**Forecasting COVID-19 cases in India**

**ABSTRACT**

COVID19, 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.

The purpose of this work is to give a contribution to the understanding of the COVID-19 in India. To this end, we developed a Prophet model for the contagion, and we used official data of the pandemic up to Jan, 2021 for identifying the parameters of this model. The non standard part of our approach resides in the fact that we considered as model parameters also the initial number of susceptible individuals, as well as the proportionality factor relating the forecasted number of positives with the actual (and unknown) number of infected individuals.

**INTRODUCTION**

Coronaviruses are a wide range of diseases, few of which lead to disease in humans and the remaining which mingle between animals and natures. Animal coronaviruses will occasionally transmit to individuals and only transmit to humans. In recent years, zoonotic coronaviruses have formed triggering humanoid outbursts for instance coronavirus ailment 2019 (COVID-19), severe acute respiratory syndrome (SARS), as well as a respiratory syndrome in the Middle East (MERS). The human disease occurs often as a lung infection, or occasionally as a stomach infection. The clinical range of disease ranges from no signs or moderate breathing problems to extreme, increasingly progressing pneumonia, severe breathing suffering condition, infected tremor, or death-induced multiple-body part catastrophe.

As of December 2020, India registered more than 1 crore established crisis of COVID-19). Around 97 Lakhs individuals are now in good health from these, whereas 1,45,000 crises have led to death.

Transmission of infectious disease is a dynamic mechanism of transmission that happens within

the crowd. Frameworks can be developed for this method to potentially examine and test the propagation mechanism of infectious diseases so that we can forecast correctly the future pattern of infectious diseases. Therefore, to monitor or reduce the damage of infectious diseases, the study and review of predictive models for infectious diseases have been a hot topic of science.

Therefore, We have studied the behaviour of covid19 in India using EDA techniques and data visualization and then Using the time series forecasting techniques like Facebook Prophet Model, SIR Model, We will be developing the regression models to predict the number of cases in India. These models clearly provide a more detailed description of the epidemic spread than collective models but their identification is significantly harder. A first reason is that they are usually characterized by a high number of parameters and variables. A second reason, perhaps more relevant, is that the network topology is unknown in most real situations and its identification is an extremely hard task. In this paper, we focus on collective models since, thanks to their relative simplicity, they can be more suitable for non expert operators and public health authorities, and they can provide simple but reliable models, even under scarcity of data.

**MATERIALS AND METHODS**

We describe the dataset used to estimate the work, prediction algorithms, and model accuracy metrics.

## **Covid-19 dataset***:*

## The dataset used in this study includes the India total & daily confirmed, deaths cases of Covid-19, collected from the official website Our World in data in the period from January, 2020, to January 2021. It comprises time-series cases from which to build our model, which we compare to other predictive models. Descriptive statistics tells the nature of data. To know the information about the data tools to summary statistics used like mean, standard error, minimum and maximum, skewness and kurtosis.

**Prediction Algorithms:**

Various time series predicting models are available as ARIMA, SARIMA, GARCH, Prophet and LSTM etc. Here we are using ARIMA, SARIMA and one machine learning model Prophet.

**The Prophet Model**:

Prophet 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. It works best with time series that have strong seasonal effects and several seasons of historical data.

The procedure makes use of a decomposable time series model with three main model components: trend, seasonality, and holidays. Similar to a generalized additive model, with time as a regressor, Prophet fits several linear and Non-linear functions of time as components. In its simplest form.

**y*(t)* = g*(t)* + s*(t)* + h*(t)* + *e(t)***

Where

g(t) is the trend models non-periodic changes (i.e. growth over time)

s(t) is the seasonality presents periodic changes (i.e. weekly, monthly, yearly)

h(t) ties in effects of holidays (on potentially irregular schedules ≥ 1 day(s))

e(t) covers idiosyncratic changes not accommodated by the model

In other words, the procedure’s equation can be written

***y(t)=piecewise\_trend(t)+ seasonality(t)+holiday\_effects(t)+noise***

**Trend:**

The procedure provides two possible trend models for g(t), "a saturating growth model, and a piecewise linear model."

**Saturating Growth Model:**

If the data suggests promise of saturation i.e. one is wrestling constraints like: cubed footage, processing power, number of people w/ Internet access— setting growth='logistic' is the move.Typical modeling of these ***nonlinear, saturating trends*** is basically accomplished

Image for post

where:

C is the carrying capacity

k is the growth rate

m is an offset parameter

**Rate of Change v. Time:**

Second, the market does not allow for stagnant technology. Advances like those seen over the past decade in handheld devices, app development, and global connectivity, virtually ensure that growth rate is not constant.

Because this rate can quickly compound due to new products, the model must be able to incorporate a varying rate in order to fit historical data.

We incorporate trend changes in the growth model by explicitly defining changepoints where the growth rate is allowed to change.

Suppose there are S changepoints at times sj, j = 1,…,S.

Prophet defines a vector of rate adjustments

Image for post

Where

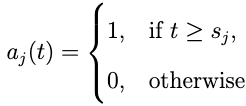
**δ***j* is the change in rate that occurs at time ***s****j*

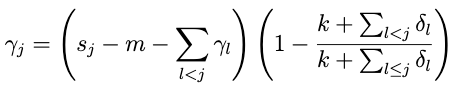
The rate at any time ***t*** is then the base rate ***k***, plus adjustments up to that time

Image for postThis is represented more cleanly by defining a vector

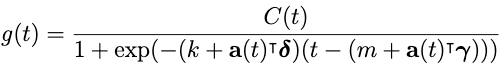
Image for post

such that



The rate at time t is then k+a(t)ᵀδ. When the rate k is adjusted, the offset parameter m must also be adjusted to connect the endpoints of the segments. The correct adjustment at changepoint j is easily computed a **

At last, the piecewise growth=‘logistic’ model is reached



An important set of parameters in our model is C(t), or the expected capacities of the system at any point in time. Analysts often have insight into market sizes and can set these accordingly. There may also be external data sources that can provide carrying capacities,such as population forecasts from the World Bank.

In application, the logistic growth model presented here is a special case of generalized logistic growth curves — which is only a single type of sigmoid curve — allowing the relatively straightforward extension(s) of this trend model to other families of curves.

**RESULTS AND DISCUSSIONS:**

In this section, the summary statistics of Covid-19 dataset of India is shown. We have developed the model which forecasts the covid 19 cases in India and we have also identified some patterns of how cases were increased/decreased daily since january 30.

**Exploratory Data Analysis(EDA):** Itis important to develop good predictive models. The daily cases in India started increasing after may 2020 as many people in India started getting back to the work. The daily cases from Jan 2020 In India is shown in the figure 1.

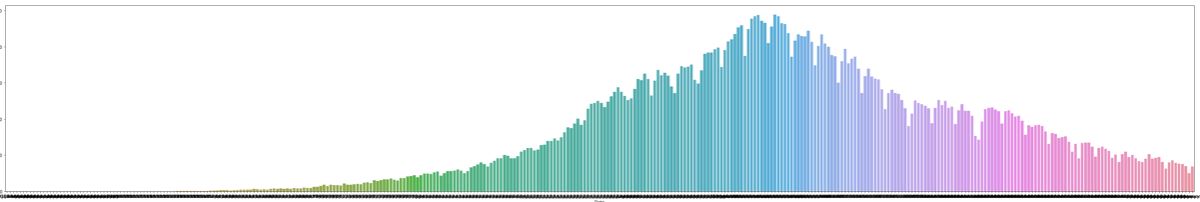


Figure 1 : Daily confirmed cases in India

In India though the highest number of daily cases have gone above 90000, The recovery rate was high and death rate was low. Initially the daily cases was low. In figure 2 the nationwide confirmed recovered and Deceased cases till January 19th are shown. We can see that the recovery rate was higher.

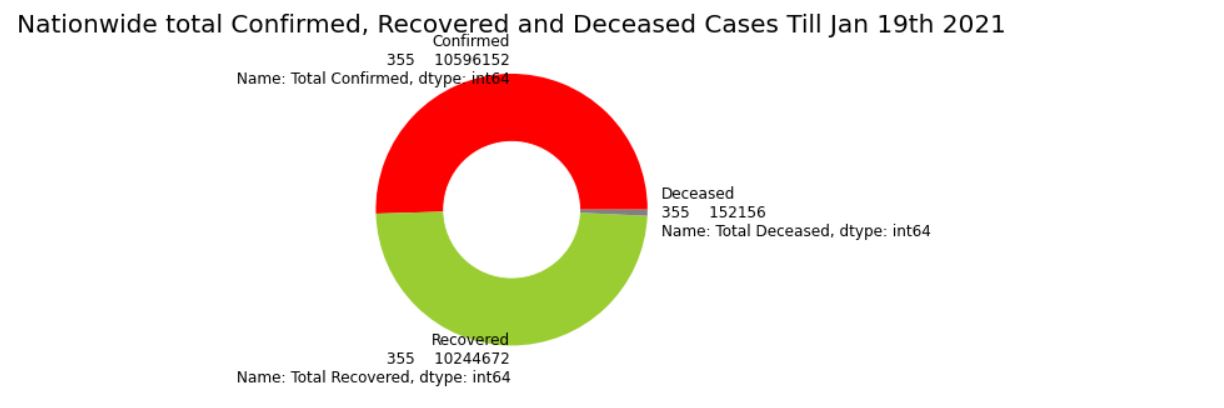


Figure 2 : Nationwide confirmed recovered and Deceased cases in India

As the number of Cases started going to peak, the number of recoveries has also increased. The cases started declining in the month of October with the same recovery rate. With this many Indian Doctors claimed that Most of the people in India have reached close to natural herd immunity and most of the Indians may not require Covid 19 Vaccine. The daily recoveries are shown in Figure 3.

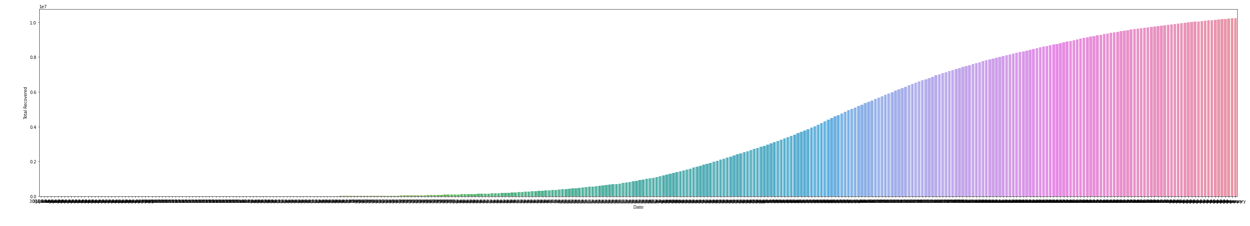


Figure 3 : Total number of recovered cases in India

**Time Series Prediction:**

We have done the future predictions of Number of cases in India using the Facebook Prophet Model. With forecasting we have also identified the number of patterns behind the number of cases increasing/decreasing everyday and every week.

In figure 4, we see that every Monday the number of cases were decreased because the number of testings were decreased during the weekends.

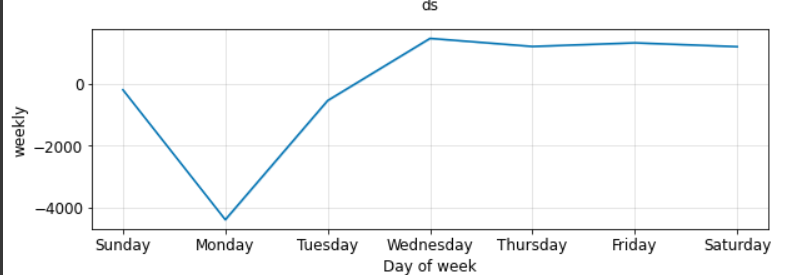


Figure 4 : Weekly analysis on Number of cases

The Figure 5 is the snapshot of our predictions on Number of cases. The data trained with the model consisted of the number of cases till 31st December 2020 and the predictions are made for the whole January.

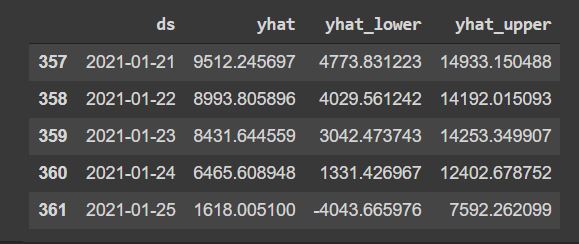


Figure 6: Number of cases forecasted with FBProphet Model

The overall result is shown in Figure 6. The predicted results are close the actual results.

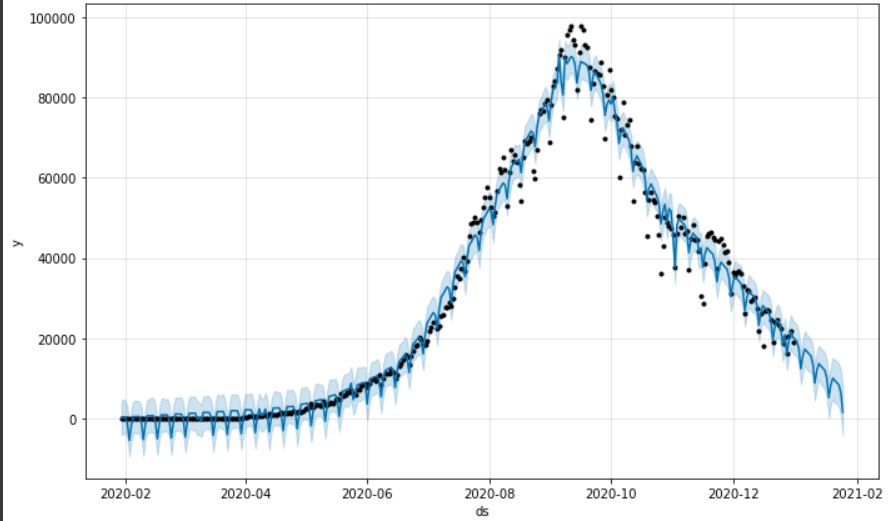


Figure 7: Forecasted Numbers vs Actual Number

**References**:

# Modelling and Forecasting of COVID-19 in India by Pradeep Mishra1\*, Abdullah Mohammad Ghazi Al Khatib2, Iqra Sardar3, Jamal Mohammed4, M Ray5, Manish Kumar6, Deepa Rawat7, SA Pandey8, Anurag Dubey9, Jos Feys10 and Kipkoech Rono11 <https://clinmedjournals.org/articles/jide/journal-of-infectious-diseases-and-epidemiology-jide-6-162.php?jid=jide>

A machine learning forecasting model for COVID-19 pandemic in India R. Sujath, Jyotir Moy Chatterjee & Aboul Ella Hassanien https://link.springer.com/article/10.1007/s00477-020-01827-8

Sean J. Taylor, Benjamin Letham, *Forecasting at Scale, Facebook*, Menlo Park, California, United States <https://peerj.com/preprints/3190.pdf>

<https://towardsdatascience.com/forecasting-with-prophet-d50bbfe95f91>