

Modeling of covid-19 cases of selected states in Nigeria using linear and non-linear prediction models

Modelowanie przypadków COVID-19 w wybranych stanach Nigerii przy użyciu liniowych i nieliniowych modeli predykcyjnych

Babatunde Abdulrauph Olarenwaju^a, Igboeli Uchenna Harrison^{b,*}

^a Department of Computer Sciences, University of Ilorin, Nigeria

^b Department of Computer Science, University of Abuja, Nigeria

Abstract

COVID-19 has stamped an indelible mark in the history of humanity as one of the recorded deadly virus that has wiped out millions of lives on planet earth many whose exact cause of death cannot be account for due to lack of knowledge. It has become a household name in every nook and cranny from developed to the underdeveloped nations of the world. Most of the prominent signs of COVID-19 like fever, cough, difficulty in breathing and accessional muscle pain can also resemble those of many other notable diseases thereby making it highly necessary to undergo a diagnostic test to be able to categorically identify COVID-19 patients. The use of medical diagnostic tests can also help determine patients who have recovered from COVID-19. Various studies abound with researchers trying to predict and even forecast the level of damage and disruption of economic activities this may have brought to almost every nation of the world. This research attempts to find out the nature of the spread of the virus using Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Networks (ANN). The essence is to ascertain the exact model to use in forecasting the future occurrence of the pandemic especially at this stage where the second wave of the pandemic is in view. The study found that both linear and nonlinear predictions models can fit the trend of the virus in Nigeria with ARIMA producing results of over 97% on a 120-day period while ANN produced results of about 98.01% in some states. We conclude that future waves of the virus in addition to other epidemics of this nature can be predicted with high degree of accuracy with ARIMA or ANN.

Keywords: ARIMA; ANN; Prediction; Pandemic

*Corresponding author

Email address: uchenna.igboeli@uniabuja.edu.ng (I. U. Harrison)

©Published under Creative Common License (CC BY-SA v4.0)

1. Introduction

Coronavirus 2019 is an infectious and transmittable viral disease caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), which emerged in Wuhan, China [1]. The epidemic which has negatively impacted almost every nation of the world, began in Guangdong province, China, in the late 2002. It spread to Hong Kong on February 21, 2003 [2], and from there to what has been described as a pandemic today. The rate at which the disease is spreading is alarming as it has being reported in almost every continent from Asia to America, Australia, Europe and Africa. With the spread of the dreaded virus into various parts of the world, the Nigeria government took some pro-active measures to prevent it from getting into the country. This mainly took the form of taking steps to enforce strict security checks at the airport as early as in January 2020. These measures yielded insufficient results as the country recorded her index case late February from a victim that imported the virus from Europe. The announcement of the presence of the virus in Nigeria immediately brought great fear and concern to all Nigerians as many began to doubt the effectiveness of the surveillance operations deployed in our airports and the country's general preparedness as the dreaded virus continued to spread around Europe,

Asia and United States in an astronomical rate. The index case of the virus here in Nigeria, had prior to testing positive, visited some other states of the apart from Lagos state. This is a proof of the inadequate level of readiness to combat the virus especially considering the level of publicity given to the pandemia worldwide. Just after the index case was detected, the NCDC launched a National Emergency Operations Centre (EOC) to oversee the national response to COVID-19. As a follow-up, the Presidential Task Force (PTF) for coronavirus control was also inaugurated on March 9, 2020 [3].

As at August 13, 2020, Nigeria has recorded about 47,743 confirmed cases of COVID-19 with 12,844 active and 33,943 cases already discharged while having about 956 deaths. This result is coming out of a total of 338,084 samples tested throughout the country [4]. The country has a total of 62 NCDC COVID-19 Laboratories scattered in all the states of the federation as at 13th of August 2020 for a population of about 200,000,000. Andres et. al., (2020) found that there exists a relationship between COVID-19 behavior and population in the region being considered. One can now argue that the few number of cases recorded in Nigeria may not adequately reflect the true position of the spread of COVID-19. Jester, Uyeki & Jernigan (2018)

[5] traced the outbreak of epidemic from 1918, 1957, 1968 and 2009 till date and noted that all the four pandemics in last 100 years have had some genes that originated from avian influenza viruses. These viruses are constantly changing and therefore require ongoing surveillance and frequent vaccine virus changes. Improvements in healthcare, scientific researches, vaccines and effective communications have however improved pandemic response. There exists various works and studies on the prediction of cases and mortality rates related to the COVID-19 dreaded disease. This research is aimed at examining the effectiveness of both the ANN and ARIMA models in predicting the rate of spread of the virus in Nigeria. Its major significance is to compare a linear (ARIMA) and a non-linear method (ANN) of time series forecasting when applied to the confirmed cases of covid-19 in Nigeria as it will help the concerned agencies in tackling the prevalence and future nature of the disease. Oladipo & Babatunde (2016) [6] used Decision Tree, Logistic Regression and ANN for the prediction of diabetes among two sets of people from china and arrived at the conclusion that machine learning algorithms can efficiently predict such an ailment that is regarded as among the top five causes of death found in the United States [7].

2. Methodology

In this study, daily COVID-19 situation reports as presented by the Nigeria Centre for Disease Control (<https://covid19.ncdc.gov.ng/>) was used. The data covers a sample of 10 states for the period of 120days between April 2020 and July, 2020. The data is normalized and for the ANN, a unipolar sigmoid function is applied to produce values between 1 and 0. For the states selected, only the observed cases, number of patients already discharged and number of deaths were used as input variables to the neural network, while the output is a single attribute representing the predicted number of cases. The implementation of the simulation was carried out using the R programming language with the predicted values written to various spreadsheets. These worksheets were exported to C# interface where the relative error values were computed and the output written to MySQL data tables.

2.1 General Form Of ARIMA(P,D,Q)

An AR(p) model is a discrete time linear equation with noise of the form:

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \varepsilon_t \quad (1)$$

Here, p is the order, $\alpha_1 \dots \alpha_p$ are the parameters or coefficients (real numbers), ε_t is an error term which is usually a white noise.

Where p = 1 then AR(1) becomes:

$$X_t = \alpha X_{t-1} + \varepsilon_t \quad (2)$$

With $|\alpha| < 1$ and $Var(X_t) = \frac{\alpha^2}{1-\alpha^2}$, it is a wide sense stationary process.

Since AR is a time series model, when we introduce the time lag operator (L) the AR becomes $Lx_t = x_{t-1}$, for all $t \in Z$ (set of real numbers). Since the time lag operator is a linear operator, the powers, positive or negative can be denoted as:

$$L^k L^k x_t = x_{t-k}, \text{ for all } t \in Z$$

With this lag operation, the AR model becomes:

$$X_t = \alpha X_{t-1} + \varepsilon_t$$

$$X_t - \alpha X_{t-1} = \varepsilon_t$$

The MA(q) model in ARIMA which denotes the moving average with orders (p and q) is an explicit formula for X_t in terms of noise of the form

$$X_t = c + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q} \quad (3)$$

Where:

$c =$ a constant from the regression modelled

$\varepsilon_t =$ the white noise in the past forecast errors

$\beta_1 \varepsilon_{t-1} =$ error in the AR values of the prediction

$\beta_q \varepsilon_{t-q} =$

last lag error in the AR used in the prediction

Combining the AR and the MA, we have:

$$X_t = c + \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \beta_1 \varepsilon_{t-1} + \dots + \beta_p \varepsilon_{t-p} + \varepsilon_t \quad (4)$$

This becomes the general form of ARIMA model as a discrete time linear equation.

2.2 Data Preparation for ARIMA

To process the data we used ARIMA(p,d,q) model. The steps involved in the use of ARIMA model are identification, parameter estimation and forecasting. These steps were repeated continually until an appropriate model is identified. The model to be selected is chosen based on the Akaike Information Criterion (AIC) [8] which is the model with the least AIC value. Based on the dataset, the data can also be made stationary by differencing before the training and forecasting.

2.3 Data Preparation for ANN

The general equation model for ANN is given as:

$$y_t = w_0 + \sum_{j=1}^q w_j f \left(w_{0,j} + \sum_{i=1}^p w_{i,j} y_{t-i} \right) + \varepsilon_t \quad (5)$$

Where:

y_t is the predicted value,

w_j are the output layer weights,

$w_{i,j}$ are input layer weights,

f is a transfer function,

q is the number of hidden nodes,

p is the number of input nodes,

ε_t is a random error at time t .

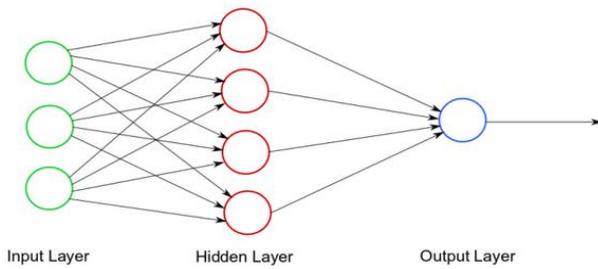


Figure 1: Overview of ANN (<https://learn.g2.com/artificial-neural-network>)

Out of the 120 days data collected for the selected states, we adopted 70% for training the algorithm, while the remaining 30% is used for testing. The 70% mapped out for training is further divided into two. While 40% is reserved for actual training, 30% is used for cross-validation. Before uploading the data into the ANN, the data is normalized. The need to normalize the data arises from the fact that the features to be used as input do not have a uniform scale. This experiment uses the min-max normalization function.

The unipolar sigmoid function is adopted and is defined as:

$$y = f(x) = \frac{1}{1 + e^{-\lambda x}} \quad (6)$$

It was used to normalize input vectors of the training data prior to processing and returns values between 0 and 1.

2.4 Validation and evaluation

The daily time series reports of confirmed cases of COVID19 from April 1, 2020 to July 30, 2020, were extracted from Nigeria Centre for Disease Control website and analyzed for the purpose of the research. This sample is made up of states that had confirmed cases from April after the index case was announced in March. For the ARIMA prediction, the times-series data for all the states were checked for stationarity using the Augmented Dickey Fuller `adf.test()` function. Each state data is differenced accordingly to arrive at a stationary set with differentials. It is evident that while some states became stationary at differences=1, others attained stationarity at difference = 2 while none was stationary at the initial stage. Ensuring that the data became stationary became necessary as a precondition for the use of use of ARIMA since the data must not have any form of trend.

Shown in Table 1 are the various models for the individual states along with their error values. The data series from the states were tested for ARIMA (0,1,1), ARIMA (2,1,1), ARIMA (1,2,1), ARIMA (2,2,1), ARIMA (0,1,2), ARIMA (0,2,2) and, ARIMA (0,2,1) models. For each model, the AIC value is computed while the model which produces the lowest AIC value is selected as the optimum model which was finally used for the prediction.

Table 1: ARIMA Models and their errors

States					
Model	Edo	River	Lagos	Osun	Ogun
0,1,1	1034.20	1039.70	1294.47	790.64	966.74
2,1,1	1028.16	1039.30	1292.90	792.75	965.98
1,2,1	1028.77	1037.30	1290.90	791.10	968.28
2,2,1	1028.16	1039.30	1292.90	792.75	965.96
0,1,2	1027.20	1037.39	1291.23	790.87	968.44
0,2,2	1027.19	1037.39	1291.23	790.87	968.44
0,2,1	1034.20	1039.70	1294.47	790.62	966.74
States					
Model	Ekiti	Kaduna	FCT	Oyo	Enugu
0,1,1	599.07	882.88	1083.40	1105.40	930.04
2,1,1	601.18	881.94	1085.80	1108.10	931.22
1,2,1	599.19	884.74	1085.40	1106.30	930.21
2,2,1	601.18	881.97	1085.80	1108.10	931.22
0,1,2	599.23	884.63	1085.40	1106.20	929.81
0,2,2	599.23	884.63	1085.40	1106.20	929.83
0,2,1	599.05	882.88	1083.40	1105.70	930.04

Table 2: ANN Models for various periods

States	120 days			
	3:2	3:1	4:1	4:2
River	0.0038	0.0055	0.0054	0.0057
Enugu	0.0025	0.0052	0.0053	0.0047
Ogun	0.0029	0.0025	0.013	0.0018
Kaduna	0.0036	0.0027	0.0072	0.0017
Lagos	0.0015	0.0012	0.0039	0.0038
FCT	0.0024	0.0038	0.0012	0.0038
Oyo	0.0024	0.0064	0.0047	0.0015
Osun	0.0068	0.0034	0.0039	0.0007
Ekiti	0.0027	0.0061	0.0016	0.0014
Edo	0.0026	0.0022	0.003	0.0029
States	90 days			
	3:2	3:1	4:1	4:2
River	0.0044	0.0036	0.0012	0.0023
Enugu	0.005	0.0055	0.0092	0.0012
Ogun	0.0125	0.0041	0.0008	0.0035
Kaduna	0.0023	0.0039	0.0036	0.0073
Lagos	0.0048	0.006	0.0189	0.0037
FCT	0.0069	0.0043	0.0104	0.0022
Oyo	0.0031	0.0021	0.0014	0.0009
Osun	0.0055	0.0127	0.0088	0.0054
Ekiti	0.0087	0.0088	0.0051	0.0042
Edo	0.0072	0.0103	0.0123	0.0058
States	60 days			
	3:2	3:1	4:1	4:2
River	0.0027	0.0054	0.0233	0.0010
Enugu	0.0058	0.0074	0.0149	0.0116
Ogun	0.0049	0.0100	0.0061	0.0030
Kaduna	0.0104	0.0043	0.0056	0.0020
Lagos	0.0036	0.0035	0.0011	0.0009
FCT	0.0067	0.0076	0.0217	0.0072
Oyo	0.0008	0.0045	0.0025	0.0013
Osun	0.0089	0.0239	0.0384	0.0099
Ekiti	0.0064	0.0080	0.0089	0.0044
Edo	0.0137	0.0010	0.0026	0.0079
States	30 days			
	3:2	3:1	4:1	4:2
River	0.0106	0.0081	0.0111	0.0078
Enugu	0.0010	0.4778	0.4534	0.4742

Ogun	0.0025	0.0081	0.0094	0.0017
Kaduna	0.0031	0.0064	0.0153	0.0021
Lagos	0.0051	0.0030	0.0047	0.0027
FCT	0.0060	0.0063	0.0061	0.0014
Oyo	0.0084	0.0202	0.0308	0.0095
Osun	0.0027	0.0010	0.0174	0.0107
Ekiti	0.0200	0.0167	0.0170	0.0146
Edo	0.0037	0.0063	0.0042	0.0045

Based on the ANN models (3,1), (3,2), (4,1), (4,2) and their errors as shown on Table 2, the model with the minimum error is selected for prediction. ANN(4:2) showed dominance over other models however. The unipolar sigmoid activation function was used for the mapping while min-max normalization was used to normalize the data.

The prediction performance is evaluated using various error valuation techniques. These include the Mean Absolute Percentage Error (MAPE) defined as follows:

$$MAPE = \frac{1}{T} \sum_{t=1}^T \left| \frac{A_t - F_t}{A_t} \right| \quad (7)$$

Where:

A_t = Actual Value

F_t = Forcasted Value

T = Number of times Summations.

3. Results

The charts in Figure 2 shows the plotting of the FCT & Lagos State 60-Days plots of the prediction errors of both ARIMA and ANN.

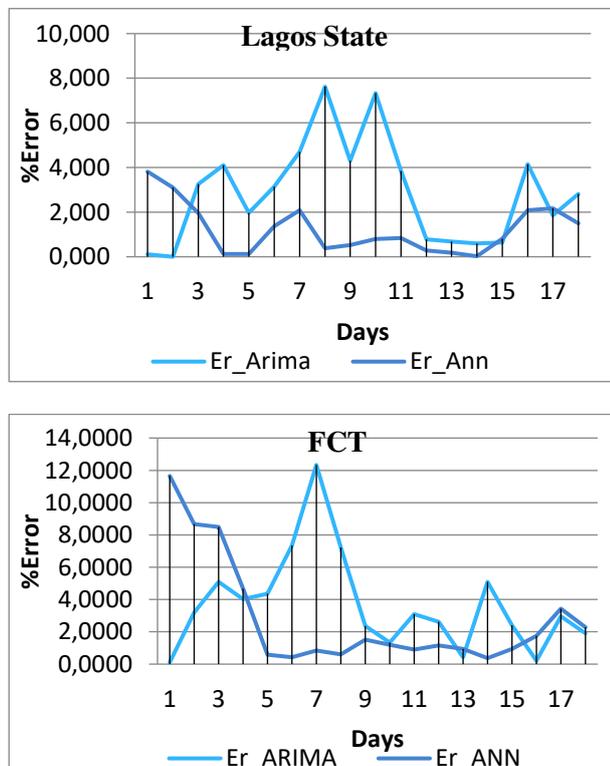


Figure 2: 60-Days Error Plots for ARIMA and ANN for FCT and Lagos States

For easy comparison, the 30% data used for the testing of ANN were also extracted from the predictions of ARIMA via the use of a C# programming module. This led to the presentation of 18-periods rather than the entire 30-days duration. The charts on figure 2 are the FCT and Lagos State 60-days error plots for both ARIMA and ANN. Presented in Table 3 is the summary of ARIMA versus ANN error values for all states modeled.

4. Discussion

For the prediction using ANN, considering each state and prediction period, the data was split into training and testing datasets consisting of 70% and 30% respectively. The input is made up of the number of patients on admission and the number already discharged.

Table 3: Summary of ARIMA & ANN error values for all states modeled

Days	State	Er_ARIMA	Er_ANN
30	Lagos	5.5082	2.3249
60		3.8885	2.9689
90		2.6091	3.6301
120		2.0348	9.6285
30	FCT	2.8343	5.6592
60		3.0526	1.3037
90		3.7146	2.6181
120		2.9871	1.9868
30	Kaduna	6.9396	13.9458
60		9.2587	16.5799
90		4.5329	2.5237
120		4.3819	3.9552
30	Enugu	0	4.5
60		10.7065	2.383
90		8.465	6.9796
120		5.8304	4.2215
30	Ogun	9.0758	15.8903
60		6.493	8.5536
90		6.1392	4.0318
120		5.0028	23.5906
30	Osun	0	25.4033
60		4.1779	1.8558
90		2.309	1.272
120		3.5205	3.6437
30	River	3.9375	20.6042
60		7.7425	14.2968
90		7.4306	30.2168
120		4.5729	27.7891
30	Oyo	6.9384	5.8372
60		3.7193	9.7848
90		5.4633	2.5101
120		2.8541	4.8947
30	Edo	5.8698	10.2801
60		7.2401	4.623
90		3.9818	5.1882
120		3.8242	7.2988
30	Ekiti	8.3333	6.8542
60		11.3526	7.999
90		13.0728	5.2975
120		5.1626	5.5611

From all the series collected it is evident that there is an upward trend in the number of confirmed cases of

COVID19 in Nigeria despite the fact that several safety measures have been put in place by the government. Lagos state has been on top of the list having recorded a total of over 19,600 cases followed by FCT with 5,758 cases. Comparing these figures with states like Zamfara, Yobe and Kogi with 79, 76 and 5 cases, one can easily be convinced that lack of enough testing facilities and experts must have influenced the low figures. Since different ARIMA(p,d,q) values were used for individual states depending on the nature of their data, it is difficult to state a universal model that fits all the states. The two states with the highest number of cases i.e Lagos State and FCT however performed incredibly well with ARIMA(1,2,1) and ARIMA(0,1,1) models respectively. On the 120-day period, the two models posted an accuracy level of 97.96% and 97.01% with is in line with the results obtained in the study by [9-11]. Though no particular ANN model produced the optimal error value across all the data modeled, it is seen that ANN(4:2) was predominant over other ANN models tested. On the 120-days period tested for Lagos State and FCT, the model produced an accuracy level of 90.37% and 98.01% respectively.

5. Conclusion

There is no known universal antiviral treatment for COVID-19 at the moment, and no vaccine has been certified by the WHO though there are various claims by different scientists of having one kind of cure or vaccine or the other available. The results from the study is a true representation of the trend in all the states of Nigeria including the Federal Capital Territory (FCT), considering the fact that that almost all geopolitical areas in Nigeria were represented with the exception of areas that do not have testing facilities during the early days of COVID-19. This shows that either of the two models (ARIMA & ANN) can be used to optimally predict the COVID-19 confirmed cases in Nigeria. We are of the opinion that ANN would have produced a better all-round result if additional input neurons were made up of more of the factors influencing positive confirmations. This might include number of laboratories in each state, number of people tested on daily basis which are not available at real time.

As the spread of the virus have continued to disrupt economic activities in the world over, the risks posed, such as restriction of traffic, closure of cities and continued closure of schools, will create a huge economic downturn and slowdown the welfare of the people. To ease the tension and the suffering of the masses, the government must intensity its effort in providing necessary financial incentives to the people during this period [12]. With the gradual relaxation of the lockdown in most cities, leading to opening up of markets, churches and schools, it becomes imperative that safety measures have to be put in place. We therefore advice that more proactive measures be taken by all stakeholders to forestall an outbreak which the country cannot contain considering the level of health facilities and experts available in the country. As the

lifting of stay-at-home and other restrictions continues to spread all through the states of Nigeria, it is important that the populace still continues to observe some degrees of safety measures. Adams et. al., (2020) [13] in a study found that there is a positive correlation between the number of deaths as a result to COVID-19 and the number of active and critical cases. This simply implies that it is not yet time for Nigerians to celebrate and off their guards as more cases are still being confirmed on a daily basis. It has been observed that people have abandoned the masks and continued to flock together in market places, religious houses, schools while social gatherings have all commenced even without strict adherence to COVID-19 protocols. In the future, other variables that could contribute to the forecasting accuracy of our models such as humidity, temperature, hygiene, sentiments among other variables can be included in the models.

Funding

The authors certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

Conflicts of interest

The authors declare no conflict of interest.

References

- [1] A. Muhammad, K. Suliman, K. Abeer, B. Nadia, S. Rabeea, COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses, *Journal of Advanced Research* 24 (2020) 91–98.
- [2] T. Tong Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV). *Perspectives in Medical Virology* (2006) 16:43-95, DOI:10.1016/s0168-7069(06)16004-8.
- [3] J. Amzata, K. Aminub, V. Kolob, A. Akinyeleb, J. Ogundairob, M. Dnjibo, Coronavirus outbreak in Nigeria: Burden and socio-medical response during the first 100 days, *International Journal of Infectious Diseases* 98 (2020) 218-224.
- [4] NCDC, Nigeria Centre for Disease Control. COVID-19 Outbreak in Nigeria Situation Report; NCDC: Abuja, Nigeria, 2020.
- [5] B. Jester, T. Uyeki, D. Jernigan, Readiness for Responding to a Severe Pandemic 100 Years After 1918, *American Journal of Epidemiology* 187(12) 2596-2602.
- [6] I. Oladipo, A. Babatunde, Data Mining Classification Techniques for the Diagnosis, Treatment and Management of Diabetes Mellitus: A Review., *Proceedings of the 1St International Conference of IEEE Nigeria Computer Chapter In collaboration with Department of Computer Science, University of Ilorin, Ilorin, Nigeria – 2016.*
- [7] B. Kaplan CDC Reveals Top 5 Causes of Death in the U.S., <https://www.orlandohealth.com/content-hub/cdc-reveals-top-5-causes-of-death-in-the-us> (accessed 13 August 2020).
- [8] H. Akaike, Information Theory and an Extension of the Maximum Likelihood Principle. In: B.N. Petrov and F.

- Csaki (eds.) 2nd International Symposium on Information Theory: (1793) 267-81 Budapest:Akademiai Kiado.
- [9] S. Roy, G.S. Bhunia, P.K. Shit, Spatial prediction of COVID-19 epidemic using ARIMA techniques in India. *Model. Earth Syst. Environ* (2020), doi.org/10.1007/s40808-020-00890-y .
- [10] Q. Yang, J. Wang, H. Ma, Research on COVID-19 based on ARIMA model—Taking Hubei, China as an example to see the epidemic in Italy, *Journal of Infection and Public Health*, (2020), Volume 13, Issue 10 Pages 1415-1418.
- [11] D. Benvenuto, M. Giovanetti, L. Vassallo, Angeletti, M. Ciccozzi, Application of the ARIMA model on the COVID-2019 (2020), *Epidemic Dataset, Data in Brief*, 29 (2020) 105340.
- [12] A. Oyelola, I. Adeshina, E. Gayawan Early Transmission Dynamics of Novel Coronavirus (COVID-19) in Nigeria, (2020). *International Journal of Environmental Research and Public Health*, 17 (2020) 3054 .
- [13] O.S. Adams, A.M. Bamanga, U.H. Yahaya, O. R. Akano, Modeling COVID-19 Cases in Nigeria Using Some Selected Count Data Regression Models, (2020), *International Journal of Healthcare and Medical Sciences*, ISSN(e): 2414-2999, ISSN(p): 2415-5233, Vol. 6, Issue. 4 (2020) 64-73.