijnm11.030>
- [10] Perone, G. (2021), Comparison of ARIMA, ETS, NNAR, TBATS, and hybrid models to forecast the second wave of COVID-19 hospitalizations in Italy, The European Journal of Health Economics. doi:10.1007/s10198-021-01347-4
- [11] UNFPA. (2007), State of world population, https://www.unfpa.org/sites/default/files/pubpdf/695_filename_sowp2007_eng.pdf
- [12] UN IGME (2020), Levels and trends in child mortality 2020, [Www.unicef.org](http://www.unicef.org). <https://www.unicef.org/reports/levels-and-trends-child-mortality-report-2020>
- [13] World Health Organization (2018), World health statistics 2018: monitoring health for the SDGs, sustainable development goals. World Health Organization, Cop.
- [14] WHO. (2015), December. Health in 2015: From MDGs to SDGs. <https://www.who.int/data/gho/publications/mdgs-sdgs>