




Hybrid Neural Network Prediction for Time Series Analysis of COVID-19 Cases in Nigeria

Adedayo F. Adedotun 

Department of Mathematics, Covenant University, 112212 Ota, Nigeria

* Correspondence: Adedayo F. Adedotun (adedayo.adedotun@covenantuniversity.edu.ng)

Received: 06-08-2022

Revised: 07-16-2022

Accepted: 08-23-2022

Citation: A. F. Adedotun, "Hybrid neural network prediction for time series analysis of COVID-19 cases in Nigeria," *J. Intell Manag. Decis.*, vol. 1, no. 1, pp. 46-55, 2022. <https://doi.org/10.56578/jimd010106>.



© 2022 by the authors. Licensee Acadlore Publishing Services Limited, Hong Kong. This article can be downloaded for free, and reused and quoted with a citation of the original published version, under the CC BY 4.0 license.



Abstract: The lethal coronavirus illness (COVID-19) has evoked worldwide discussion. This contagious, sometimes fatal illness, is caused by the severe acute respiratory syndrome coronavirus 2. So far, COVID-19 has quickly spread to other countries, sickening millions across the globe. To predict the future occurrences of the disease, it is important to develop mathematical models with the fewest errors. In this study, classification and regression tree (CART) models and autoregressive integrated moving averages (ARIMAs) are employed to model and forecast the one-month confirmed COVID-19 cases in Nigeria, using the data on daily confirmed cases. To validate the predictions, these models were compared through data tests. The test results show that the CART regression model outperformed the ARIMA model in terms of accuracy, leading to a fast growth in the number of confirmed COVID-19 cases. The research findings help governments to make proper decisions on how to prepare for the outbreak. Besides, our analysis reveals the lack of quarantine wards in Nigeria, in addition to the insufficiency of medications, medical staff, lockdown decisions, volunteer training, and economic preparation.

Keywords: Autoregressive integrated moving averages (ARIMAs); Classification and regression tree (CART); COVID-19; Prediction

1. Introduction

The extreme acute respiratory coronavirus 2 syndrome (COVID-19) has prompted a global alert. The COVID-19 virus primarily spreads through saliva droplets or nasal discharge when an infected individual coughs or sneezes [1, 2]. The top five affected countries were the US, Brazil, India, Russia, and Spain. Sahai [3] examined the time series data on the overall number of infected patients from these five nations. A 77-day out-of-sample forecast was produced using ARIMA models. By July 31st, India and Brazil would have 1.38 million and 2.47 million, respectively, while the US would have 4.29 million, according to their analysis. In the same vein, Anne [4] used a time series model to predict the short-term transmission of the exponentially growing COVID-19 time series, with the aid of simulation. Taiwo et al. [5] used the autoregressive integrated moving average (ARIMA) model to model and forecast Nigerian confirmed and death cases as a result of the COVID-19 pandemic. This model predicts the number of cumulative instances over time and is validated using Akaike information criterion (AIC) statistics. ARIMA (1,2,0) and ARIMA (1,1,0) were selected to model the confirmed and death cases of COVID-19, respectively. Based on the results of the ARIMA model-building, the two models were demonstrated to be suitable for modeling and forecasting Nigerian COVID-19 data. The predicted values showed that, over the following three months, the number of cumulatively confirmed deaths and cases of COVID-19 in Nigeria may range from 189,019 to 327,426 and from 406 to 3,043, respectively (May 30, 2021). The ARIMA models predicted an alarming daily increase in the number of confirmed COVID-19 death cases in Nigeria.

Ribeiro et al. [6] forecasted the time series one, three, and six days ahead of the COVID-19 cumulative cases in ten Brazilian states using the following tools: Autoregressive integrated moving average (ARIMA), cubist regression (CUBIST), random forest (RF), ridge regression (RIDGE), support vector regression (SVR), and stacking-ensemble learning. In general, these models could provide credible predictions with errors ranging from 0.87% percent to 3.51%. Ceylan [7, 8] also developed auto-regressive integrated moving average (ARIMA) models to project COVID-19 occurrences in Italy, Spain, and France. The relevant data were collected from the

official website of the World Health Organization, from February 21st through April 15th, 2020. Several ARIMA models with different parameters were created. With the lowest mean average percentage errors (MAPEs) (4.7520, 5.8486, and 5.6335), ARIMA (0,2,1), ARIMA (1,2,0), and ARIMA (0,2,1) models were the best prediction tools for Italy, Spain, and France, respectively. ARIMA models are suitable for forecasting COVID-19 prevalence in Italy, Spain, and France. Their findings shed light on the disease patterns, and help assess the epidemiological stage of these locations.

By employing Kuwait as a case study, Alabdulrazzaq et al. [9] assessed and tested the accuracy of an ARIMA model over a reasonable timespan. The best-fit model was employed in Kuwait's progressive prevention plan to forecast confirmed and recovered COVID-19 cases. At a 95% level of significance, the findings were compared to the actual values reported after the forecast period had passed. The Pearson's correlation coefficient between the prediction points and the actual recorded data was determined to be 0.996. This suggests an unbreakable connection between the two sets. Xu et al. [10] integrated data envelopment analysis (DEA) with four different machine learning (ML) approaches to examine the effectiveness and performance of the COVID-19 response in the US. The performance of the COVID-19 response was predicted using environmental variables such as social distance, health policy, and socioeconomic indices. The performance was assessed using Classification and Regression Tree (CART), Boosted Tree (BT), Random Forest (RF), and Logistic Regression (LR). The 23 states had an average efficiency score of 0.97, indicating that they are efficient. Furthermore, the BT and RF models produced the best prediction results, while CART outperformed LR. The most significant factors influencing efficiency, in order, were urbanity, physical inactivity, the total number of tests per person, population density, and hospital beds.

To forecast the COVID-19 outbreak, Ardabili et al. [11] compared machine learning and soft computing approaches. Out of the many machine languages tested, only two models achieved the promising results. This paper acts as a preliminary benchmark to demonstrate machine learning's research potential. Pinter et al. [12] illustrated the usefulness of the hybrid machine learning approach in predicting COVID-19 in Hungary. The researchers proposed to project the time series of infected people and death rate, through a hybrid machine learning strategy using the multi-layered perceptron-imperialist competitive algorithm (MLP-ICA) and adaptive network-based fuzzy inference system (ANFIS). The forecasts predict that the pandemic and overall morale will have greatly declined by late May. The validation process lasts for 9 days and produces good outcomes, demonstrating the model's accuracy. The model is predicted to maintain its accuracy as long as there is no significant disturbance. This paper provides an early benchmark to highlight the promise of machine learning research.

For a number of nations, Chakraborty and Ghosh [13] created short-term (real-time) predictions of upcoming COVID-19 cases as well as risk evaluations (in terms of case fatality rates) for a few particularly badly affected nations. The approach is based on a Wavelet-based forecasting model and an autoregressive integrated moving average model. In the first task, the researchers adopted the optimal regression tree to identify crucial factors that significantly affect case fatality rates across nations. The analysis of early risk estimates for 50 severely affected countries undoubtedly yielded in-depth insights from this data-driven investigation.

Univariate time series models, machine learning, and epidemiologic compartment models have all been used in numerous studies to forecast COVID-19 transmission rates and analyze their effects on public health, urban mobility, and the environment. The goal of this work is to model and forecast one-month confirmed cases of COVID-19 in Nigeria utilizing daily confirmed cases. The CART models and ARIMAs were employed to assure the prediction accuracy and utilize their intrinsic power to explore big data.

2. Methodology

2.1 Data and Descriptive Statistics

The data used for the analysis was accessed online from <https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/owid-covid-data.xlsx>. Under the Creative Commons by license, the data is fully available and licensed. A list of the COVID-19 data preserved by Our World of Data is the full COVID-19 dataset. It is updated regularly and provides reports on reported cases, deaths, and tests, as well as other factors of possible concern.

ARIMA models were created using the methods detailed in Box and Jenkins' classic work, and a CART model was then utilized to form a decision tree. Based on a 4:1 ratio, a training set (45) and a testing set (5) were produced. Modeling was done with the training set, and verification was done with the testing set.

The established models' effectiveness and robustness were assessed using areas under the curve (AUCs) and a confusion matrix. The testing set was used to calculate sensitivity and specificity based on the model attributes. Minitab was used for all statistical analyses, with a significance level of $p < 0.05$.

The ARIMA (p, d, q) model represents the autoregressive integrated moving average (ARIMA) model.

The autoregressive (AR) and moving average (MA) models are combined to create ARMA models. Consider the stochastic process X_t , which is written as

$$X_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (1)$$

where, $\{\varepsilon_t\}$ is a purely random process.

This equation can be rewritten using the lag operator, L , as

$$\varphi(L)X_t = \theta(L)\varepsilon_t \quad (2)$$

where: $\varphi(L)$ and $\theta(L)$ are polynomials of orders p and q , respectively, and are defined as

$$\varphi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p \quad (3)$$

$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q \quad (4)$$

The roots of $\varphi(L) = 0$ must lie outside the unit circle for stationarity, and the roots of $\theta(L)$ must again lie outside the unit circle for invertibility of the MA component. As a result, we have a combination of the autoregressive and moving average processes' "stability" conditions.

The model-building process consists of three steps: identification, parameter estimate, and diagnostic testing. Identification: The ARIMA model orders p , d , and q are used to specify the number of parameters to estimate.

The Box-Jenkins ARIMA approach, on the other hand, can only be used on stationary time series. As a result, determining whether the time series data is stationary is the first stage in creating a Box-Jenkins model. The fundamental justification for obtaining stationary data, according to Gujarati and Porter [14], is that any model derived from this data can be viewed as stable or stationary, providing a valid basis for predicting.

After stationarity has been established, the order (p and q) of the autoregressive and moving average terms must be determined. The autocorrelation (ACF) and partial autocorrelation (PACF) plots are the most fundamental methods for achieving this.

For the ARIMA model, a software package was employed. The parameters were estimated using the Akaike information criterion (AIC) and Bayesian information criterion (BIC) values. A model with the lowest AIC, BIC, and Q-statistics, as well as a high R-square, could be considered suitable for predicting [15]. The model is regarded as unacceptable in an application if the computed p-value associated with the Q-Statistics is modest (p-value) [14]. As a result, the analytical procedure should be repeated until a satisfactory model is found.

The first step in creating an ARIMA model is to determine whether the variable being forecasted is stationary in time series. By stationary, we imply that the values of a variable change around a constant mean and variance across time. We won't be able to build the ARIMA model until this series is stationary. To create an ARIMA (p, d, q) model with " d " as the order of differencing, we must first difference the time series " d " times to generate a stationary series. When differencing, exercise caution because excessive differencing will result in an increase in the standard deviation rather than a decrease. Starting with the lowest order (of the first order, $d=1$) differencing and testing the data for unit root problems is the best strategy. As a result, we obtained a first-order differencing time series.

We now look at the regression tree. The data is $D = \{(x^{(i)}, y^{(i)})\}_{i=1}^N$, where $x^{(i)} \in X$ and $y^{(i)} \in \mathfrak{Y}$. Typically, $X = \mathbb{R}^d$ and $\mathfrak{Y} = \mathbb{R}^K$. The goal of the regression tree algorithm is to construct a function $f : \mathfrak{X} \rightarrow \mathfrak{Y}$ such that the error is small:

$$\sum_i |f(x^{(i)}) - y^{(i)}|^2 \quad (5)$$

The way to do it is to construct a tree and define a constant value on each subregion corresponding to the terminal node of the tree. Thus f constructed this way is a piecewise constant function.

In particular, any node $t \in T$ corresponds to a subset of \mathfrak{X} . On each node t , define the average y -value $\bar{y}(t)$ of the data on the node t by

$$\bar{y}(t) = \frac{1}{N(t)} \sum_{x^{(i)} \in t} y^{(i)} \quad (6)$$

which is an estimator of $E[Y | X \in t]$. We also define the (squared) error rate $r(t)$ of node t by

$$r(t) = \frac{1}{N(t)} \sum_{x^{(i)} \in t} (y^{(i)} - \bar{y}(t))^2 \quad (7)$$

It is nothing but the variance of the node t , which is also an estimator of

$$Var(Y|X \in t) = \sigma^2(Y|X \in t) \quad (8)$$

We define the cost $R(t)$ of the node t by

$$R(t) = r(t)p(t) \quad (9)$$

Recall that $p(t) = \frac{N(t)}{N}$. Therefore

$$R(t) = \frac{1}{N} \sum_{x \in t} (y^{(i)} - \bar{y}(t))^2 \quad (10)$$

Let s be a split of a node t . Define the decrease $\Delta R(s, t)$ of the cost by s as

$$\Delta R(s, t) = R(t) - R(t_L) - R(t_R) \quad (11)$$

The splitting rule at t is s^* such that we take the split s^* among all possible candidate splits that decrease the cost most. Namely,

$$\Delta R(s^*, t) = \max_s \Delta R(s, t) \quad (12)$$

One may use this splitting rule for the split of the classification tree. This way, we can grow the regression tree to T_{max} . As before, one quick rule of thumb is that one stops splitting the node if the number of elements of the node is less than the preset number. Once T_{max} is found, we can prune back. The pruning method for the regression tree is the same for the classification tree except that we define $R^{ts}(T)$ and $R^{cv}(T)$ differently.

3. Results and Discussion

3.1 Time Plot and Stationarity

From Figure 1 below, it was observed that the pattern of the graph indicates series non-stationarity. There is an upward trend for covid cases. The autocorrelation plot in Figure 2 indicates significant spikes up to lags 35, a downward trend from lag to lag, and a slight cut-off from lag 35 which also indicates an element of non-stationarity. The partial autocorrelation also tails after lag 1 with a significant spike at lag 1.

From Figure 3, it was observed that the pattern of the graph indicates series non-stationarity. There is an upward trend for covid cases. The autocorrelation plot in Figure 4 indicates significant spikes up to lags 35, a downward trend from lag to lag, and a slight cut-off from lag 35 which also indicates an element of non-stationarity. The partial autocorrelation also tails after lag 1 with a significant spike at lag 1.

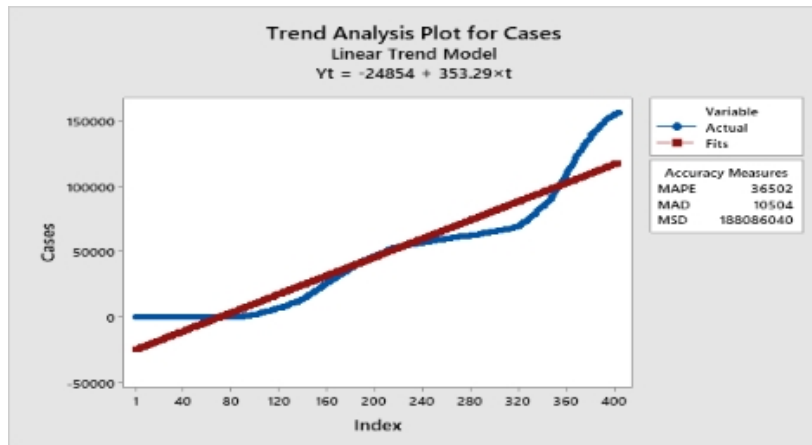


Figure 1. Daily covid cases

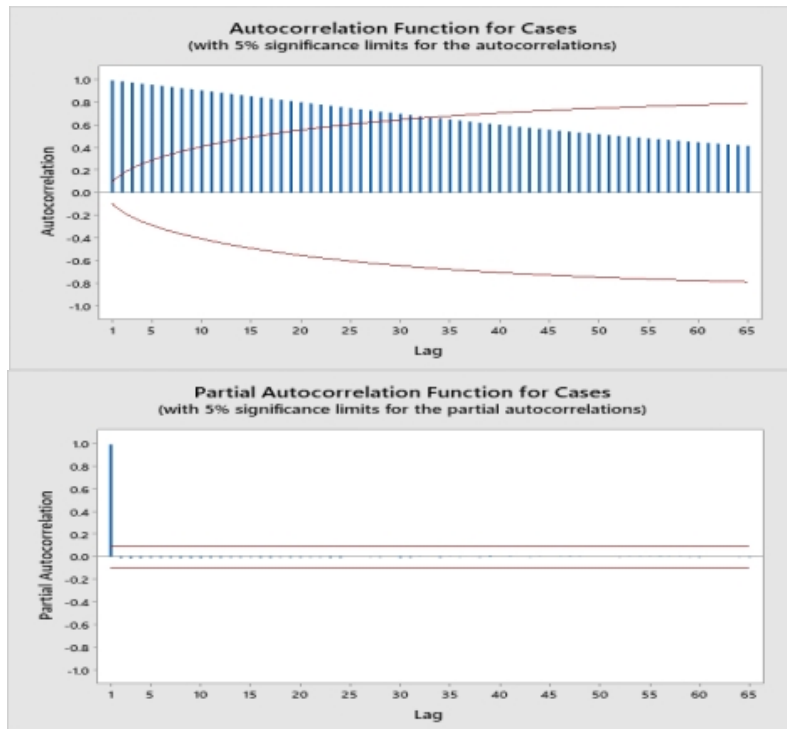


Figure 2. ACF and PACF for covid cases

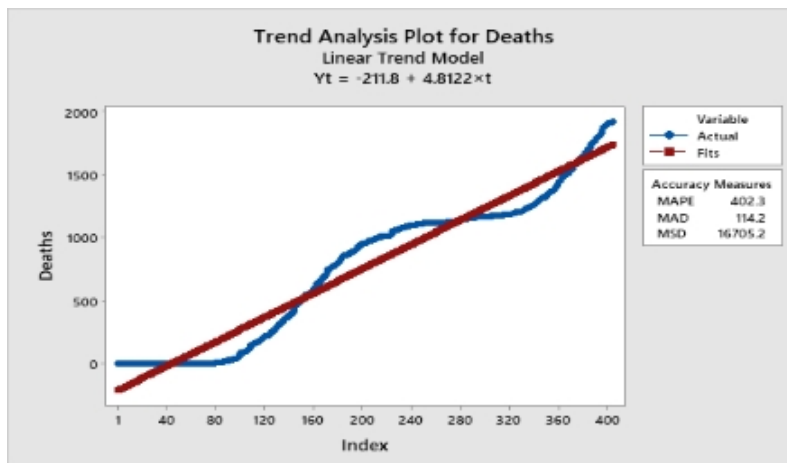
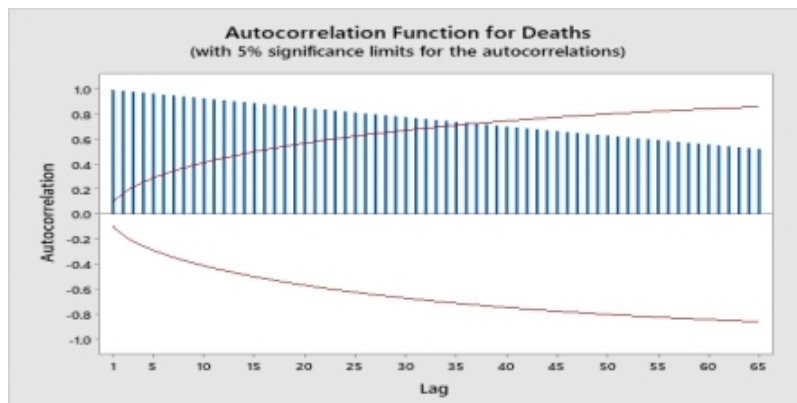


Figure 3. Daily covid deaths



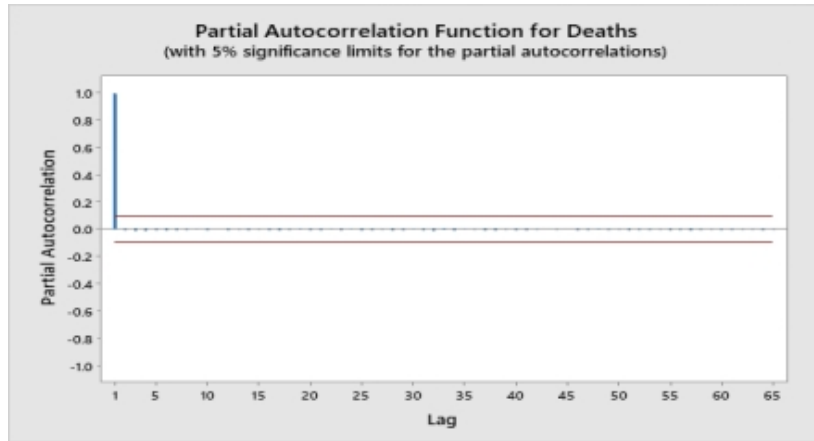


Figure 4. ACF and PACF for covid deaths

3.2 ARIMA Modelling

Having made the series stationary, the decision was made on reasonable values of the orders of the Autoregressive (AR(ϕ)), ordinary differencing, Moving Average (MA(θ)).

After trying different ARIMA models of various orders, to choose the best model, we look for the model with the least AIC. Brockwell and Davis (1991) in their research suggest that AIC is the primary criterion in selecting the orders of a time series.

After various trials, it was discovered that the ARIMA model (3,1,0) for covid cases gives the minimum MSE. This is observed in Table 1.

After various trials, it was discovered that the ARIMA model (1,1,0) for covid deaths gives the minimum MSE. This is observed in Table 2.

The ultimate aim of building any time series model is forecasting. If this objective is not achieved, the work is incomplete. Forecasts were made for the possible number of covid cases and deaths. Based on the chosen model, the forecast for 12 months is seen in Table 3.

Table 1. Final estimates of parameters for covid cases

| Type | Coef | SE Coef | T-Value | P-Value |
|----------|--------|---------|---------|---------|
| AR 1 | 0.4756 | 0.0463 | 10.28 | 0.000 |
| AR 2 | 0.1987 | 0.0507 | 3.92 | 0.000 |
| AR 3 | 0.2748 | 0.0463 | 5.94 | 0.000 |
| Constant | 16.91 | 8.41 | 2.01 | 0.045 |

Differencing: 1 regular difference

Table 2. Final estimates of parameters for covid deaths

| Type | Coef | SE Coef | T-Value | P-Value |
|----------|--------|---------|---------|---------|
| AR 1 | 0.5257 | 0.0425 | 12.37 | 0.000 |
| Constant | 2.252 | 0.238 | 9.45 | 0.000 |

Differencing: 1 regular difference

Table 3. Forecasts for covid cases

| Period | Forecast | 95% Limits | | Actual |
|--------|----------|------------|--------|--------|
| | | Lower | Upper | |
| 405 | 156347 | 155997 | 156697 | |
| 406 | 156660 | 156037 | 157282 | |
| 407 | 156991 | 156081 | 157900 | |
| 408 | 157319 | 156065 | 158573 | |
| 409 | 157645 | 156021 | 159270 | |
| 410 | 157974 | 155958 | 159991 | |
| 411 | 158304 | 155872 | 160736 | |
| 412 | 158634 | 155767 | 161500 | |
| 413 | 158965 | 155647 | 162282 | |
| 414 | 159296 | 155512 | 163080 | |
| 415 | 159628 | 155365 | 163892 | |
| 416 | 159961 | 155206 | 164716 | |

Table 4. Forecasts for covid deaths from period 404

| Period | Forecast | 95% Limits | | Actual |
|--------|----------|------------|---------|--------|
| | | Lower | Upper | |
| 405 | 1921.46 | 1912.08 | 1930.83 | |
| 406 | 1927.10 | 1910.00 | 1944.21 | |
| 407 | 1932.32 | 1908.28 | 1956.37 | |
| 408 | 1937.32 | 1907.13 | 1967.51 | |
| 409 | 1942.20 | 1906.54 | 1977.86 | |
| 410 | 1947.02 | 1906.44 | 1987.59 | |
| 411 | 1951.80 | 1906.76 | 1996.84 | |
| 412 | 1956.57 | 1907.43 | 2005.71 | |
| 413 | 1961.32 | 1908.38 | 2014.27 | |
| 414 | 1966.08 | 1909.57 | 2022.58 | |
| 415 | 1970.83 | 1910.97 | 2030.69 | |
| 416 | 1975.58 | 1912.54 | 2038.61 | |

The result in Table 4 shows that there is likely to be an increase in the number of covid cases as well as its corresponding deaths.

3.3 Node CART® Regression: Total_deaths Versus Total_cases

The result of the response information in Table 5 showed that the mean of the variable to be 1027.05, with a standard deviation of 668.719. The kurtosis value is 5.18358 and this implied the series is not normally distributed.

Figure 5 depicts a trend in which the R^2 statistic rises quickly for the first few nodes before leveling out. The researchers wish to look at the performance of some of the even smaller trees that are similar to the tree in the results because this chart reveals that the R^2 value is generally steady between trees with around 45 nodes and trees with approximately 70 nodes.

Figure 6 illustrates a tree diagram of the k-fold cross-validation study, which shows all cases from the entire data set. The table of fits and error statistics, as well as the topic categorization criteria, provide further information about the terminal nodes.

The values for the training and test statistics are near. Table 6 indicates that the tree is not overfitted because the R^2 statistic is nearly as high as the 45-node tree, the study then decides to investigate the associations between the predictor factors and the response values using the 45-node tree.

Figure 7 illustrates the scatterplot of response fits versus actual values. The graphs demonstrate that the predicted values are extremely close to the actual ones.

Table 5. Response information

| Mean | StDev | Minimum | Q1 | Median | Q3 | Maximum |
|---------|---------|---------|-------|--------|--------|---------|
| 1027.05 | 668.719 | 0 | 439.5 | 1113 | 1481.5 | 2065 |

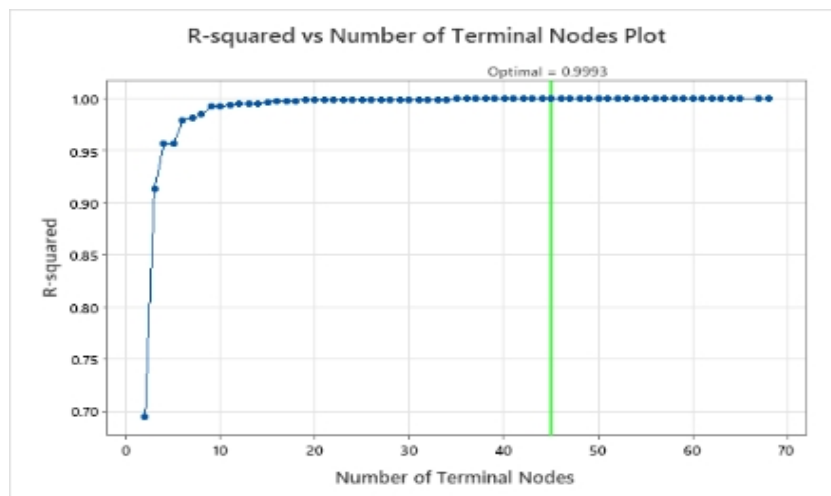


Figure 5. R-squared vs number of terminal nodes plot

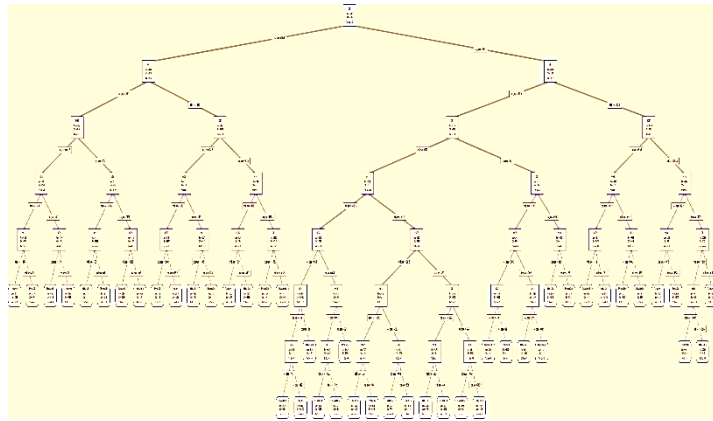


Figure 6. Optimal tree diagram

Table 6. Model summary

| | | |
|---|----------|----------|
| Total predictors | 1 | |
| Important predictors | 1 | |
| Number of terminal nodes | 45 | |
| Minimum terminal node size | 3 | |
| Statistics | Training | Test |
| R-squared | 0.9997 | 0.9993 |
| Root mean squared error (RMSE) | 11.3639 | 17.2089 |
| Mean squared error (MSE) | 129.1391 | 296.1451 |
| Mean absolute deviation (MAD) | 8.2412 | 12.1177 |
| Mean absolute percent error (MAPE) | 0.0719 | 0.0829 |

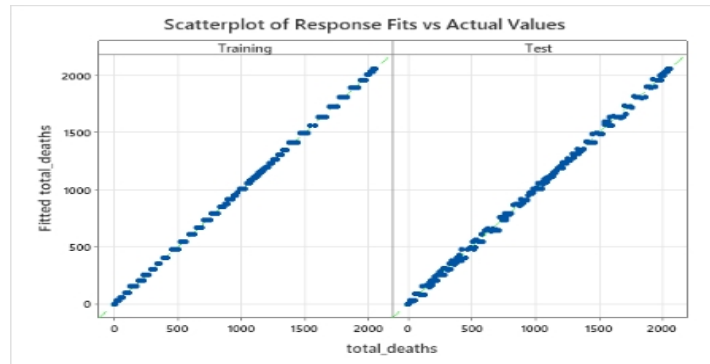


Figure 7. Scatterplot of response fits versus the actual values

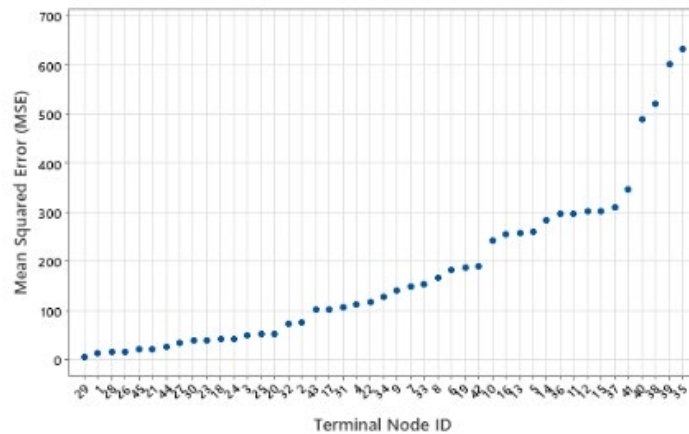


Figure 8. Scatterplot of MSE versus terminal node

The terminal node's MSE in Figure 8 demonstrates that node 8 is the least precise of the terminal nodes. You can have a higher level of trust in the accuracy of the fits for nodes with lower MSE values. If there is a way to lessen or explain the variation, the examples in terminal node 8 have the best chance of improving the tree.

Figure 9 displays a plot of residuals by the terminal node that reveals the fit is too large for a tiny cluster of patients in Terminal Node 8. The researchers look into why some of these patients use services for a shorter period than the average patient in their group.

Figure 9 also shows clusters or outliers are shown in the plot of residuals by the terminal node. In Terminal Node 1 and Terminal Node 7, there is one residue that looks to be significantly larger than the others.

The results reveal that the tree's performance on new data is similar to that on training data. Similar trends may be seen in the points for the training and test data sets.

Table 7 presents the fit and error data for each node in each row. In order of least to highest inaccuracy, the best nodes are listed first. The mean response of the cases in node 29 has the best fit value of 1179.78. It is chosen because the MSE, MAD, and MAPE are least compared to the other best terminal nodes.

Table 8 shows the criteria for classifying subjects into the best 5 terminal nodes. This implies that each row of the table lists the values of the predictors for a terminal node. For Node 29 with fit, 1179.78, the predicted total cases will be between 6,7697 and 71,006. For Node 1, with fit value of 2.55, the predicted total cases will be less than or equal to 467. For Node 28 with fit, 1165.73, the predicted total cases will be between 641377 and 67697. For Node 26 with fit, 1127, the predicted total cases will be between 61,088 and 62,297. For Node 45 with fit, 2,060, the predicted total cases will be greater than 162,438.

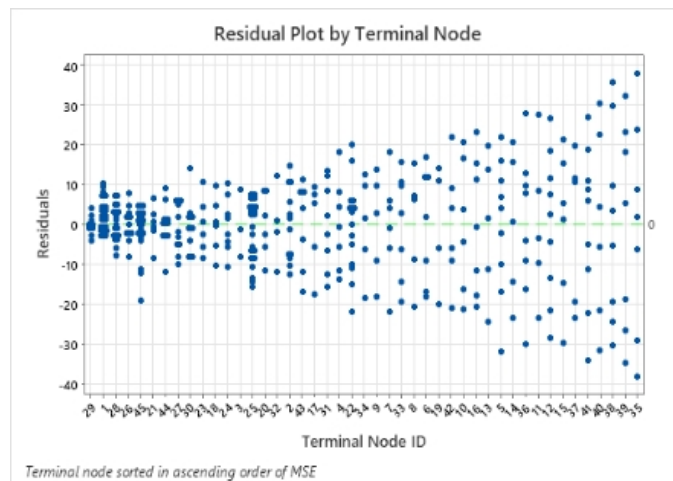


Figure 9. Residual plot by terminal node

Table 7. Fits and error statistics for best 5 terminal nodes

| Terminal Node | Count | Fit | StDev | MSE | MAD | MAPE |
|---------------|-------|---------|---------|---------|---------|----------|
| 29 | 9 | 1179.78 | 2.29868 | 5.2840 | 1.80247 | 0.001528 |
| 1 | 49 | 2.55 | 3.75288 | 14.0841 | 3.03207 | 0.827121 |
| 28 | 22 | 1165.73 | 4.08080 | 16.6529 | 3.52066 | 0.003019 |
| 26 | 12 | 1127.00 | 4.10284 | 16.8333 | 3.33333 | 0.002957 |
| 45 | 44 | 2060.00 | 4.52769 | 20.5000 | 3.00000 | 0.001459 |

Table 8. Criteria for classifying subjects into best 5 terminal nodes

| Terminal Node | Fit | Criterion |
|---------------|---------|----------------------------------|
| 29 | 1179.78 | 67697.5 < total_cases <= 71006.5 |
| 1 | 2.55 | total_cases <= 467.5 |
| 28 | 1165.73 | 64137 < total_cases <= 67697.5 |
| 26 | 1127.00 | 61088 < total_cases <= 62297.5 |
| 45 | 2060.00 | total_cases > 162438 |

4. Conclusion

The COVID-19 cases and deaths exhibited non-stationarity. The Autoregressive Integrated Moving Average (ARIMA) proposed by Box-Jenkins was employed to analyse COVID-19 cases and deaths from March 2020. The study is mainly to model and forecast the monthly covid cases and deaths for twelve months. Moreover, several models were developed but based on minimum corrected Akaike Information Criteria (AIC) value, estimation of

necessary parameters and series of diagnostic tests were performed. It was observed that ARIMA (3,1,0) model was the best model for modelling covid cases, while ARIMA (1,1,0) model was the best model for modelling covid deaths.

The CART model gave a 99.97% accuracy score for the training set and a 99.93% accuracy score the test set. Each round square box represents a node, with a number at the top indicating the node's ID. The value on the box's first line represents the mean of all observations in that node (if it is a leaf node, the mean is utilized for prediction). Unpruned trees produce better forecasts than trimmed ones. The tree model predicts the same thing for all observations that occur under the same leaf node, which approximates the underlying pattern to a large extent. A larger sample size would be necessary to obtain a more precise estimate. By tracking the overall drop in the optimization criterion, an aggregate measure may be developed to emphasize the significance of each feature in the model.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflict of Interest

The authors declare that they have no conflicts of interest.

References

- [1] G. O. Odekina, A. F. Adedotun, and O. F. Imaga, "Modeling and forecasting the third wave of COVID-19 incidence rate in nigeria using vector autoregressive model approach," *J. Niger. Soc. Phys. Sci.*, vol. 4, no. 1, pp. 117-122, 2022. <https://doi.org/10.46481/jnsps.2021.431>.
- [2] G. O. Odekina, A. F. Adedotun, and O. A. Odusanya, "Vector autoregressive modeling of COVID-19 incidence rate in Nigeria," *Int J. Des. Nat. Ecodyn.*, vol. 16, no. 6, pp. 665-669, 2021. <https://doi.org/10.18280/ijndne>.
- [3] A. K. Sahai, N. Rath, V. Sood, and M. P. Singh, "ARIMA modelling & forecasting of COVID-19 in top five affected countries," *Diabetes Metab. Syndr. Clin. Res. Rev.*, vol. 14, no. 5, pp. 1419-1427, 2020. <http://dx.doi.org/10.1016/j.dsx.2020.07.042>.
- [4] "ARIMA modelling of predicting COVID-19 infections," *medRxiv*, 2020.
- [5] A. I. Taiwo, A. F. Adedotun, and T. O. Olatayo, "Nigerian COVID-19 Incidence Modeling and Forecasting with Univariate Time Series Model," In *International Series in Operations Research & Management Science*, Athens, 2020, Springer, vol. 320. https://doi.org/10.1007/978-3-030-87019-5_8.
- [6] M. H. D. M. Ribeiro, R. G. da Silva, V. C. Mariani, and L. dos Santos Coelho, "Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil," *Chaos. Soliton. Fract.*, vol. 135, Article ID: 109853, 2020. <https://doi.org/10.1016/j.chaos.2020.109853>.
- [7] Z. Ceylan, "Estimation of COVID-19 prevalence in Italy, Spain, and France," *Sci. Total Environ.*, vol. 729, Article ID: 138817, 2020. <https://doi.org/10.1016/j.scitotenv.2020.138817>.
- [8] M. Alazab, A. Albara, M. Abdelwadood, A. Ajith, J. Vansh, and A. Salah, "COVID-19 prediction and detection using deep learning," *Int J. Comput. Inf. Syst. Ind. Manage. Appl.*, vol. 12, pp. 168-181, 2020.
- [9] H. Alabdulrazzaq, M. N. Alenezi, Y. Rawajfih, B. A. Alghannam, A. A. Al-Hassan, and F. S. Al-Anzi, "On the accuracy of ARIMA based prediction of COVID-19 spread," *Results Phys.*, vol. 27, Article ID: 104509, 2021. <http://dx.doi.org/10.1016/j.rinp.2021.104509>.
- [10] Y. Xu, Y. S. Park, and J. D. Park, "Measuring the response performance of US States against COVID-19 using an integrated DEA, CART, and logistic regression approach," *Healthcare*, vol. 9, no. 3, pp. 268-268, 2021. <http://dx.doi.org/10.3390/healthcare9030268>.
- [11] S. F. Ardabili, A. Mosavi, P. Ghamisi, F. Ferdinand, A. R. Varkonyi-Koczy, U. Reuter, T. Rabczuk, and P. M. Atkinson, "COVID-19 outbreak prediction with machine learning," *Algorithms*, vol. 13, no. 10, pp. 249-249, 2020. <http://dx.doi.org/10.3390/a13100249>.
- [12] G. Pinter, I. Felde, A. Mosavi, P. Ghamisi, and R. Gloaguen, "COVID-19 pandemic prediction for Hungary: a hybrid machine learning approach," *Mathematics*, vol. 8, no. 6, pp. 890-890, 2020. <http://dx.doi.org/10.3390/math8060890>.
- [13] T. Chakraborty and I. Ghosh, "Real-time forecasts and risk assessment of novel coronavirus (Covid-19) cases: A data-driven analysis," *Chaos. Soliton. Fract.*, vol. 135, Article ID: 109850, 2020. <http://dx.doi.org/10.1016/j.chaos.2020.109850>.
- [14] "Basic Econometrics," Inc New York, 2009.
- [15] R. Biswas and B. Bhattacharyya, "ARIMA modeling to forecast area and production of rice in West Bengal," *J. Crop Weed*, vol. 9, no. 2, pp. 26-31, 2013.