

# Forecasting the Stability of COVID-19 on Indian Dataset with Prophet Logistic Growth Model

Nishtha Phutela

BML Munjal University

Arushi G Bakshi

ByteDance

Sunil Gupta (✉ [s.gupta@ddn.upes.ac.in](mailto:s.gupta@ddn.upes.ac.in))

University of Petroleum and Energy Studies <https://orcid.org/0000-0003-3862-1980>

Goldie Gabrani

BML Munjal University

---

## Research Article

**Keywords:** COVID-19 Pandemic, Prophet Model, Prediction, Logistic curve.

**Posted Date:** June 2nd, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-32472/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Recently COVID-2019, a highly infectious disease has been declared as Pandemic by WHO, and since then the researchers all over the world are making attempts to predict the likely progression of this pandemic using various mathematical models. In this paper, we are using logistic growth model to find out the stability of this pandemic and Prophet Model to forecast the total number of confirmed cases that would be caused by COVID-19 in India.

## Specification Table

Subject	Decision Sciences
Specific subject area	The data prediction is focus on stability of confirmed cases caused by COVID-19 in India using Prophet Logistic S-shaped Curve.
Type of data	Table Graph Figure
How data were acquired	Coronavirus Source Data: <a href="https://covid.ourworldindata.org/data/owid-covid-data.csv">https://covid.ourworldindata.org/data/owid-covid-data.csv</a>
Data format	The raw data is collected in CSV file, which is cleaned for prediction and analysis. The CSV file has been uploaded.
Parameters for data collection	The total number of confirmed cases in India for the period from January 30th, 2020 to April 29th, 2020 were collected and used to find the corresponding carrying capacity in Prophet Logistic Growth Model.
Description of data collection	Daily data on COVID-2019 for India during the above-mentioned period was collected from the hyperlinks given at the official website on COVID-19: <a href="https://www.mygov.in/corona-data">https://www.mygov.in/corona-data</a> <a href="https://ourworldindata.org/coronavirus-source-data">https://ourworldindata.org/coronavirus-source-data</a> <u>Two different website links were used to maintain the authenticity in the data, and prediction. It will show the exact figure in term of stabilizing the curve.</u>  Analysis of the data was done to evaluate the number of new confirmed cases of COVID-2019.
Data source location	Oxford Martin School, University of Oxford, UK.
Data accessibility	The raw data for this is taken from GitHub repository of coronavirus source directory. <a href="https://covid.ourworldindata.org/data/owid-covid-data.csv">https://covid.ourworldindata.org/data/owid-covid-data.csv</a>

## Value Of The Data

- This data offers a new approach to the scientific community, researchers, and academicians to forecast that when a pandemic is likely to stabilize.
- This data gives valuable insights for the progression of COVID-19, hence the organizations dealing with infectious diseases can use this data for early prevention, to monitor, manage and control the spread of pandemic.
- This data supports decision-makers working for different domains of society like nation's leaders determining what policies to implement, pharmaceutical interventions, environmental impact and effect on economy.

## Introduction To Data Description

The data of COVID-2019 from December 31<sup>st</sup> 2019 to May 19<sup>th</sup> 2020 was collected from official website <https://ourworldindata.org/coronavirus-source-data>, the data is available in form of CSV file [1]. The dataset consisted of 19 features. Out of these, the feature 'total confirmed cases' was forecasted using Prophet Model. These values demonstrated an increasing trend over time. Hence, these were treated independently as univariate time series and a forecast was made for the number of confirmed cases after 19<sup>th</sup> May. Figure 1 shows the increase in the actual number of confirmed cases over time as well as the number of confirmed cases as predicted from the logistic growth model. This plot has been made using the parameters mentioned in Table 1. Here a is a hyperparameter (hp), b is the multiplicative factor (mf) and c is the maximum number of people who might get infected by this virus, which is carrying capacity in our case(cc). Figure 2 shows the actual number of confirmed cases as well as the forecasted confirmed cases. We have shown that after 10<sup>th</sup> July, the growth of confirmed cases will reach a stable point (refer Table 6).

Table 1: Parameter Computing using Nonlinear Least Square Estimation Method

a (hp)	b(mf)	c(cc)
11331.051527372085	0.08525654103333764	194401.71436567788

## Experimental Design, Materials, And Methods

We have used logistic curve to model the spread of COVID-19. A Prophet logistic curve is a Sigmoidal-S-shaped curve used to model functions which can be divided into three phases viz. (i) first phase: increases gradually during the start period, (ii) second phase: grows more quickly in the intermediate period and (iii) third phase: gradually at the end period, reaching a saturation level (referred to as carrying capacity) [2][3][4]. The Sigmoidal curve of this type is often used to find biological growth patterns which exhibit exponential growth period to start with and subsequently reaches a maximum level. The pattern followed by COVID-19 is similar to this curve as large number of people get infection in early stages and further growth subsequently slows down due to implementation of various preventive measures.

Prophet model gives very good outcomes when data shows either non-linear trends or non-linear growth patterns. Fundamentally, Prophet uses an additive regression concept, which can simply be modelled as a piecewise linear or as a logistic growth curve [5]. We tried using both these options and found that the logistic curve trend gives much better estimations to model the spread of COVID-19. Even intuitively, modelling the COVID-19 spread through logistic method is more realistic. The motivation of this analysis is to identify if the COVID-19 in India is still growing exponentially or the growth had started following a flat curve [6]. It is necessary to fit India's data to the logistic model because if any country is in the initial part of the curve, it indicates that the COVID-19 is still growing at a fast rate whereas if it is in the second part of the curve, it implies that growth is towards its saturation point. To ascertain the same, growth of the COVID-19 from the available data is modelled with logistic function and forecast of the growth over the next few days is done with Prophet.

### **Implementation Steps:**

a. The data frame used in Prophet needs two columns-(i) ds: to store date time series and (ii) y: to store the corresponding values of the time series in the data frame. The parameter carrying capacity (c) represents the maximum number of infections that can be caused by the virus or the saturation point of virus growth which is estimated via logistic model using the equation:

$$a. \quad y(t) = c / (1+a*e^{-bt})$$

b. demonstrate COVID-19 spread using this equation three parameters a, b and c are computed using Nonlinear Least Squares Estimation method by using Scipy Curve fit optimization library of python. The logistic curve is then fit to the dataset using these parameters and it is then compared to the actual values of confirmed cases at each time instant (refer Table 1 and Figure 1). To

c. Now before feeding dataset to the prophet model to forecast growth, we need to find the estimated carrying capacity, c which is computed as follows:

Rule 1: When the fastest growth day is still ahead, it implies that the growth is still increasing. In this case add ten days after finding the fastest growth day.

Rule 2: When the fastest growth day is in the past, it means that growth has stabilized. In this case use the current day and add ten days to find the estimated highest number of infections.

d. The fastest growth day identified through our logistic model is 18<sup>th</sup> May which is in the past as we have analyzed data till 19<sup>th</sup> May, therefore rule 2 is applicable to our model. The parameter c is 71436567788, which means the maximum limit for the number of infections in India would be 194401.71436567788. Prophet model was the fit using this value of c (194401.71436567788 as shown by the horizontal line at the top (refer Figure 2)). From the forecast curve (Figure 2) it can be inferred that the growth of the COVID-19 in India is expected to stabilize after July 10<sup>th</sup> (refer Table 2).

Table 2: Growth Stabilization of COVID-19 in India

Carrying Capacity	Expected Stabilization Period
194401.71436567788	After July 10 <sup>th</sup>

In order to evaluate the model-fit, we used cross-validation. The first case of COVID-19 was confirmed in India on 30<sup>th</sup> January, and the date of data collection is 19<sup>th</sup> May, so we had 110 days data. The initial 100 days (train data) were used to train the model and the remaining 10 days (test data) were used to evaluate the model (refer Table 3). The outcomes obtained are shown in Table 4.

Table 3: Parameters of Cross Validation

Initial	Period	Horizon
100 days	3 days	10 days

Predicted vs actual with lower and upper Confidence Interval (Table 4), the data up to 9<sup>th</sup> May was considered for training hence the cutoff date is 2020-05-09. The effects of the error diagnostics of the model are shown in Table 5. Minimum error is reported at a horizon of 8 days, meaning thereby that the number of confirmed cases can be more accurately predicted by looking at 8 days in the past.

Table 4: Cross validation: Predicted (yhat) vs. Actual (y) values of Confirmed Cases

ds	yhat	yhat_lower	yhat_upper	y	cutoff
0 2020-05-10	63471.235561	62377.885236	64548.208157	62939	2020-05-09
1 2020-05-11	67163.717254	66019.337365	68221.623076	67152	2020-05-09
2 2020-05-12	70954.393625	69872.823539	72003.486816	70756	2020-05-09
3 2020-05-13	74809.373630	73731.487633	75764.416473	74281	2020-05-09
4 2020-05-14	78744.944047	77679.830223	79873.291466	78003	2020-05-09

Table 5: Error Diagnostics of the Confirmed Cases

horizon	mse	rmse	mae	mape	coverage
0 1 days	2.832747e+05	532.235561	532.235561	0.008456	1.0
1 2 days	1.372940e+02	11.717254	11.717254	0.000174	1.0
2 3 days	3.936003e+04	198.393625	198.393625	0.002804	1.0
3 4 days	2.791787e+05	528.373630	528.373630	0.007113	1.0
4 5 days	5.504810e+05	741.944047	741.944047	0.009512	1.0
5 6 days	5.782167e+05	760.405647	760.405647	0.009277	1.0
6 7 days	6.984032e+05	835.705200	835.705200	0.009724	1.0
7 8 days	3.446930e+01	5.871056	5.871056	0.000065	1.0
8 9 days	1.251006e+06	1118.483742	1118.483742	0.011630	0.0
9 10 days	3.806742e+06	1951.087498	1951.087498	0.019291	0.0

## Significant Inferences from Data Analysis:

The lockdown 4.0 is till May 31<sup>st</sup>. The confirmed cases data has fit a logistic distribution. This data has been collected in a controlled environment (lockdown, social distancing etc.). From the logistic curve it can be inferred that India is in the second part of the logistic growth curve therefore the number of confirmed cases is going to stabilize soon (provided the controlled conditions sustain). Therefore, the preventive measures have been successful in controlled setup. The maximum people that can be infected by this virus (as inferred from the logistic model) is 194401.71436567788. This means that if the controlled conditions continue to prevail, the growth of the COVID-19 will become stable after 10<sup>th</sup> July and the confirmed cases will come close to the carrying capacity. Also, evident from the figures below, the confirmed cases on May 31<sup>st</sup> will be 1,44,335 (refer Table 6) which is much less than the carrying capacity of logistic (194401.71436567788), therefore the strict preventive measures should be continued at least till 10<sup>th</sup> July.

Table 6: Forecast of Confirmed Cases in India

<b>ds</b>	<b>yhat</b>	<b>yhat_lower</b>	<b>yhat_upper</b>
2020-05-20	103235.600307	102129.776456	104270.847174
2020-05-21	107313.841507	106247.197462	108389.310530
2020-05-22	111352.038596	110258.201211	112480.461904
2020-05-23	115345.353552	114244.679963	116461.425297
2020-05-24	119361.977780	118282.030369	120475.057813
2020-05-25	123352.736953	122261.982145	124425.779672
2020-05-26	127184.675683	126078.611273	128202.811087
2020-05-27	130714.141752	129593.929953	131815.502842
2020-05-28	134275.602206	133206.843681	135397.311858
2020-05-29	137718.900756	136613.884397	138853.724724
2020-05-30	141046.637487	140033.698549	142223.147066
2020-05-31	144335.028957	143237.983242	145378.998661

## Conclusion

In this paper, the data provides a simple way to understand the spread of COVID-19 and to authenticate whether various preventive measures taken like social distancing, quarantine, contact tracing are effective or not. Author have forecasted the confirmed cases caused due to coronavirus in India. Figure 2 shows forecast curve of confirmed cases. We have shown that after 10<sup>th</sup> July, the growth of confirmed cases will reach a stable point.

## Declarations

### Competing Interests

The authors confirm that they have no known involvement in any organization and other entity with any financial interest. The authors also declare that they have no known personal relation that could influence

the work done in the paper.

## References

1. Coronavirus Source Data, Oxford Martin School, 2020. <https://covid.ourworldindata.org/data/owid-covid-data.csv>
2. [https://facebook.github.io/prophet/docs/quick\\_start.html](https://facebook.github.io/prophet/docs/quick_start.html)
3. <https://towardsdatascience.com/forecasting-with-prophet-d50bbfe95f91>
4. <https://www.analyticsvidhya.com/blog/2018/05/generate-accurate-forecasts-facebook-prophet-python-r/>
5. Sean J. Taylor, Benjamin Letham, *Forecasting at Scale*, Facebook, Menlo Park, California, United States <https://peerj.com/preprints/3190.pdf>
6. Işıl Yenidoğan, Aykut Çayır, Ozan Kozan, Tuğçe Dağ, Çiğdem Arslan, Bitcoin Forecasting Using ARIMA and PROPHET, 2018 3rd International Conference on Computer Science and Engineering (UBMK). DOI: 10.1109/UBMK.2018.8566476

## Figures

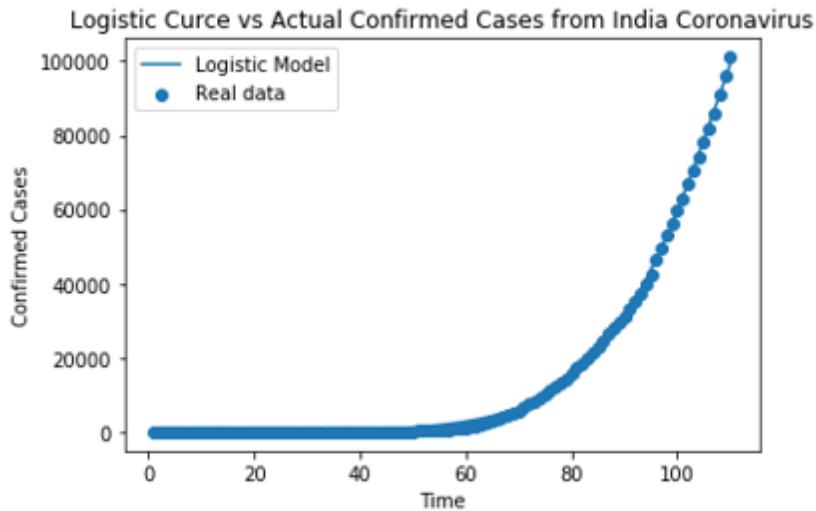
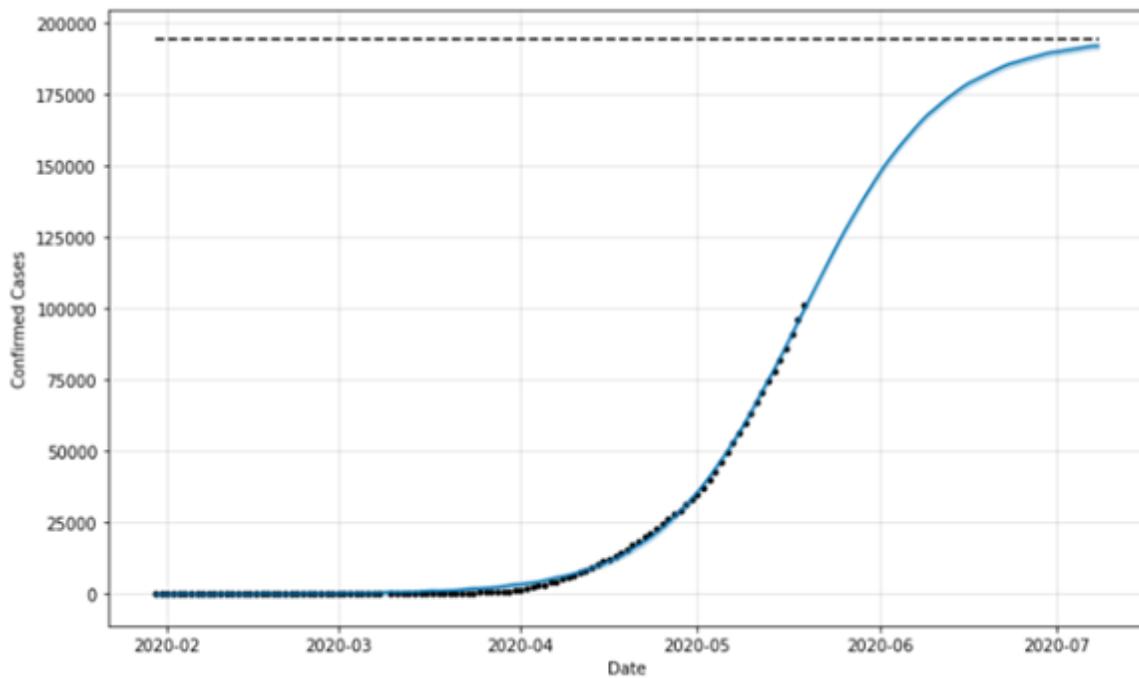


Figure 1

Logistic Curve of Confirmed Cases



**Figure 2**

Forecast Curve of Confirmed Cases