***ORIGINAL RESEARCH ARTICLE***

# Attributable Fraction and Forecasting for Covid-19 Confirmed Cases in Nigeria Using Facebook-Prophet Machine Learning Model




## ABSTRACT

**Aims:** The motivation is to know the attributable fraction among Nigerians, who were tested positive for covid-19 and forecast the covid-19 cases.

**Place and Duration of Study:** We extracted data from (https://covid19.ncdc.gov.ng/) on 8th September,2021 and covid.19analytics package on 7th September, 2021, from Data Repository by Johns Hopkins University Center for Systems Science and Engineering , Status of Cases in Toronto – City of Toronto , COVID-19: Open Data Toronto ,COVID-19: Health Canada , Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, COViD-19 Vaccination and Testing records from "Our World In Data" and Pandemics historical records from Visual Capitalist. Data in Nigeria contained the number of samples tested, confirmed cases, active cases, discharged cases and deaths.

**Methodology:** Attributable fraction was used to compute the proportion of patients who tested positive to Covid-19. By using the time for regressor, Prophet model will fit many non-linear and linear functions of time components. Prophet uses the Fourier series to get flexible model to forecast and fit the seasonality effects. A fast solution for L-BFGS which stands for Limited memory Broyden-Fletcher-Goldfarb-Shannon algorithm, is used with Stan backend for the prediction problem.

**Results:** As at Saturday 11th September 2021,7:18am Nigeria local time, a total of 2884034 samples have been tested for covid-19, with 198239 confirmed cases,9871 active cases,185780 discharged cases and 2588 deaths. The attributable fraction for covid-19 in Nigeria was 0.0687. The r square is very high (0.999), and the p value is very low (2.2e-16).

**Conclusion:** The attributable fraction gives the percentage of the patients who tested positive to covid-19, among the 2884034 samples tested. It implies that the remaining percentage of patients who tested negative to covid-19 only exhibit covid-19 symptoms or were exposed to the virus. The confirmed cases were found to be highest on Saturdays with the lowest on Tuesdays.


*Keywords: Forecasting, Covid-19, Attributable Fraction, Facebook-Prophet, Machine Learning*

## 1. INTRODUCTION

Covid -19 has become a big challenge since the outbreak with many countries still battling with the deadly virus. This transmittable and infectious disease is caused by a recent coronavirus [1]. Researchers have studied the trend, pattern

and predicted the covid-19 cases using different methods [2]. A forecast model that uses improved version of adaptive neuro-fuzzy inference system (ANFIS) was developed by [3]. [4] created a mathematical model for covid-19 built on the practical assumptions and formed an autonomous system with first-order differential equation. [5] applied deep learning methods, these methods can understand structures, trends, dependencies, and structures. Method like multi-layer perception can be used for multivariate inputs and multi-step prediction. Machine learning was used, and prophet logistic growth model was used for the prediction [6].

**Comment [DSSA1]:** Coomment No.1 :Repeated,so delete the word « structure»

[7] developed a model called susceptible-exposed -infectious -recovered type epidemiological, to monitor the state of the virus. Parameters that can be interpreted were proposed with a modular regression model, autoregressive integrated moving average was employed because the model is simple [8]. [9] and [10] applied ARIMA, but [10] in addition applied Prophet package which is based on an additive regression time series forecasting algorithm developed by Facebook was used for the forecasting in addition to ARIMA [11] stated that coronavirus disease is caused by acute respiratory syndrome (SARS-CoV-2), [12] also applied autoregressive integrated moving ~~average.Inaverage.~~ In the study of covid-19, [13] examined the relationship between the confirmed, death and recovered was assumed to be linear. [14] used autoregressive model of order two after first difference to model cocoa production. The attributable fraction and predisposing factors were modeled using logistic regression by [15].

The motivation is to know the percentage of patients that had the diseases after being tested, because a lot of people assume that they have covid-19 when they have the symptoms associated with the virus, this is not true this is because the symptoms of covid-19 are common to other diseases. People have been told to always go for test if they exhibit the symptoms of covid-19. We need to know the covid-19 attributable fraction after being tested. There is need to forecast the covid-19 cases, this information will really help to plan for the deadly virus. In this study, attributable fraction for covid-19 in the case of Nigeria was computed and predictions were made.

## 2. MATERIAL AND METHODS

### Attributable Fraction

This is the proportion of incidents found in the population that are attributable to the risk factor.

### Prophet Model

Facebook built an additive regression time series, it is an algorithm for forecasting [6] and [16]. To apply it, a fast solution for L-BFGS which stands for Limited memory Broyden-Fletcher-Goldfarb-Shannon algorithm, can be used with Stan backend for the prediction problem.

The prophet model is given by $y(t) = g(t) + s(t) + h(t) + \varepsilon_i$ (1)

Where $g(t)$ denotes trend (linear or logistic growth curve for developing non-periodic variations find time series), $s(t)$ is the seasonal changes (weekly or yearly seasonality), $h(t)$ is the irregular effects(holidays) and $\varepsilon_i$ is the error term for unusual change in the model.

By using the time for regressor, Prophet will fit many non-linear and linear functions of time components.

**Trend**

The trend will be developed and fit a piece wise linear curve, over the trend or the non-periodic portion of the time series. With this, we can take care of the missing data or anything that can affect the spike. Prophet uses the Fourier series to get flexible model, this is to forecast and fit the seasonality effects. These effects of seasonality are estimated by

$$s(t) = \sum_{n=1}^{N}(a_n cos\left(\frac{2\pi nt}{P}\right) + b_n sin\left(\frac{2\pi nt}{T}\right)) \quad (2)$$

P is the ~~period,and~~period, and $[a_1, b_1, \ldots, a_N, b_N]$ will be estimated.

**Description and Sources of Data**

The data were extracted from (https://covid19.ncdc.gov.ng/) on 8th September,2021 and covid.19analytics package on 7th September, 2021.This package allows us to have access to live data anywhere in the world, from Novel Corona Virus COViD-19 (2019-nCoV) Data Repository by Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) , Status of Cases in Toronto – City of Toronto , COVID-19: Open Data Toronto ,COVID-19: Health Canada , Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, COViD-19 Vaccination and Testing records from “Our World In Data” (OWID) and Pandemics historical records from Visual Capitalist.

Data on 36 states and the capital of Nigeria (Abuja) were extracted, the available data contained the number of samples tested, confirmed cases, active cases, discharged cases and deaths.

## 3. RESULTS AND DISCUSSION

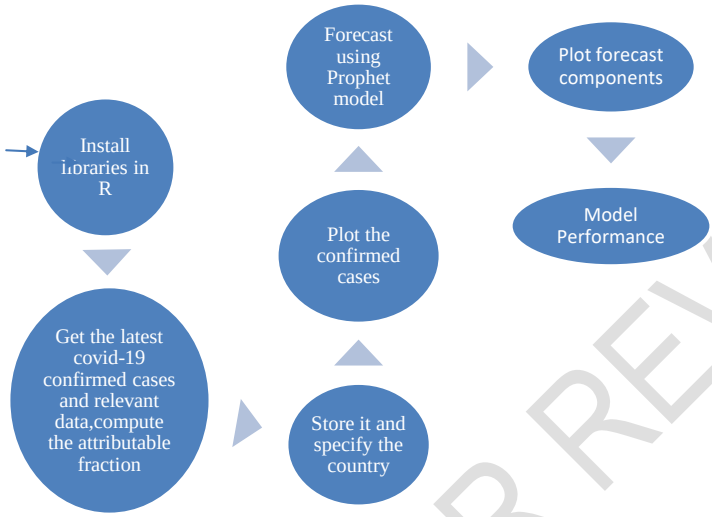


**Fig. 1. Schematic representation of the procedures.** This explains the procedures used for the analysis, after the relevant installation of R packages, the diagram explains how to carry out the research

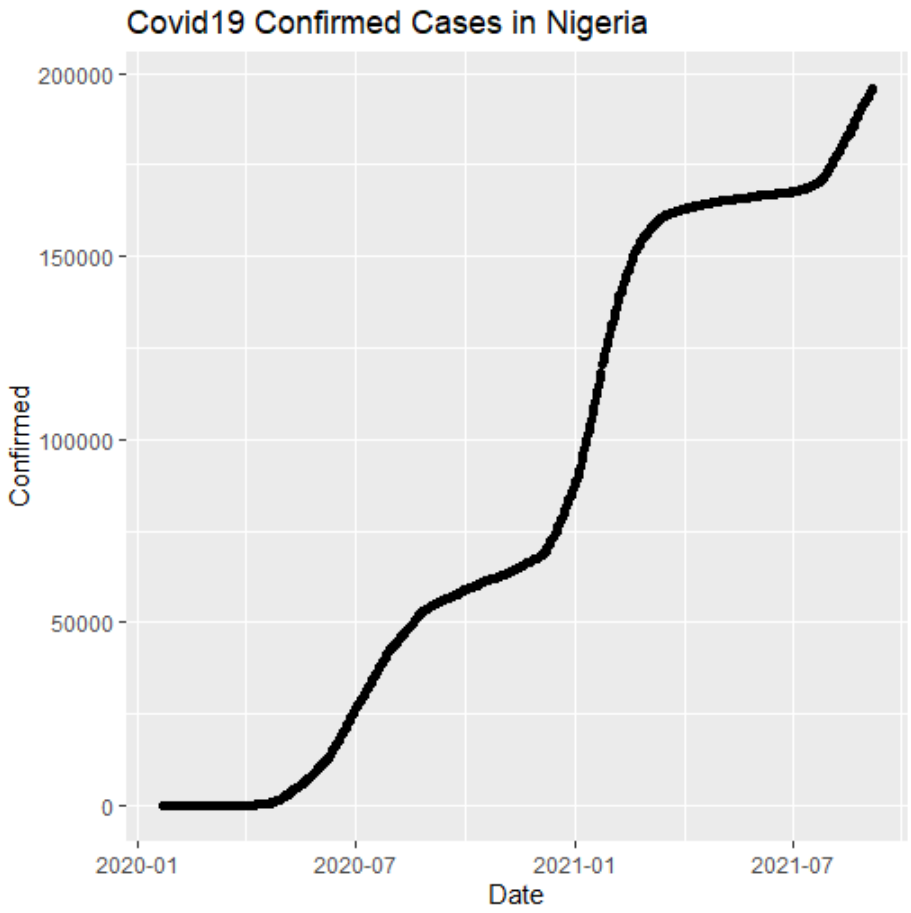


**Fig. 2. Confirmed Cases for Covid-19 in Nigeria.** The graph shows the covid-19 confirmed cases from 28th February, 2020 to 7th September,2021.

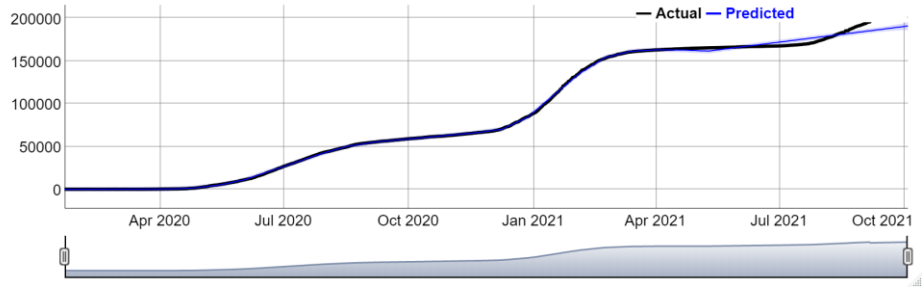


**Fig. 3. Actual confirmed cases versus Predicted cases.** The graph shows the actual confirmed cases against the predicted confirmed cases for the next 28 days, the predicted confirmed cases are the ones in blue.

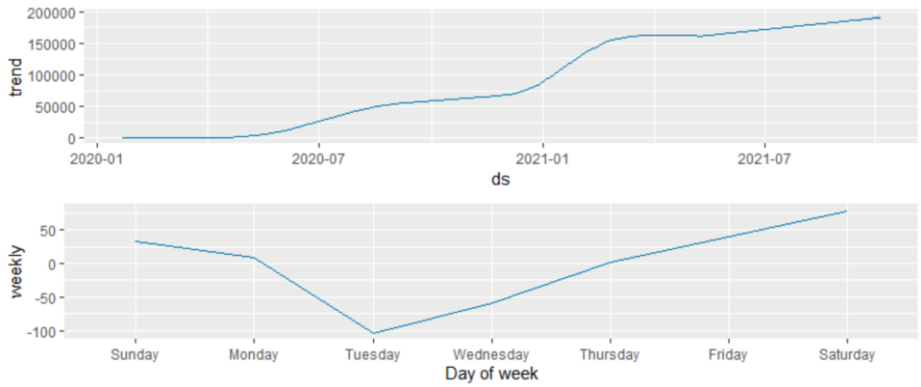


**Fig. 4. Plots of the Forecasts Components.** This graph shows the weekly trend, we can know which day(s) the week has the highest or lowest Covid-19 confirmed cases.

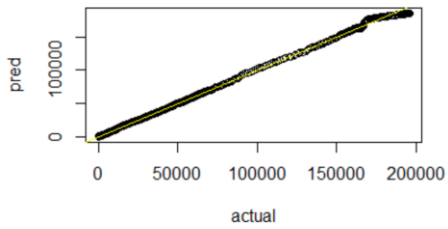


**Fig. 5. Plot of Predicted versus Actual for the model performance.** This graph shows how well the model has performed.

**DISCUSSION**

The new confirmed case on the 10th of September 2021 was 466 and 3 deaths were recorded in Nigeria. The 466 new occurred in 13 states- Lagos (134), Rivers (82), Edo (69), Gombe (39), FCT (32), Kaduna (21), Plateau (20), Benue (19), Kwara (17), Delta (16), Akwa Ibom (10), Bayelsa (5) and Kano (2). As at Saturday 11$^{th}$ September 2021,7:18am Nigeria local time, a total of 2884034 samples have been tested for covid-19, with 198239 confirmed cases,9871 active cases,185780 discharged cases and 2588 deaths. The attributable fraction for covid-19 in Nigeria was 0.0687, with 95% confidence interval between 68.40% and 68.99%, we were 95% confident that the patients who had covid-19 lies in this range. From figure 2, the confirmed cases column is cumulative numbers, which can give us the sum at any date, we were able to disable yearly and daily seasonality because we do not have enough data to capture the seasonality. Figure 3 shows the predicted the covid-19 confirmed cases for the next 28 days from 7$^{th}$ September 2021 and the conditions are valid if the current conditions did not change. The confidence bound was also included because it can be lower or higher. From figure 4, the average, the confirmed cases were found to be lowest on Tuesdays, there were fluctuations in the confirm cases, it may be due to under or over reporting of covid-19 cases or the procedures for being tested for covid-19 cases. We have many confirmed cases on Sunday, Friday, and Saturday. This does not imply that Nigerians have a higher risk of getting covid-19 on these days. Figure 5 shows a linear pattern, and no many under or over estimation, the r square is very high (0.999), and the p value is very low (2.2e-16), which means we have a high confidence that it is statistically significant.

## 4. CONCLUSION

The attributable fraction for covid-19 confirmed cases in Nigeria is 68.7%. This gives the percentage of the patients who tested positive to covid-19, among the 2884034 samples tested. We can conclude that the remaining percentage of patients who tested negative to covid-19 only exhibit covid-19 symptoms or were exposed to the virus. The confirmed cases were found to be highest on Saturdays with the lowest on Tuesdays, there is need to investigate further the reason.

## CONSENT

Not Applicable

## ETHICAL APPROVAL

Not Applicable

## REFERENCES


(1) Joshua IO. Understanding COVID 19 attributable fraction in Nigeria. Curr Tre Biosta & Biometr. 2020 Jun 22;2(5):276-7.

(2) Sarkar K, Khajanchi S, Nieto JJ. Modeling and forecasting the COVID-19 pandemic in India. Chaos, Solitons & Fractals. 2020 Oct 1;139:110049.

(3) Al-qaness MAA, Saba AI, Elsheikh AH, Elaziz, MA, Ibrahim RA, Lu S, Hemedan AA et al. Efficient artificial intelligence forecasting models for COVID-19 outbreak in Russia and Brazil. *Process Safety and Environmental Protection*. https://doi.org/10.1016/j.psep.2020.11.007. 2021;149,399-409

(4) Battineni G, Chintalapudi N, Amenta F. Forecasting of COVID-19 epidemic size in four high hitting nations (USA, Brazil, India and Russia) by Fb-Prophet machine learning model. Applied Computing and Informatics. 2020 Dec 10.

(5) Devaraj J, Elavarasan RM, Pugazhendhi R, Shafiullah GM, Ganesan S, Jeysree AK, Khan IA, Hossain E. Forecasting of COVID-19 cases using deep learning models: Is it reliable and practically significant?. Results in Physics. 2021 Feb 1;21:103817.

(6) Taylor SJ, Letham B. Forecasting at scale. The American Statistician. 2018 Jan 2;72(1):37-45.

(7) Hamzah FB, Lau C, Nazri H, Ligot DV, Lee G, Tan CL, Shaib MK, Zaidon UH, Abdullah A, Chung MH. CoronaTracker: worldwide COVID-19 outbreak data analysis and prediction. Bull World Health Organ. 2020 Mar 19;1(32).

(8) Gaur S. Global forecasting of covid-19 using ARIMA based FB-Prophet. International Journal of Engineering Applied Sciences and Technology. 2020;5(2):463-7.

(9) Ibrahim RR, Oladipo OH. Forecasting the spread of COVID-19 in Nigeria using Box-Jenkins modeling procedure. medRxiv. 2020 Jan 1.

(10) Abioye AI, Umoh MD, Peter OJ, Edogbanya HO, Oguntolu FA, Kayode O et al. Forecasting of COVID-19 pandemic in Nigeria using real statistical data. *Communications in Mathematical Biology and Neuroscience*, https://doi.org/10.28919/cmbn/5144. *2021*, 1–12.

(11) Kenneth GE, Maxwell O, Amechi SC. COVID-19 Outbreak: An Asthmatic Asymptomatic. Journal of Applied Medical Sciences. 2020;9(2):29-34.

(12) Kumar N, Susan S. Covid-19 pandemic prediction using time series forecasting models. In2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) 2020 Jul 1 (pp. 1-7). IEEE.

(13) Hoseinpour Dehkordi A, Alizadeh M, Derakhshan P, Babazadeh P, Jahandideh A. Understanding epidemic data and statistics: A case study of COVID-19. Journal of medical virology. 2020 Jul;92(7):868-82.

(14) Ajare EO, Ibidoja OJ. Autoregressive Model for Cocoa Production in Nigeria. International Journal of Science for Global Sustainability. 2016 Dec 28;2(4).
(15) Ibidoja OJ, Ajare EO, Fowobaje KR, Jolayemi ET. Malaria Attributable Fraction and some of its Predisposing Factors. International Journal of Science for Global Sustainability. 2016 Sep 26;2(3).
(16) Abdulmajeed K, Adeleke M, Popoola L. Online forecasting of COVID-19 cases in Nigeria using limited data. Data in Brief. 2020 Jun 1;30:105683.