International Journal of Statistics and Applied Mathematics 2020; 5(4): 103-106

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2020; 5(4): 103-106

www.mathsjournal.com
Received: 17-05-2020
Accepted: 20-06-2020

Indhuja M
Department of Statistics,
PSG College of Arts and Science
Coimbatore, Tamil Nadu, India

Sindhuja PP
Department of Statistics,
PSG College of Arts and Science
Coimbatore, Tamil Nadu, India

Corresponding Author:
Indhuja M
Department of Statistics,
PSG College of Arts and Science
Coimbatore, Tamil Nadu, India

Prediction of covid-19 cases in India using prophet

Indhuja M and Sindhuja PP

Abstract
The substantial spike in daily new cases infected with the novel Coronavirus (COVID-19) worldwide is alarming and many researchers are now carrying out several analyses to forecast the future pattern of this pandemic. In this study, we have forecasted the expected daily number of COVID-19 cases in India for the next four weeks using Prophet. The results have demonstrated that by the end of the second week of August, the rate of proliferation of COVID-19 might reach upto 1,500,000 cases per day in India. It is observed that the performed model of fit is accurate within a certain range, and extreme preventive and control measures are suggested in an effort to avoid such a situation.

Keywords: COVID-19, Prophet, India

1. Introduction
In the last weeks of 2019, many local hospitals in Wuhan, China, have reported an unusual number of patients admitted with severe pneumonia without knowing the cause and did not respond to any form of vaccine or medicine [5]. The agent causing the viral-related disease was identified as SARS-CoV-2 because of its similarity to SARS CoV [9]. In addition, these cases increased exponentially, primarily because of human to human transmission, and became a global pandemic. Currently there is no scientific evidence to substantiate the origin of the virus. The rate of contagion and infection spread is expeditious when compared with precedent viral infections.

Due to rapid proliferation of COVID-19, many countries including India have announced complete national lock down. As a result, nearly 3.5 billion of the world's population have gone into self-isolation [5]. The World Health Organization (WHO) confirms that the incubation period for COVID-19 outbreak is 14 days. The basic reproduction number "R naught" or R0 is a contagiousness indicator or infectious transmissibility of parasite agents [8]. In epidemic sciences and health literature, R0 is highly encountered to understand a slow outbreak of disease. For instance, When R0 equals one, it implies that an average person affected by the disease could transmit it over a single individual. According to WHO, R0 for COVID-19 is confirmed to be around 2.0-2.5 [8]. For India, R0 was calculated to be about 1.83 in April. The R value went up to 1.19 on 7 July from 1.11 around 26 June, just a week after 'Unlock 2' came in [13].

India currently has the largest number of confirmed cases in Asia, and has the third highest number of confirmed cases in the world with the number of total reported cases crossing the 100,000 mark on 19 May and 200,000 on 3 June. As of 14 July 2020, the Ministry of Health and Family Welfare (MoHFW) has confirmed a total of 906,752, 571,459 (including 1 migration) and 23,727 deaths in the country [6]. Although the recovery rate is north of 50%, active cases have continued to burgeon. Due to the rapid spread of the virus in a short time period, it is necessary to carry out intensive studies of the same. Subsequently, for this, it is important to precisely monitor and glean data regarding cases in order to increase the accuracy of future predictions.

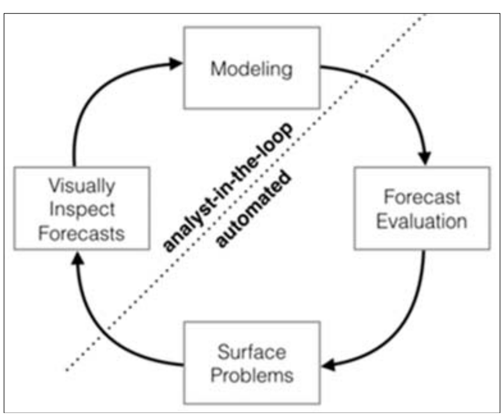
There are many studies on the prediction of epidemic diseases, nevertheless producing high quality prediction is not an easy task for most analysts as forecasting is a specialized data science skill requiring substantial experience [6]. Prophet is an open source library published by Facebook that is based on decomposable (trend+seasonality+holidays) models [1]. It is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. This works best with the time series data that have clear seasonal effects [13]. In this study, the COVID-19 cases in India in the next four weeks have been forecasted with Prophet using R analytics platform. The weekly trend of the cases have also been found using the Prophet package. The viral spread in India has also been compared with other countries across the globe.

Similar adaptations include Cayir *et al*. (2018) for forecasting bitcoin using ARIMA and Prophet models [2], Samal *et al*. (2019) for Time Series based Air Pollution Forecasting using SARIMA and Prophet Model [8] and Andrej Baranovskij (2020) [3] has carried out a study using Growth Modeling and Forecasting techniques for the COVID-19 cases with Prophet. This is partly related to the fact that time series forecasting using Prophet has not been explored yet.

2. Materials and Methods

Prophet is an open source software available in Python and R for forecasting time series data. Prophet is published by Facebook's Core Data Science team. It depends on a contribution model where nonlinear trends are fit with weekly and yearly seasonality and plus holidays. Prophet is strong in dealing with missing data, capturing the shifts in the trend and large outliers. In addition, it arrives at a reasonable estimate of the mixed data without spending manual effort [11]. Purely automatic prediction techniques are not flexible to combine useful assumptions because of their fragility. Furthermore, high quality estimates are not easy to make, requiring special data science skills. All these are determined as working motivations for the Prophet, because it aims to make high-quality predictions easier. Prophet is optimized for business forecasts that are observed on Facebook. For example, time, daily, weekly observations of history, within a year, large outliers, trend changes, missing observations and trends with non-linear growth curves [3].

The Prophet framework has its own special data frame to handle time series and seasonality data easily. The data frame needs two basic columns. One of these columns is "ds" and this column stores date time series. The other column is "y" and it stores the corresponding values of the time series in the data frame. Thus, the framework can work on seasonal time series quite efficiently, also providing options to handle seasonality of the dataset. These options are yearly, weekly and daily seasonality. With provision of these options, a data analyst can choose the available time granularity for the forecast model on the dataset [3].



We use a decomposable time series model with three main components: trend, seasonality, and holidays. They are combined in the following equation:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

g(t): piecewise linear or logistic growth curve for modeling non-periodic changes in time series
s(t): periodic changes (e.g. weekly/yearly seasonality)
h(t): effects of holidays (user provided) with irregular schedules
ϵ_t: error term accounts for any unusual changes not accommodated by the model

Using time as a regressor, Prophet aims to fit several linear and nonlinear functions of time as components. The model parameters g(t), s(t), h(t), ∈t are piecewise linear curves for modeling non-periodic changes in time series, periodic changes, the effects of holidays with irregular schedules, and error term accounts for any unusual changes not accommodated by the model respectively. To fit the proposed model with seasonality effects and forecast based on it, it uses a Fourier series which provides a flexible model. Seasonal effects s(t) can be represented as in Equation (1),

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

where p represents a regular period [1].

3. Results and Discussion

To measure performance of models and to find the best splitting ratios, we have used different metrics for all models. These metrics can be listed as below:

$$R^2 = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad (2)$$

Where $y_i$the predicted value and y is is the mean of the real values. R2 measures goodness of fitting.

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (3)$$

Where N is the number of samples, y_i is predicted value and $\hat{y}_i$ is the real value.

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \quad (4)$$

Where N is the number of samples, y_i is predicted value and $\hat{y}_i$ is the real value. RMSE, RMSPE and, MAPE are derived from MSE and MAE [2].

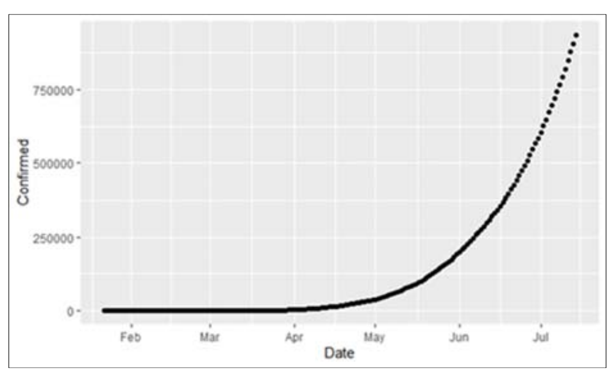


Fig 1: Raw plot of the Confirmed COVID-19 Cases from March 22nd to July 16th in India.

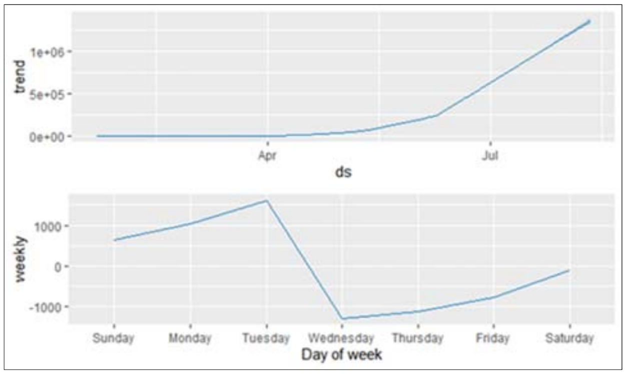


Fig 2: Weekly trend of the Confirmed Cases in India

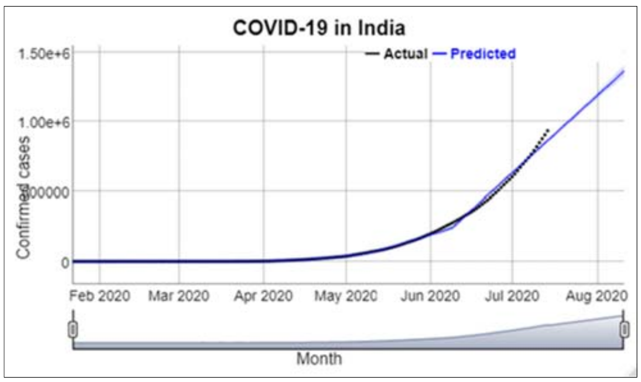


Fig 3: Actual and Predicted cases in India.

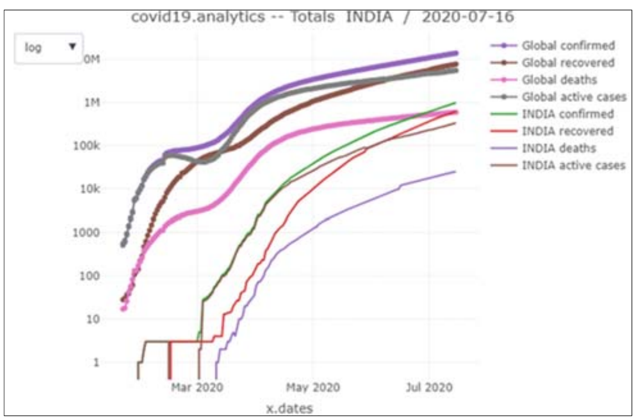


Fig 4: Log plot comparing India with the global conditions.

The forecast has been performed for the Confirmed COVID-19 cases in India till the second week of August. The raw plot (Fig.1) indicates the increasing trend in the confirmed COVID-19 cases with each day attaining the maximum number of cases. From (Fig.2), the weekly pattern for the confirmed cases has been observed with the maximum number of cases reported mostly on Tuesday. The plot diagram (Fig.3) shows that by the end of the second week of August, more than 1,500,000 cases might be witnessed. The slight variations with the actual and predicted cases after June is primarily due to the relaxations in lockdown norms. The log plot (Fig.4) depicts that COVID-19 Confirmed, Recovered, Death and Active cases in India follow a similar pattern as that of the other nations across the globe. The plots have revealed that the trend in cases registered at Indian hospitals has approached its peak in the last two weeks of July. The likely cause being that most people traveling back to their home lands due to few relaxations amidst the lockdown. As a result of this migration of people, viruses could disseminate and expose the symptoms on or after the incubation period. The possible residuals were plotted and statistical analysis was performed using ‘R’ version 3.6.0.

Table 1: The table also provides the goodness of fit statistics, where the R^2 value (0.996) suggests that it could be the best fit for the model.

MODEL	R^2	RMSE	MAE	MPE	MAPE	MASE
Prophet	0.996	14074.8	7216.1	27.5863	88.764	0.040947

A very high Root Mean Square Error and the correlation R^2 values have been found making the model less susceptible to uncertainties. It is seen that the table shows a higher RMSE value and the R^2 value close to 1.

The table also provides the goodness of fit statistics, where the R^2 value (0.996) suggests that it could be the best fit for the model. Also there is less scope of improvement, as the unexplained variability is only 0.04%. Further, Mean Absolute Error indicates that the predicted cases are not so far from the observed cases.

Performance evaluation of the forecast is estimated in terms of Root Mean Square Error (RMSE), R Squared, and the other error values are found to be within the satisfactory limits. The R Squared value (Table I) shows how close the data are to the fitted regression line, implying a very good fit. The model forecasted a steady rise in the number of infected cases in the days to come. This implies that the country's existing measures will not be sufficient to control the ongoing epidemic. Based on the forecasting model and assuming the current conditions of the COVID-19 outbreak, a rapid and exponential increase in the number of cases is expected.

4. Conclusion

In the present study, the data have been considered with respect to the confirmed daily cases of COVID-19 in India. Based on the reported data, the predicted daily number of cases using Prophet will reach upto 1,500,000 cases per day in the next four weeks in India. This prediction has been made in the context of existing conditions. However, a few more preventive measures could improve the current situation. Although the death rate is comparatively low, it is certain that the virus spread is significantly increasing day by day.

The trend in proliferation of cases evinces that, India being the second most populated country in the world, has controlled the spread of the virus to a greater extent when compared to the other highly populated countries. However, it is important to note that the country is reaching its new peak in maximum number of cases almost every day. Thus, the future of the COVID-19 transmission will depend largely on government regulations and encouragement for the individuals to carry out self-isolation.

5. References

1. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2018/05/generate-accurate-forecasts-facebook-prophet-python-r/
2. Aykut Cayir, Bitcoin Forecasting Using ARIMA and PROPHET, 2018. https://www.researchgate.net/publication/329400738_Bitcoin_Forecasting_Using_ARIMA_and_PROPHET Andrej Baranovskij COVID-19 Growth Modeling and Forecasting with Prophet, 2020. https://medium.com/katanaml/covid-19-growth-modeling-and-forecasting-with-prophet-2ff5ebd00c01
3. Anuradha Tomar *et al*. Prediction for the spread of COVID-19 in India and effectiveness of preventive measures.
4. Chen Z-L, Zhang Q, Lu Y, Guo ZM, Zhang X, Zhang WJ, *et al*. Distribution of the COVID-19 epidemic and correlation with population emigration from wuhan, China. Chin Med J (Engl), 2020.
5. COVID-19 Pandemic in India. https://en.wikipedia.org/wiki/COVID-19_pandemic_in_India.
6. Facebook research. https://research.fb.com/blog/2017/02/prophet-forecasting-at-scale/
7. Krishna Rani Samal *et al*. Time Series based Air Pollution Forecasting using SARIMA and Prophet Model, 2019.
8. Nalini Chintalapudi *et al*. COVID-19 virus outbreak forecasting of registered and recovered cases after sixty day lockdown in

Italy: A data driven model approach. Journal of Microbiology, Immunology and Infection. 2020; 53:396-403.
9. Peiris JSM, Poon LLM. Severe Acute respiratory Syndrome (SARS). In: Encyclopedia of virology; 2008. https://www.mendeley.com/catalogue/27b71d32-45a5-3370-94e3-6f02ed0071de/
10. Prophet Forecasting At Scale. https://research.fb.com/prophet-forecasting-at-scale/
11. Risman Baf A. Potential treatments for COVID-19; a narrative literature review.
12. Arch Acad Emerg Med. 2020; 8(1):e29.
13. Sean J Taylor and Benjamin Letham. Forecasting at scale. The American Statistician, (just-accepted), 2017.
14. The Print. https://theprint.in/health/indias-r-value-for-covid-19-drops-back-to-1-11-after-three-month-high-last-week/459874/
15. Which countries are under lockdown - and is it working?. https://www.telegraph.co.uk/coronavirus/