

# COVID-19 India Outbreak Trend Analysis and Predictions using Multi Time Series Models

Jyoti Khurana

Assistant Professor

Information Technology Department

Maharaja Surajmal Institute of Technology, Delhi,  
India

Anmol Ashri

Student

Information Technology Department

Maharaja Surajmal Institute of Technology, Delhi,  
India

## ABSTRACT

COVID-19 has been declared as the pandemic and is spreading at an alarming rate. The current study describes the situation outbreak in India. COVID-19 is a time-series data set which contains both non-linear and non-stationary patterns, and it is highly recommended to use the model which can extract the design using sequential networks. This study mainly focuses on the minimum number of hospital beds required for COVID-19 patients along with the prediction of confirmed covid-19 cases and total predicted deaths due to prevailing pandemic till 4 February 2021 using multi-time series forecasting models: (Facebook) Fb Prophet Model and Auto-Regressive Integrated Moving Average Model (ARIMA). The performances of the two models are compared on the basis of error metrics: root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE). The study also uses exploratory data analysis to report the current situation of covid-19 in India. The data used in the study is taken from datasets available on Kaggle.com and covers up time period till 6 December 2020. All the data visualization, analysis and prediction are made using Python 3 in Jupyter notebook. The performed study can help government and healthcare communities to initiate appropriate measures to control this outbreak in India.

## Keywords

COVID-19, Time Series, Fb Prophet, ARIMA, Data Visualization, Predictions

## 1. INTRODUCTION

With the initial outbreak of the coronavirus originated from Wuhan in China around December 2019, the virus has spread throughout the world within the next few weeks [1]. The exponential growth of the infection among the world's different countries has presented an unprecedented challenge. The World Health Organisation (WHO) has declared this contagious disease as pandemic [2]. The world is moving towards a very stressful stage with the spread of novel coronavirus. Different countries in the world are suffering from this pandemic and causing disaster to the public health system and bringing enormous disruption to the economy and society [3]. With the prevailing circumstances, it has become significant for the healthcare professionals and policymakers to plan and adequately equipped for the conditions that may arise due to the rapid spread of the COVID-19.

Countries are using various Information, Communication, and Technological (ICT) aspects to control the situation. Different ICT tools, applications, and services are used to predict future conditions and plan accordingly [4]. Artificial Intelligence (AI) can play a significant role in fighting against the

pandemic situation. Different Time-series AI models can be used for estimating and predicting the spread rate. A Laplacian based model has been developed to study the transmission of Pine Wilt disease [5]. A mathematical model for analysing and understanding human liver behaviour and the spread of dengue was predicted [6]. A model was designed using SIRS [7] to resist the spread of the syncytial virus in infants.

Every infectious disease epidemic reveals specific patterns and such patterns needed to be studied based on transmission dynamics of such outbreaks. Paramount methods to eliminate such contagious infections rely on the processes to be followed to predict the future of such diseases. India is a vast country with 1.38 billion populations. Each state of this country is at a different stage of the epidemic. There is massive pressure on the administration and health care professionals for accommodating different patients with the symptoms of coronavirus. It is the essential requirement to understand the growth pattern of the virus; as such information will help to handle the upcoming situation. This information will play a crucial role in planning in advance to overcome the situation of panic. So, in order to facilitate such information timely, time-series data-driven prediction models must be used to predict the number of cases for the aid of future health planning [3].

The data is growing and changing dynamically, so using statistical and epidemiological methods will result in uncertainty [8,9]. The researchers have used various deep learning and machine learning-based models to predict the number of cases of COVID-19 infections [10] in order to facilitate in advance with the requirements such as extensive testing, number of beds in the hospital and others to control the spread and reduce the mortality rate.

In this paper, we have used multi time series models such as Fb Prophet Model, ARIMA Model for prediction of the COVID-19 total number of cases in India in up to 4<sup>th</sup> February 2021. These models will allow dealing with the data with different veracity as making a decision based on an individual model would be critical. With the help of these models we have tried to predict a total number of COVID-19 cases in India, the total number of mortality due to the significant COVID-19 outbreak in India and minimum whole numbers of hospital beds would be needed for COVID-19 patients in India in order to control the further outbreak of this epidemic in India. In the absence of a permanent cure of COVID-19, these approximations will provide an intuition into the resource allocation to keep this epidemic under control. Besides focusing on the dynamics of the COVID-19 outbreak, the practical purpose of this analysis will provide government

officials with realistic approximations which can be helpful in deriving policy making to control the magnitude of this epidemic. Such assessment of mortality due to COVID-19 and number of required hospital beds for COVID-19 patients will help to anticipate the extent of this epidemic in India. They will provide some critical information for planning the health care system in India to control this epidemic.

Further, the paper is organized into the following sections. Section 2 will discuss the mathematical background of the prediction models. Section 3 will illustrate the dataset used for prediction and visualization of the COVID-19 outbreak in India. Experimental results and visualization details will be discussed in Section 4. Section 5 will conclude the paper.

## 2. PREDICTION MODELS

Time-series is simply a series of data points ordered in time. In a time-series, time is often the independent variable, and the goal is usually to make a forecast for the future. It is a forecasting area that focuses on analysing past observations of random variables to develop a model that captures the underlying relationship and its pattern. This approach is usually beneficial when there is no knowledge or little knowledge of the data generating process. Over past decades, a lot of research has produced a significant development and improvement of time series forecasting models.

In this section, the time-series models that have been used for predictions are detailed as:-

### 2.1 Auto-Regressive Integrated Moving Average Model (ARIMA)

One of the most well-known and widely used families of time series models include the Auto Regressive Integrated Moving Average (ARIMA) model. This model is a well-known implementation of Box-Jenkin methodology [11]. It is a time-series data analysis, based on a three-class component to manage stationary and non-stationary time datasets.

The model is generally denoted by ARIMA (p, d, q), where the parameter p and q are the order of the Auto-Regressive (AR) and Moving Average (MA) models and the parameter d is the level of differencing. The first component is AR(p) of time series model deals with the dependent observation for predicting future events. The Second component MA(q) handles different white noise annotations representing past forecast errors for the estimation of the outcomes depending on the future value. This combination allows the ARIMA model to deal with the time-series non-stationary values. Both combinations will make an ARIMA model. But as in the mean of the observed data, the variance is not constant, so the third component is used as Integrating (I(d)) to convert the observations using differencing series [12].

Further, the model can be mathematically expressed as:-

$$y'_t = \mu + \delta_1 y'_{t-1} + \delta_2 y'_{t-2} + \dots + \delta_p y'_{t-p} + \rho_1 y'_{t-1} + \rho_2 y'_{t-2} + \dots + \rho_\varphi y'_{t-\varphi} + \epsilon_t \quad (1)$$

where  $y'_t$  denotes the actual value of the variable under consideration at time t,  $y'_{t-1}$  is the random error at time t. The  $\delta_p$  and  $\rho_\varphi$  are coefficients of the ARIMA model. Building an ARIMA model involves generally three iterative steps: Achieving stationarity (A model that shows stationarity is one that shows there is constancy to the data over time);

Autocorrelation function (ACF) and the partial autocorrelation function (PACF) plots are used for parameters (p and q) estimation and model diagnostics are checked to find the 'best' fitted forecasting model using Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) [6].

### 2.2 Facebook Prophet Model (Fb Prophet)

Prophet is an open-source tool from Facebook initially developed to forecast time series data related to businesses' time series problems. It is based on a decomposed additive model where non-linear trends are fit with seasonality; it also takes into account the effects of holidays. The trend shows the tendency of the data to increase or decrease over a long period, and it filters out the seasonal variations. Seasonality is the variation that occurs over a short period, and is not prominent enough to be called a "trend".

Mathematically, Prophet model is represented as:

$$y(t) = g(t) + s(t) + h(t) + e(t) \quad (2)$$

where:

- $g(t)$  refers to trend (changes over a long period of time).
- $s(t)$  refers to seasonality (periodic or short term changes).
- $h(t)$  refers to effects of holidays to the forecast.
- $e(t)$  refers to the unconditional changes that are specific to a business or a person or a circumstance. It is also called the error term.
- $y(t)$  is the forecast.

Fundamentally it is a regression model with explicable parameters that often work well with their default values. It also allows the user to choose the components concerning their forecasting and give better results based on necessary adjustments [14]. To forecast trend prophet use two models: Logistic growth model and the other one is the piecewise linear model. If data to be forecasted has saturating and non-linear data (grows non-linearly) and after reaching the saturation point shows little to growth or shrink and only exhibits some seasonal changes then logistic growth model is the best option. If the data shows linear properties and growth or shrink trends, then a piecewise linear model is a better choice.

The logistic growth model is fit under following statistical equation:-

$$g(t) = \frac{C}{(1+e^{-k(t-m)})} \quad (3)$$

where

- C is the carry capacity
- k is the growth rate
- m is an offset parameter

Piecewise linear model is fit using the following statistical equations :-

$$y = \beta + \beta_1 \times x, x \leq c; \quad (4)$$

$$y = \beta - \beta_2 \times c + (\beta_1 + \beta_2) \times x, x > c \quad (5)$$

where  $c$  is the trend change point and  $\beta_1$  and  $\beta_2$  are dummy variables in (4) and (5). To find seasonality, it depends on the Fourier series to provide a flexible periodic effect model [15]. To account for holiday, a predefined data of past and future holiday's events is required, holiday effects are assumed to be independent so assimilating them to the model becomes insignificant [14].

### 3. DATASET

In this paper, the prediction is made for calculating the total number of confirmed cases, death cases and the minimum number of hospital beds required by the patients to fight against COVID-19. All these datasets are collected from the starting date of the epidemic in India to 06/12/2020. We have explored two datasets for the prediction and visualization of the COVID-19 outbreak.

#### 3.1 Novel Coronavirus 2019 dataset

Due to the long duration of COVID-19 pandemic, the data were continuously changing and collected periodically by researchers at John Hopkins University and made publically available on [16]. The dataset contains the data from a date - 20/1/2020. The data includes the number of confirmed cases, number of cured patients and mortality rate which is maintained on a time-series basis. The data is provided with the country name, code, city/province, sex, age and reporting date of confirmed COVID-19 patients worldwide.

#### 3.2 COVID-19 in India Dataset

The dataset is extracted and made from the data provided by covid19india.org for the individual level details and testing details in India and rest of data is provided by Indian Ministry of Health & Family Welfare [17]. The dataset contains the data from date 30/1/2020 on this date the first confirmed case of COVID-19 was reported in India from Kerala. The dataset comprises time-series data including the count of confirmed cases, count for cured patients and death rate of India, and individual states. This dataset also provides us with personal testing details such as reported date, age, gender, detected city and detected state. It also provides the number of COVID-19 daily tests done by each state in India and the number of running COVID-19 labs headed by ICMR (Indian Council of Medical Research). Dataset also provides data on the number of hospital beds in each state and data on the medical infrastructure of each state. Additionally, it also includes information on population at a state level. We have used this dataset for our COVID-19 visualization, forecasting and prediction of the total number of COVID-19 cases in India with the prediction of the total number of deaths and predicted the number of hospital beds required in India to tackle the COVID-19.

## 4 VISUALIZATION OF COVID-19

### 4.1 Visualization of COVID-19

To better understand the current position of the COVID-19 in India, this section explores the visualization of the pandemic in India from the data taken in consideration. Data is visualised on the basis of different criteria such as number of total cases, number of total recovered cases and number of

total deaths occurred from COVID-19 spread in India. The section covers different aspects of pandemic to understand the spread in a better way. We have used Python 3.0 programming language to show all visualization with the help of Jupyter Notebook which is an open source web application used by us to run all Python code.

#### 4.1.1 Rate of spread of COVID-19 in India

First case of COVID-19 in India was reported on 30/1/2020 in Kerala and from that day the number of COVID-19 cases is increasing gradually in India. We have used the COVID-19 India data till 20/7/2020. Number of total confirmed COVID-19 cases in India till 06/12/2020 is 96,77,203 ; total number of recovered COVID-19 cases in India till 06/12/2020 is 91,39,901 and total number of deaths due to COVID-19 outbreak in India till 06/12/2020 is 1,40,573 as shown in figure 1.

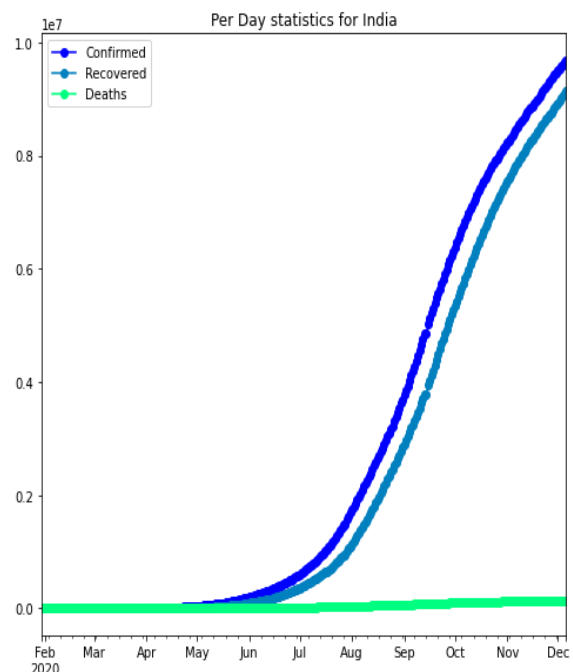


Figure 1. Rate of spread of COVID-19 in India

The effect of pandemic in different States of India is represented in Figure 2. It can be evaluated that Maharashtra leads in the total number of confirmed cases (18,59,367) among all the states in India. It also leads in the total number of deaths(47827) due to the current outbreak. Karnataka and Andhra Pradesh are the states after Maharashtra in number of confirmed cases. Assam, Odhisha, Jharkhand have the same highest recovery rate (total number of recovered COVID-19 cases /total number of COVID-19 cases in the state) among all the states in India with a recovery rate of 0.98 followed by Telengna, Bihar which have a recovery rate of 0.97. Maharashtra, Punjab have an equal death rate (total number of total deaths due to COVID-19 in state/total number of confirmed COVID-19 cases in the state) of 0.03 which is highest among other states and greater than India's death rate of 0.01.

State	Confirmed	Deaths	Recovered	Recovery Rate	Death Rate
Maharashtra	1859367	47827	1737080	0.93	0.03
Maharashtra***	1723135	45325	1581373	0.92	0.03
Karnataka	895284	11880	858370	0.96	0.01
Andhra Pradesh	872839	7042	860368	0.99	0.01
Tamil Nadu	792788	11822	770378	0.97	0.01
Kerala	644696	2472	582351	0.9	0
Delhi	597112	9763	565039	0.95	0.02
Uttar Pradesh	558173	7967	528832	0.95	0.01
West Bengal	507995	8820	475425	0.94	0.02
Odisha	321913	1784	316970	0.98	0.01
Rajasthan	284116	2468	260773	0.92	0.01
Telangana	275261	1480	266120	0.97	0.01
Chhattisgarh	249699	3025	227158	0.91	0.01
Haryana	246679	2624	232108	0.94	0.01
Bihar	239322	1300	232563	0.97	0.01
Gujarat	221493	4110	203111	0.92	0.02
Madhya Pradesh	217302	3358	200664	0.92	0.02

Figure 2. State Wise Covid-19 Trend in India

#### 4.1.2 Healthcare facilities in India

With the outbreak of pandemic in every state of the country crossing different stages, it has become essential for the healthcare policy makers to conduct proper tests among the population of different States. It will help to control the epidemic timely and helps to insure the real scenario. Till 06/12/2020 India has done about 14,77,87,656 COVID-19 tests and in terms of testing India stands at second position [18] among all countries affected. Figure 3 shows, COVID-19 testing done by each state. Figure 3 shows that Uttar Pradesh among all the states in terms of COVID-19 testing with 2,06,21,450 followed by Bihar (1,56,64,720) and Tamil Nadu being at third position with 1,26,05,290 number of tests. India has a diverse healthcare system with both public and private health sectors playing essential roles. Accommodation of inundation of COVID-19 patients, will require rapid augmentation of current magnitude or modifications in the health care system of India. India has 37,725 [19] number of hospital facilities in the public sector and 7,39,024 [19] number of hospital beds in the public sector.

Statewise testing

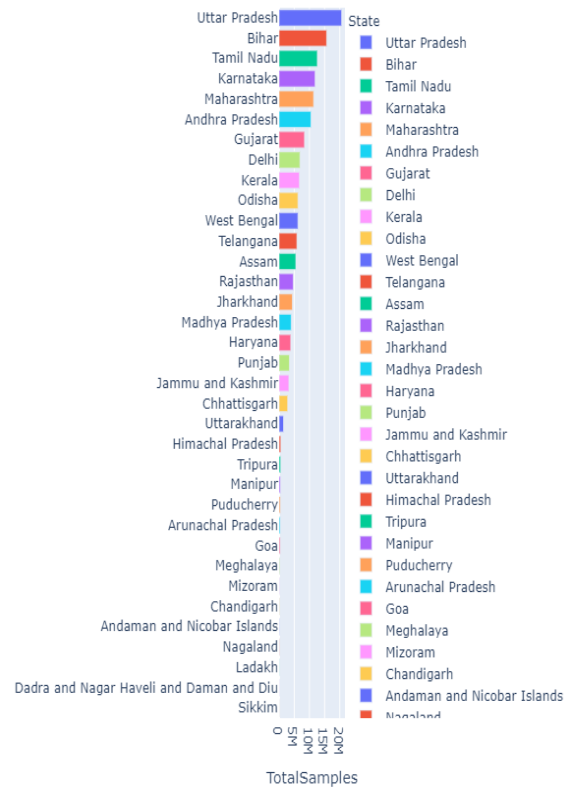


Figure 3. Distribution of Tests conducted in each state

## 4.2 Experiment Results and Analysis

In this section, prediction of COVID-19 for total number of confirmed cases, minimum number of hospital beds required for the COVID-19 patients in India and number of mortality due to COVID-19 outbreak in India till 04/02/2021. The experiments are performed on the data [16] from 30/1/2020 to 06/12/2020 using ARIMA and Fb Prophet model.

The accuracy of the models is calculated using root mean square error (RMSE)[21], mean absolute error (MAE)[21] and mean absolute percentage error (MAPE)[22].

### 4.2.1 Prediction for total number of Confirmed cases in India

Firstly, prediction is done using the Fb Prophet model. Table 1 shows the results of the prediction carried out for the 04/02/2021 as date timestamp. We have also added change points (These are important points to identify the period where the data set trend changes).

Table 1. Predictions from 26/1/2021 to 04/2/2021 using Fb Prophet model

Date Timestamp	Actual Prediction	Lower Approximation	Upper Approximation
2021-01-26	11937940	10096140	12788990
2021-01-27	11983990	11007960	12863350
2021-01-28	12028340	11032410	12947210

2021-01-29	12073090	11047820	13026520
2021-01-30	12117840	11071420	13087300
2021-01-31	12161140	11092830	13161320
2021-02-01	12199070	11097660	13195470
2021-02-02	12242030	11112870	13262780
2021-02-03	12288090	11149220	13328920
2021-02-04	12332430	11160570	13407490

The actual prediction shows that the number of confirmed COVID-19 cases in India based on the trend calculated by the model by 04/02/2021 would be 1,23,32,430 as exact predictions while 1,11,60,570 lowest value of prediction and 1,34,07,490 maximum value of prediction. The Figure 5 indicates the relationship of the total confirmed COVID-19 cases in India with time where the x-axis shows time and y-axis shows the numerical value of confirmed COVID-19 cases. The black dotted line indicates original values and the blue solid line indicate the predicted values of the COVID-19 confirmed cases of India.

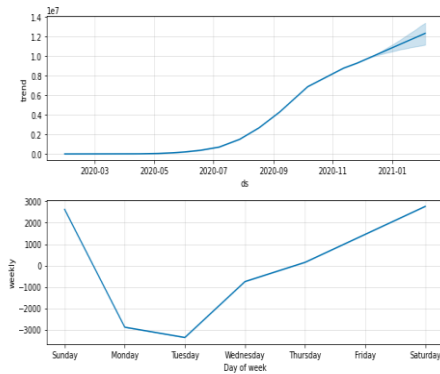


Figure 4 .Representation of Trend and weekly analysis

Figure 4 above indicates the trend of confirmed cases and weekly analysis of COVID-19 cases till 04/02/2020. The weekly analysis represents the number of cases that are maximum on Sunday and minimum on Tuesday. Secondly, prediction was done on the basis of the ARIMA time series model. Autocorrelation function (ACF) graph and partial autocorrelation (PACF) as shown in figure 6(b) were used to estimate the parameters of ARIMA model. Through experimentation, the best fitted forecasting model using Akaike Information Criterion (AIC)[23] and the Bayesian Information Criterion(BIC)[23], the values for the parameters that ARIMA(3,1,2) is best fitted for which both AIC and BIC are lowest(AIC=6468.671 ,BIC=6494.736) and it's model summary shown in table 2. Using this we have predicted the total number of confirmed COVID-19 cases in India upto 04/02/2021 shown in figure 6 which includes graphical representation with 95 % CI(Confidence Interval). For better visualization we have shown prediction using a python library called matplotlib, as shown in figure 7.

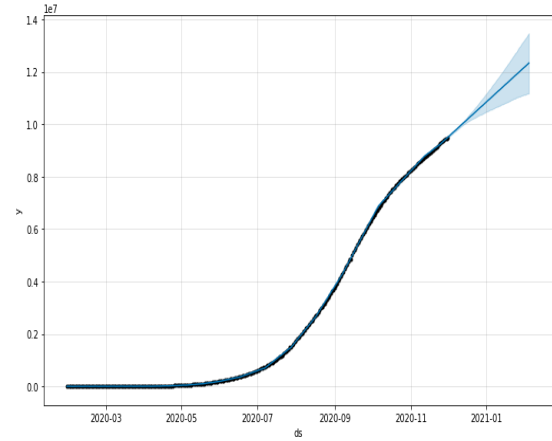


Figure 5 .Prediction of total numbers of confirmed cases in India using FB prophet

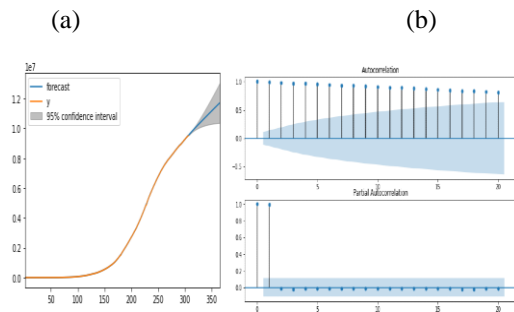


Figure 6. (a) Prediction of COVID-19 confirmed cases using ARIMA model (b) ACF and PACF plots.

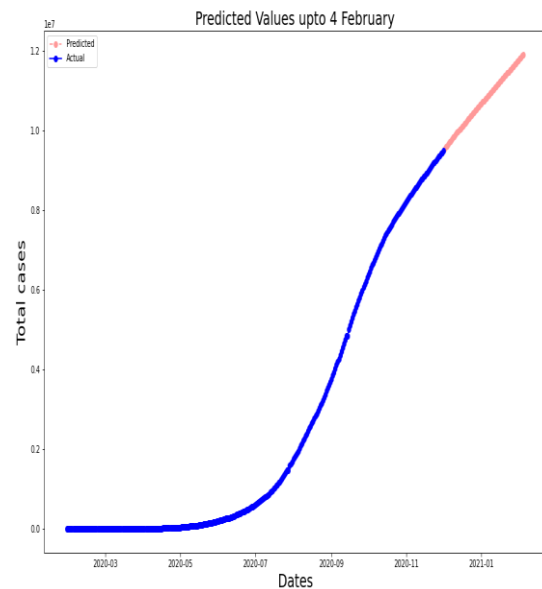


Figure 7. Prediction of COVID-19 confirmed cases using ARIMA till 04/02/2021

From figure 7, it can be predicted that the exact number of COVID-19 cases based on the data used are 1,19,16,970 with 1,03,77,872 lower limit value and 1,34,56,074 higher limit value. To evaluate the performance of the models, we have split our data into train set which contain the data till 1/12/2020 and a test set which contains data from 2/12/2020

to 6/12/2020 is used to evaluate error metrics of our model. We have calculated error metrics from 2/12/2020 to 6/12/2020. We have shown RMSE, MAE, MAPE of both models in table 3.

**Table 2. Model Summary ARIMA(3,2,1)**

ARIMA(3,2,1)	Std error	Z	P> z
AR(1)	0.065	24.918	0.000
AR(2)	0.110	-4.344	0.000
AR(3)	0.077	-1.929	0.054
MA(1)	0.037	-49.614	0.000
MA(2)	0.029	30.580	0.000

**Table 3. Evaluation of the Models on the basis of total number of confirmed cases in India**

Serial Number	MODEL	RMSE	MAE	MAPE
1	ARIMA	24497.2	24497.2	0.002542
2	FB Prophet	32671.8	32671.8	0.003459

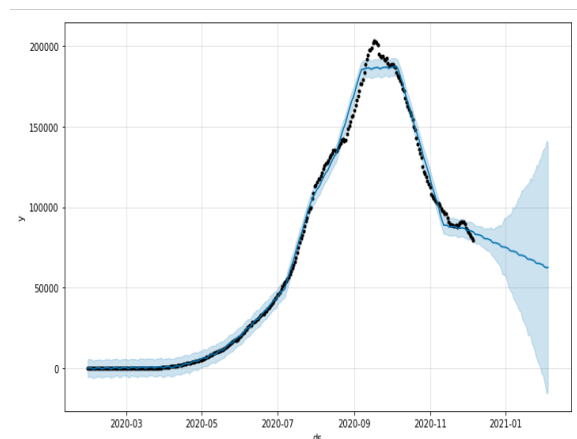
#### 4.2.2 Prediction for Minimum number of hospital beds required

COVID-19 has put humongous pressure on the medical infrastructure of India. As the cases are increasing, demand for hospital beds are also increasing with it. According to WHO not all COVID-19 patients need to be hospitalized, only 20 percent of them need special treatment or hospitalized so based on this we have predicted the minimum number of beds required for COVID-19 patients to fight against the on-going pandemic [20]. Nevertheless more investigation as factors like individual patients health report ,age distribution of patients are not taken into consideration in this prediction. We have calculated this by calculating total active COVID-19 cases in India and it is assumed 20 percent from active cases need hospital support according to WHO [1]. At first we showed predictions using the FB prophet Model . Below table which shows the prediction of minimum number of hospital beds in India from 26/1/2021 to 06/2/2021. Table 4 shows that the number of confirmed COVID-19 cases in India based on the trend calculated by the model by 04/2/2021 would be 62,412 as exact predictions while 0(-14964) lowest value of prediction and 1,40,075 maximum value of prediction. There are some values in table which are shown 0 because in prediction curve shown in Figure 8, lower approximation becomes negative but in real world scenario negative value is not considered so instead of those negative values ,0 value is considered.

**Table 4. Predictions from 26/01/2021 to 04/02/2021 using Fb Prophet model. Value 0 is considered to show realistic predicted values instead of negative values**

Date Timestamp	Actual Prediction	Lower Approximation	Upper Approximation
2021-01-26	65032	5610	126316
2021-01-27	65022	3600	125143
2021-01-28	64965	2388	128905
2021-01-29	64670	176	129974
2021-01-30	64339	0(-3581)	132624
2021-01-31	64084	0(-4002)	135320
2021-02-01	63041	0(-7913)	136253
2021-02-02	62479	0(-11600)	137202
2021-02-03	62470	0(-14720)	141148
2021-02-04	62412	0(-14964)	140075

The figure 8 indicates the relationship of the minimum number of hospital beds with time where the x-axis shows time and y-axis shows the numerical value of minimum number of hospital beds. The black dotted line indicates original values and the blue solid line indicates the predicted values of minimum number of hospital beds. Almost the predicted and the original values are going parallel as shown in the figure. Decreasing curve in the figure 8 is due to the daily decrease in the total number of active Covid-19 cases in India which further affects our prediction of minimum number of hospital beds required for Covid-19 patients.



**Figure 8. Prediction of minimum number of hospital beds using Fb Prophet model**

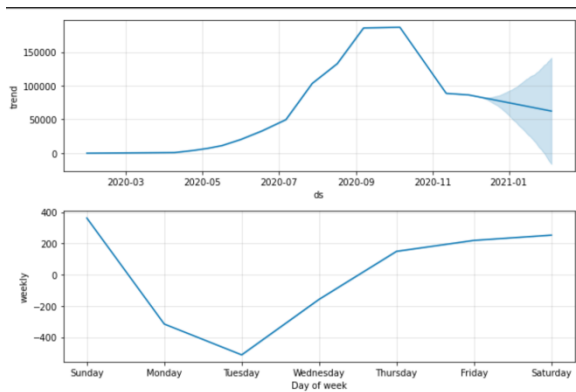


Figure 9. Representation of trend and weekly analysis

Figure 9 indicates the trend of our prediction for the number of hospital beds and weekly analysis of the number of hospital beds till 04/02/2020. The weekly analysis was maximum on Sunday, then the weekly analysis decreased till tuesday then it increased again. Secondly, we have predicted the number of hospital beds, using the ARIMA time series model. Autocorrelation function (ACF) graph and partial autocorrelation (PACF) as shown in figure 10(b) were used to estimate the parameters of ARIMA model. Through experimentation, the best fitted forecasting model using Akaike Information Criterion (AIC)[23] and the Bayesian Information Criterion(BIC)[23], we have found by the experimentation that ARIMA(0,2,1) is best fitted for which both AIC and BIC are lowest (AIC= 5059.281, BIC=5070.442) and its model summary shown in table 5. Using this we have predicted minimum number of hospital beds required for COVID-19 patients in India upto 04/02/2020 shown in figure 10(a) which includes graphical representation with 95% CI (Confidence Interval). For better visualization we have shown prediction using python library called matplotlib shown in figure 11.

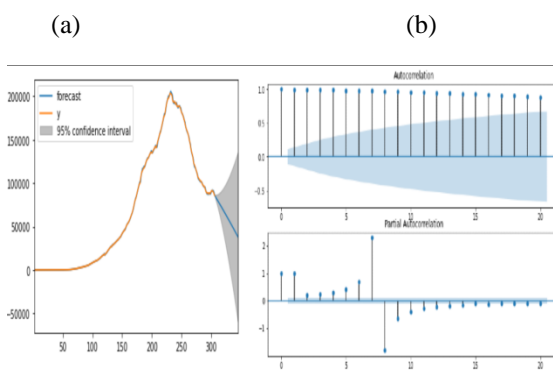


Figure 10. (a) Prediction of minimum number of hospital beds left side in which y shows actual number of required hospital beds for COVID-19 patients in India (b) ACF and PACF plots

We can evaluate from figure 11 that the exact number of predicted cases based on the data used are 3726 with 0 lower limit value (as lower limit value is negative so value 0 is considered for realistic prediction) and 135688 higher limit value. To validate the accuracy of our models we have split our data into train set which contain the data till 1/12/2020 and test sets which contain data from 2/12/2020 to 6/12/2020

are used to evaluate error metrics of our model. We have calculated error metrics RMSE, MAE, MAPE shown in table 6.

Table 5 Model Summary of ARIMA(0,2,1)

ARIMA(0,2,1)	Std error	Z	P> z
MA(1)	0.045	-14.727	0.000

Table 6 Evaluation of the Models on the basis of minimum number of hospital beds for covid-19 patients in India

Serial Number	Model	RMSE	MAE	M APE
1	ARIMA	341.6	341.6	0.00424
2	FB Prophet	7718.1	7718.1	0.09532

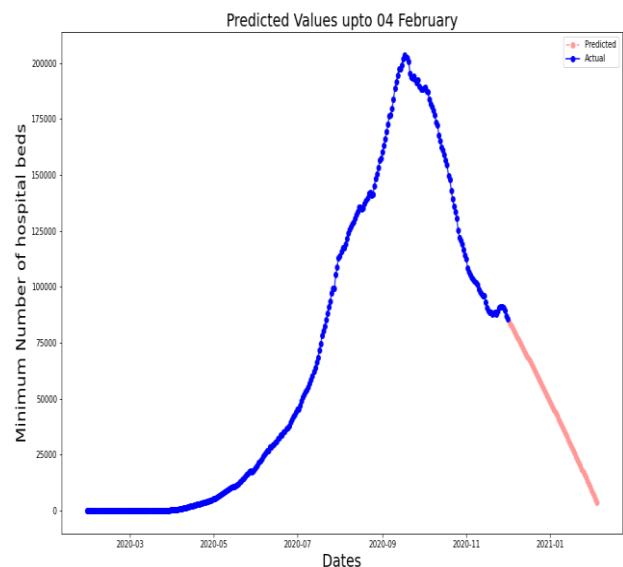


Figure 11. Prediction of minimum number of covid-19 beds using ARIMA till 30/9/2020

#### 4.2.3 Total number of mortality due to COVID-19 outbreak in India

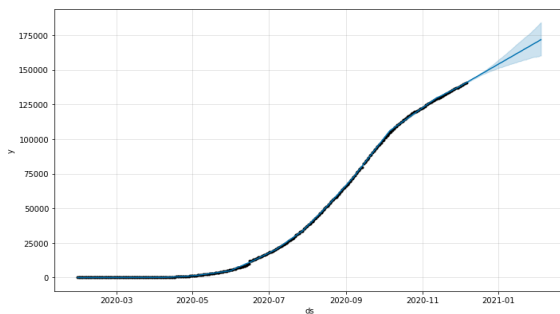
Prediction of mortality of any epidemic is the most important process as in the absence of permanent cure of an epidemic estimating mortality becomes one of the crucial steps. Same lies with COVID-19 as till now there is no permanent cure of COVID-19 will help the government to control it. We have predicted mortality due to COVID-19 till 04/02/2020 in India with the help of time series model. At first we showed predictions using the FB prophet Model. Below table which shows the prediction of mortality due to COVID-19 outbreak in India from 26/01/2021 to 04/02/2021 ds indicates Date/Time stamp, yhat shows the exact prediction, yhat\_lower shows lowest value of prediction while yhat\_upper shows maximum value of prediction. Table 7 shows that the number of confirmed COVID-19 cases in India based on the trend calculated by the model by 04/02/2020 would be 171720 as

exact predictions while 160206 lowest value of prediction and 184261 maximum value of prediction.

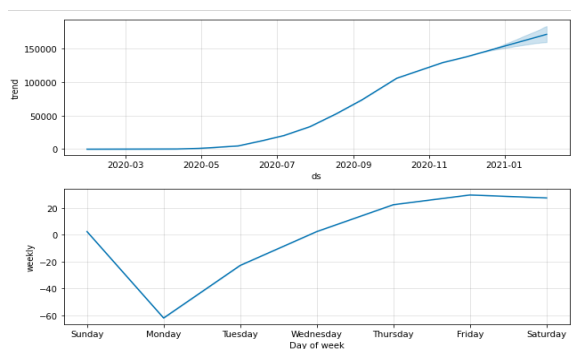
**Table 7. Predictions from 26/1/2021 to 04/02/2021 using Fb Prophet model**

Date Timestamp	Actual Prediction	Lower Approximation	Upper Approximation
2021-01-26	167008	158355	176294
2021-01-27	167551	158591	177003
2021-01-28	168090	158937	178023
2021-01-29	168615	159118	178893
2021-01-30	169132	159208	179588
2021-01-31	169626	159337	180633
2021-02-01	170080	159613	181498
2021-02-02	170638	159559	182343
2021-02-03	171181	160109	183430
2021-02-04	171720	160206	184261

The figure 11 indicates the relationship of the mortality due to COVID-19 outbreak in India with time where the x-axis shows time and y-axis shows the numerical value of mortality due to COVID-19. The black dotted line indicates original values and the blue solid line indicate the predicted values of mortality due to COVID-19 in India. Almost the predicted and the original values are parallel we have seen in figure 11.



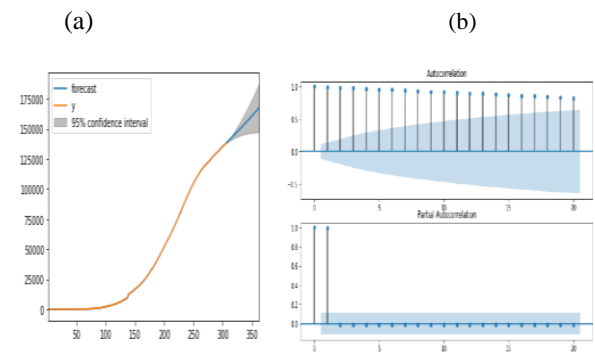
**Figure 12. Prediction of total number of mortality due to COVID-19 in India using FB Prophet Model till 04/02/2021**



**Figure 12. Representation of trend and weekly analysis**

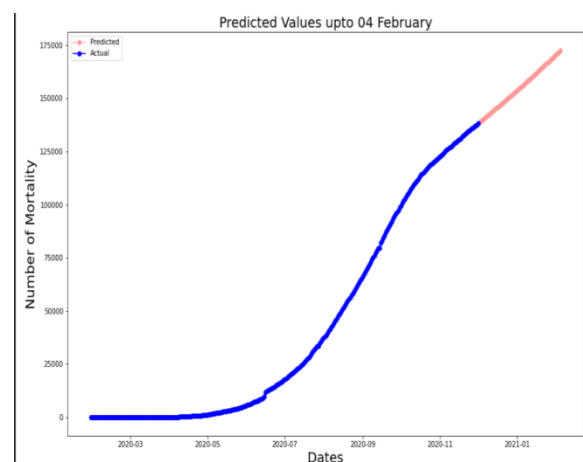
Figure 12 indicates the trend of our prediction of the mortality due to COVID-19 outbreak in India and weekly analysis of mortality due to COVID-19 in India till 04/02/2021. The weekly analysis was maximum on Friday and reached minimum on Monday as we have seen in the figure 12. Next we have shown prediction using the ARIMA

time series model. Autocorrelation function (ACF) graph and partial autocorrelation (PACF) as shown in figure 13(b) were used to estimate the parameters of ARIMA model. Through experimentation, the best fitted forecasting model using Akaike Information Criterion (AIC)[23] and the Bayesian Information Criterion (BIC)[23], we found that ARIMA(1,2,2) is best fitted for which both AIC and BIC are lowest (AIC=4032.89, BIC=4040.33) and its model summary is shown in table 8. Using this we have predicted total number of mortality due to COVID-19 in India upto 04/02/2020 shown in figure 13(a) which includes graphical representation with 95% CI (Confidence Interval). For better visualization we have shown prediction using python library called matplotlib shown in figure 14.



**Figure 13. (a) Prediction of mortality due to COVID-19 left side in which y shows actual number of mortality due to COVID-19 in India (b) ACF and PACF plots.**

We can evaluate from figure 14 that the exact number of predicted cases based on the data used are 172449 with 147440 lower limit value and 191730 higher limit value. To validate the accuracy of our models we have split our data into train set which contains the data till 1/12/2020 and test set which contains data from 2/12/2020 to 6/12/2020 is used to evaluate error metrics of our models. We have calculated error metrics RMSE, MAE, MAPE shown in table 9.



**Figure 14. Prediction of mortality due to COVID-19 in India ARIMA till 04/02/2020**



**Table 8 Model Summary ARIMA(1,2,2)**

ARIMA(1,2,2)	Std error	Z	P> z
AR(1)	0.089	9.061	0.000
MA(1)	0.087	-20.392	0.000
MA(2)	0.073	11.162	0.000

**Table 9 Evaluation of the Models on the basis of mortality due to COVID-19 in India**

Serial Number	MODEL	RMSE	MAE	MAPE
1	ARIMA	58	58	0.00042
2	FB Prophet	418.912	418.912	0.002993

## 5 CONCLUSION

In this paper, we focused on presenting the visualization of the current breakdown of the pandemic in India. Time-series prediction models such as the Facebook Prophet and ARIMA models are used to predict and analyze the total number of COVID-19 cases and total number mortality due to COVID-19 in India till 04/02/2021. The models are also used to estimate the minimum number of hospital beds required for the patients. The results of the two models are compared. Our investigation based on the experimentation, identified that the ARIMA model performs better than the Fb Prophet model except in predicting the total number of COVID-19 cases. These predictions clearly show that COVID-19 in India will increase at an alarming rate in the months of August and September. This study will allow the health authorities to make timely decisions and the government to efficiently control this pandemic in India.

## 6 REFERENCES

[1] Wang, Li-sheng, Wang, Yi-ru, Ye, Da-wei, Liu, Qing-quan, 2020b. A review of the 2019 Novel Coronavirus (COVID-19) based on current evidence. *Int. J. Antimicrob. Agents* <https://doi.org/10.1016/j.ijantimicag.2020.105948> (in press)

[2] WHO, 2020. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public>. Worldometers, 2020.

[3] Tomar A, Gupta N. Prediction for the spread of COVID-19 in India and effectiveness of preventive measures. *Science of The Total Environment*. 2020 Apr 20;138762.

[4] Zhao, S. and Chen, H., 2020. Modeling the epidemic dynamics and control of COVID-19 outbreak in China. *Quantitative Biology*, pp.1-9.

[5] Shah K, Alqudah MA, Jarad F, Abdeljawad T. Semi-analytical study of pine wilt disease model with convex

rate under caputo–febrizio fractional order derivative. *Chaos Solitons Fractals* 2020;135:109754.

[6] Jajarmi A, Arshad S, Baleanu D. A new fractional modelling and control strategy for the outbreak of dengue fever. *Physica A* 2019;535:122524.

[7] Jajarmi A, Yusuf A, Baleanu D, Inc M. A new fractional hrsv model and its optimal control: a non-singular operator approach. *Physica A* 2019:123860.

[8] Liestøl K, Andersen PK. Updating of covariates and choice of time origin in survival analysis: problems with vaguely defined disease states. *Stat Med* 2002;21(23):3701–14.

[9] Krätschmer V. Strong consistency of least-squares estimation in linear regression models with vague concepts. *J Multivar Anal* 2006;97(3):633–54.

[10] Bandyopadhyay, S. K., Dutta, S., Machine learning approach for confirmation of Covid-19 cases: positive, negative, death and release. medRxiv 2020.

[11] Makridakis S, Hibon M. ARMA models and the Box–Jenkins methodology. *Journal of Forecasting*. 1997 May;16(3):147-63.

[12] Solo V. The order of differencing in ARIMA models. *Journal of the American Statistical Association*. 1984 Dec 1; 79(388):916-21.

[13] Hyndman RJ, Athanasopoulos G. *Forecasting: principles and practice*. OTexts; 2018 May 8.

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

[15] Harvey AC, Shephard NG. *Estimation and testing of stochastic variance models*. Suntory and Toyota International Centres for Economics and Related Disciplines, LSE; 1993.

[16] Dataset link used in this paper. <https://www.kaggle.com/sudalairajkumar/novel-coronavirus-2019-dataset> .

[17] Dataset link used in this paper. <https://www.kaggle.com/sudalairajkumar/covid19-in-india>

[18] worldometer : <https://www.worldometers.info/coronavirus/>

[19] MoHFW website. Hospital in the country :Ministry of Health and Family Welfare

[20] WHO Media Statement : [https://www.who.int/indonesia/news/detail/08-03-2020-knowing-the-risk-for-COVID-19#:~:text=Most%20people%20\(about%2080,are%20at%20greater%20risk](https://www.who.int/indonesia/news/detail/08-03-2020-knowing-the-risk-for-COVID-19#:~:text=Most%20people%20(about%2080,are%20at%20greater%20risk).

[21] T. Chai , R. R. Draxler. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature.

[22] Ummul Khair, Hasanul Fahmi, Sarudin Al Hakim,Robbi Rahim. Forecasting Error Calculation with Mean Absolute Deviation and Mean Absolute Percentage Error

[23] Kenneth P. Burnham, David R. Anderson. Multimodel Inference: Understanding AIC and BIC in Model Selection.