International Journal of Future Generation Communication and Networking
Vol. 13, No. 4, (2020), pp. 2479–2488

COVID-19: Prediction of Confirmed Cases, Active Cases and Health Infrastructure Requirements for India

Shruti Chikode, Nutan Hindlekar, Pranali Padhye, Narayana Darapaneni*, Anwesh Reddy Paduri*
Great Learnings, Gurugram

chikode44@gmail.com, hindlekarnutan@gmail.com, padhyepranali@gmail.com, darapaneni@gmail.com, anwesh@greatlearning.in

Abstract

To develop a holistic model for COVID-19 forecasting to determine the hospitalization needed in each Indian state. The "covid19 in India dataset from kaggle has been used to study the pattern of infection spread and to forecast the number of cases of covid19 patients in coming months. The dataset has different excel files which include the data of the „"Confirmed", „"Cured" and "Death" count in each state of India. Also, the dataset includes the number of beds in each state which is helpful to compare the current active cases vs. number of beds. The two models which have been used for forecasting the COVID-19 are "Prophet" and "ARIMA". The outcome of this forecasting is then have been used to predict the number of beds would require providing the treatment to the infected people. The Prophet model has been implemented to forecast the number of beds required in high density states to accommodate the rising number of patient's treatment. The study shows us that the number of cases would be exponential in coming months in absence of the vaccine and the preventive measures such as lockdown, social distancing, etc.
The model gives us the prediction that the COVID cases are rising exponentially. The number of predicted cases of infected people could help the government and public healthcare facilities to increase their capacity of healthcare services to provide the best care to the infected people.

Keywords *– Covid-19, ARIMA, Prophet, Hospital beds forecasting, India*

1. Introduction

COVID-19 is a new strain of coronavirus that has not been previously identified in humans. This pandemic is the defining global health crisis of our time and the toughest challenge the world has faced since World War II. The spread of the virus is increasing globally on the daily basis and the numbers of cases are rising exponentially. Prime minister of India had declared lockdown all over the country to maintain the social distancing and to lower the rate of infection [2]. Government implemented some actions to achieve the goal of flattening the curve.
The COVID-19 is the cause of an outbreak of respiratory illness first detected in Wuhan, Hubei province, China. Since December 2019, cases have been identified in a growing number of countries. Coronaviruses are a large family of viruses that are known to cause illness ranging from the common cold to more severe diseases such as Severe Acute Respiratory syndrome (SARS) and Middle East Respiratory Syndrome (MERS). People suffering from COVID19 may face symptoms like common cold, runny nose, cough, sore throat, possibly a headache and maybe a fever, which can last for a couple of days. Depending on the patient's age, the death rate with SARS ranged from 0-50% of the cases, with older people being the most vulnerable. When it comes to human-to-human transmission of the viruses, often it happens when someone meets an infected person's secretions, such as droplets in a cough. Depending on how virulent the virus is, a cough, sneeze or handshake could cause exposure. The virus can also be transmitted by touching something an infected person has touched and then touching your mouth, nose, or eyes.

ISSN: 2233-7857 IJFGCN

So, our motive is to build a Forecasting model for new cases or incidence/infection rates by area to enable hospitals/health officials to better plan resourcing and lower the mortality rate by using Time series models named ARIMA and Prophet. Briefly, the research is categorized based on following requirements: 1. Private and Government sector beds, 2. ICU Beds, 3. Ventilators
Auto-Regressive Integrated Moving Average (ARIMA) One of the most well-known and widely used families of time-series models. Moving Averages can work better with stationary data. For a time, series that contains a trend or seasonal or non-stationary data, the forecasting technique that should be considered is Auto Regressive Integrated Moving (ARIMA). Here, in our case, ARIMA model has applied on COVID-19 data to forecast number of confirm cases that are currently infected. Also, by checking the trend of the existing data, this model has been used here for future predictions of confirmed cases (confirmed cases comprised of sum of Active, Recovered and Death). To determine whether the series is stationary or not, consider the graph of ACF. If a graph of ACF of the time series values either cuts off quickly or dies down quickly, then the time series values should be considered stationary. If a graph of ACF dies down extremely slowly, then the time series values should be considered non-stationary. If the series is not stationary, it can often be converted to a stationary series by differencing. That is, the original series is replaced by a series of differences. Differencing is done until a plot of the data indicates the series varies about a fixed level, and the graph of ACF either cuts off quickly or dies down quickly. This is done to get the values of three parameters (p,d,q). Here p indicates the order of the autoregressive part, d indicates the amount of differencing, and q indicates the order of the moving average part. If the original series is stationary, d = 0 and the ARIMA models reduce to the ARMA models.
To avoid such a lengthy method, we have used Auto ARIMA model here which works well irrespective of the trend of the data to give the values of (p,d,q). These values are then passed into ARIMA model to for future prediction of the data.

Such an approach comes with two clear benefits. Firstly, it offers a high level of interpretability, as, based on the assumptions of the model, the relationship between the independent variables and the dependent variables are well-understood and therefore easily explained. The second benefit concerns model selection, which for ARIMA models can be performed in an automated way to maximize prediction accuracy. Another benefit of ARIMA models is their ability to accommodate systems governed by dynamics that change over time by updating the model based on recent events to predict the future state of the system. However, since the ARIMA models cannot deal with nonlinear patterns or relationships, their approximation of complex real-world problems and dynamics is not always satisfactory.

Prophet is a time series forecasting model developed by Facebook. The method uses an easily decomposable time-series model consisting of three main components: trend, seasonality, and holidays. In its core, it is regression model or linear model with interpretable parameters, while also allowing the user to choose the components that concern their forecasting problem and effortlessly apply the necessary adjustments. To forecast trend, prophet uses two models: a saturating growth model and a piece-wise linear model. To capture seasonality, prophet relies on Fourier series to provide a flexible model of periodic effects, while, to account for holidays, a predefined list of past and future holiday events is required. Here, Prophet Model is used to predict the current infected cases of COVID-19. Also, it has used for future predictions of the same. It can be conveniently used for future predictions as it gives range of forecasted values as yhat_lower, yhat, yhat_upper where yhat is prediction of target variable (which is COVID-19 infected individuals in our case). Also, it worked well while calculating the infected cases state-wise which further used for calculation of the hospital facilities (hospital beds, ICU and ventilator count) for every state individually.

2. Materials and Methods

ISSN: 2233-7857 IJFGCN

As explained in the introduction, the purpose of this study is to build a holistic model for forecasting the COVID-19 infection rate and to predict the number of beds which would be required to provide the treatment for the patients who are in critical or severe situation.

2.1 Dataset

We are working on COVID19 India dataset which is live data and updated regularly by Kaggle. we have got the updated data regarding COVID19 active, recovered, and deceased individuals on daily basis.

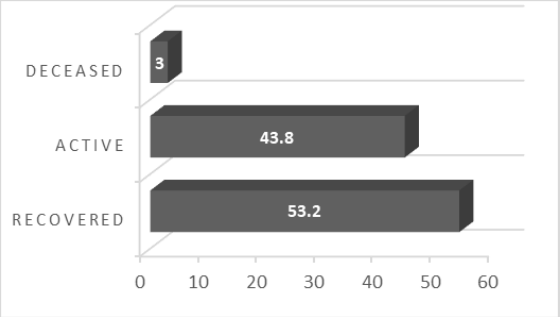

Figure 1. Current Contribution by Active, Recovered and Deceased Cases

Also, we have collected information regarding hospitalization facilities including hospital beds, ICUs and ventilators. The information provided is for public and private section. Below is the comparison for both the sectors.

Public Sector	***ICU Beds***	***25.1%***
(37.6%)	*Ventilators*	*12.5%*
Private Sector	*ICU Beds*	*41.6%*
(62.4%)	*Ventilators*	*20.8%*

Table 1. ICU/Ventilators Public V/s Private Sector

2.2 COVID-19 Forecasting Models

Here 2 models have been used to forecast the COVID-19 cases till August 15th.

1. ARIMA Model

2. Prophet Model

2.2.1 ARIMA Model

Autoregressive Integrated Moving Average (ARIMA) time-series model used for the forecasting of new COVID-19 Cases [4]. For ARIMA, we need data stationary to get the accurate predictions. If the data is non-stationary (with upward or downward trends), then we need to make it stationary either by rolling or log transformation. We can perform Dickey Fuller test to check if the series is stationary. ARIMA model represented with the help of some parameters, and expressed as,

ARIMA (p, d, q)

Where p, d and q are non-negative integers.

p: stands for the order of auto-regression,

ISSN: 2233-7857 IJFGCN

d: represents the degree of trend difference (the number of times the data have had past values subtracted) for the stationary of the trend

q: signifies the order of moving average.

From the stationary data, Our next task is to find out the values of above mentioned three parameters (p,d,q). Once got the values of p,d,q, fit ARIMA model on train data. Forecast the test data values and plot the visualization of train, test, and predicted test data values on single graph.

Here, we have used Auto-ARIMA model which worked well on the existing data (Irrespective of its trend). This model analyzed the nature of data and its trend to give the correct values of p,d,q. Using these values in the ARIMA model, we made the prediction of current and future COVID19 infected individuals and got the prediction close to the actual values.

(Source used for comparison: https://www.covid19india.org/).

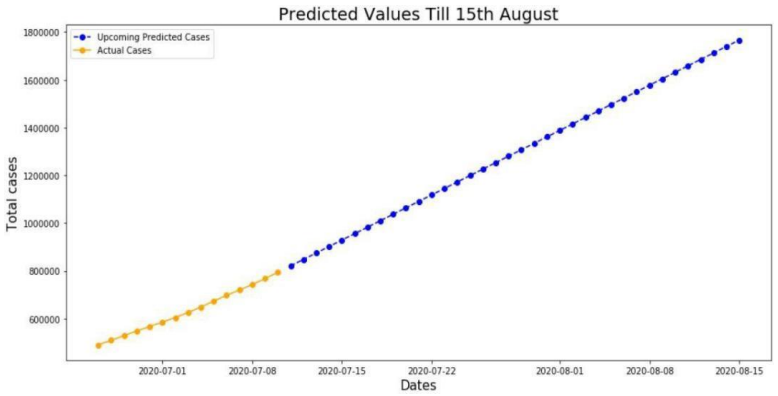

Figure 2. Predicted Values till 15th August

Prediction of confirmed cases given by ARIMA on 15th August: 17.66 Lakhs

2.2.2 The Prophet Forecasting Model

Prophet was developed for the purpose of creating the high-quality forecasting [6]. It addresses the common business time series difficulties such as seasonal effects caused by human behavior, changes in trends and outliers. This library utilizes the additive regression model y(t) comprising the following components:

$$y(t)=g(t)+s(t)+h(t)+\epsilon t, \quad (1)$$

where:

Trend g(t)g(t) models non-periodic changes.

Seasonality s(t)s(t) represents periodic changes.

Holidays component h(t)h(t) contributes information about holidays and events.

Prophet allows analysts to manually set change points of the growth rate and capacity values at different points in time. For instance, analysts may have insights about dates of past releases that prominently influenced some key product indicators. This trend model is a simple Piecewise Linear Model with a constant rate of growth. It is best suited for problems without saturating growth.

ISSN: 2233-7857 IJFGCN

We have used linear curve to predict the spread of COVID-19. The curve exhibit exponential growth period and subsequently reaches a maximum level. The pattern followed by COVID-19 is similar to this curve as large number of people gets 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 modeled as a piecewise linear or as a logistic growth curve. The motivation of this analysis is to predict the future cases of COVID-19.

Implementation Steps:

Confirmed cases Prediction:

The data frame used in Prophet required two columns-(i) "ds": to store date time series and (ii)"y": to store the corresponding values of the time series. The linear curve is then fit to the dataset using these data frame and it is then compared to the actual values of confirmed & active cases at each time instant. From the forecast curve (Figure ***2.2.2(1)***) it can be inferred that the growth of the COVID-19 in India is expected to increase after Aug 15th.

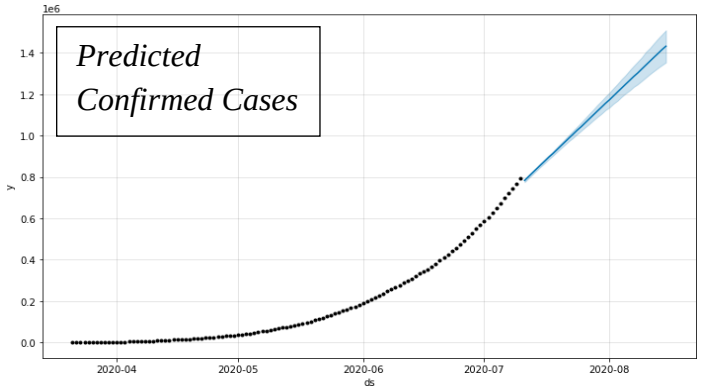

Figure 3. Predicted Confirmed Cases

In order to evaluate the model-fit, we use dataset form 20th March, after lockdown to get proper growth rate. From 20th March to 30th June dataset were used to train the model and remaining days were used to evaluate and cross-validate the model.

ds	yhat	yhat_lower	yhat_upper
11-08-2020	1410751.361	1358658.291	1463942.442
12-08-2020	1430457.945	1372530.289	1487957.236
13-08-2020	1450470.265	1390273.765	1510867.899
14-08-2020	1470788.373	1406907.128	1533850.067
15-08-2020	1490967.076	1424054.612	1558935.697

Table 2. Predicted Confirmed Cases up to 15th Aug

As per Prophet Model We can see that by 15th Aug 2020 more than 14,90,967L confirmed cases are predicted and upper limit of around 15,58,935L.

Active cases Prediction:

ISSN: 2233-7857 IJFGCN

After confirmed cases prediction, the Prophet Model is then fit to the state wise dataset and predicts active cases state wise to predict hospital beds, ICU and ventilators requirement.

Predicted Active Cases State wise up to 15th Aug:

State	Date	predicted_Active_cases
Andhra Pradesh	15-08-2020	24542
Chhattisgarh	15-08-2020	1133
Delhi	15-08-2020	47124
Gujarat	15-08-2020	9527
Haryana	15-08-2020	8588
HP	15-08-2020	549
Karnataka	15-08-2020	25101
Kerala	15-08-2020	4100
Madhya Pradesh	15-08-2020	2913
Maharashtra	15-08-2020	153668
Odisha	15-08-2020	5360
Puducherry	15-08-2020	1207
Punjab	15-08-2020	1995
Rajasthan	15-08-2020	4917
Tamil Nadu	15-08-2020	92079
Telangana	15-08-2020	27842
Chandigarh	15-08-2020	118
J & K	15-08-2020	5023
Ladakh	15-08-2020	533
Uttar Pradesh	15-08-2020	14537
Uttarakhand	15-08-2020	1154
West Bengal	15-08-2020	10801
Bihar	15-08-2020	4283
Manipur	15-08-2020	1440
Mizoram	15-08-2020	120
Andaman / Nicobar	15-08-2020	112
Goa	15-08-2020	1656
Cases being reassigned to states	15-08-2020	11032
Assam	15-08-2020	6320
Jharkhand	15-08-2020	1173
Arunachal Pradesh	15-08-2020	335
Tripura	15-08-2020	745
Nagaland	15-08-2020	371
Meghalaya	15-08-2020	24
Daman and Diu	15-08-2020	510
Sikkim	15-08-2020	126

Table 3. Predicted Active Cases up to 15th Aug

2.3 Hospital Beds Forecasting.

ISSN: 2233-7857 IJFGCN

For forecasting the hospital beds, the prophet gave much accurate results than ARIMA.

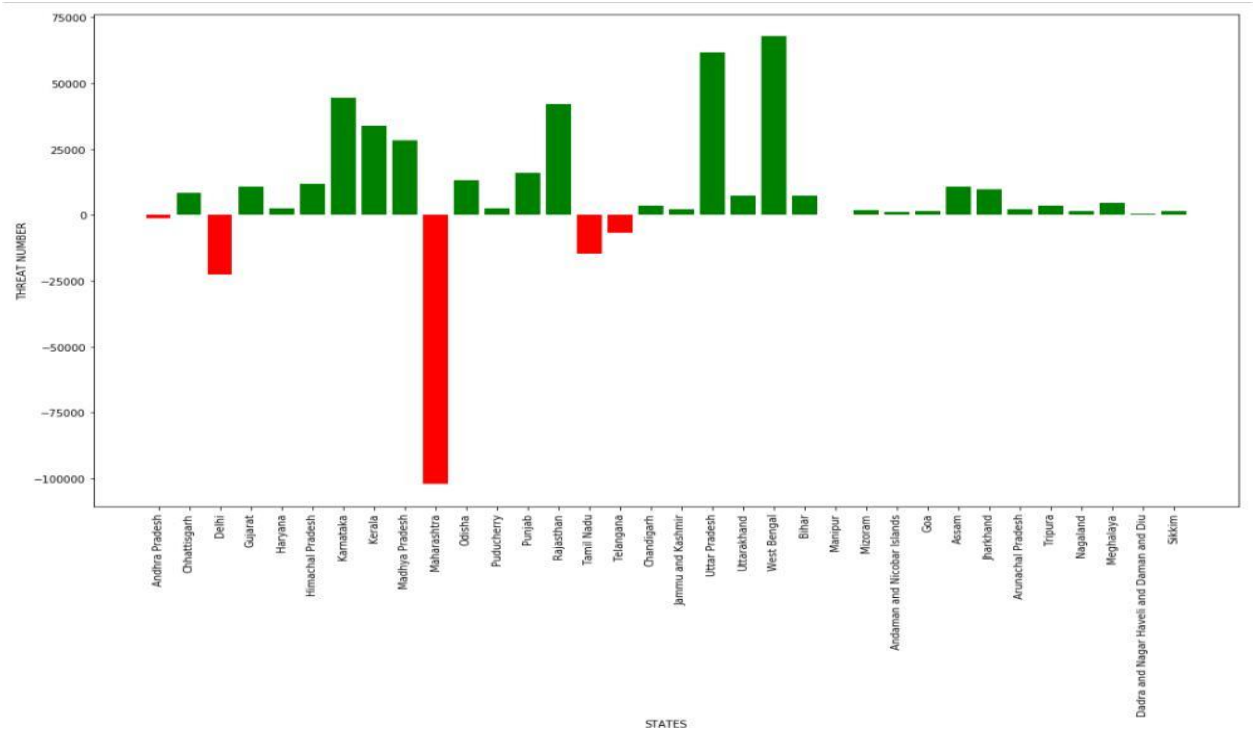

Figure 4. Hospital Beds V/s State

For Hospital forecasting we assume that total of **80% of beds in public sector** and **20% beds in private sector** will be allocated to treat the COVID-19 infected patients. Also as per the observation it seems not each infected person need to be hospitalized. The asymptomatic or mild cases will be home quarantined to reserve the beds for more critical or sever cases. We assumed that **60%** of infected people would require to be admitting in hospital and occupying either the bed, ICU or ventilator. It has been observed that only **10%** of infected patients need to be admitted in the **ICUs** and **5%** people would require the ventilators. The remaining patients will be occupying the general beds.

The data of forecasted infected cases for August 15 and the number of required beds according to the trend for each state is summarized in below table.

3. Result and Discussion

As per the above shown Figure 2.3.1, Prophet Model has calculated the predicted number of hospitals required for the upcoming COVID19 infected cases. **Green Bar** indicates that numbers of hospital beds present in that region are more than number of infected cases whereas **Red Bar** shows number of infected individuals exceeds the bed count in those respective regions [7].

Although, prophet model has predicted hospital bed counts for all the states, priority would be given to the regions in the Red zone as per the infected number of cases.

As per Prophet Model below State wise active cases & hospital requirements are predicted. Highlighted states required more beds/ICU/ventilators as per exponential growth rate from the linear curve it can be inferred that the numbers of confirmed cases are going to increase. Therefore, we need

ISSN: 2233-7857 IJFGCN

to take care of the preventive measures [8]. The maximum number of people that can be infected by this virus as per this model are 15,05,809L up to 15th August 2020.

This means that if the conditions are out of control then the growth of the COVID-19 will be increased and we will face shortage of medical facilities.

Number of hospital facilities required in future predicted by the Prophet model can be summarized in below table:

State	Current Caes_12Jul	Future Cases_15Aug	Total available beds	Total available ICUs	Total available ventilators	Future required beds	Future required ICUs	Future required ventilators
Andhra Pradesh	12026	24542	1527	763	30529	667	-11879	334
Chhattisgarh	794	1133	457	228	9133	-341	-8150	-170
Delhi	31435	47124	1126	563	22521	3588	17544	1793
Gujarat	7880	9527	1254	627	25076	-210	-16198	-104
Haryana	5475	8588	699	349	13972	145	-6800	72
HP	369	549	532	266	10647	-477	-10179	-239
Karnataka	12601	25101	4713	2356	94254	-2450	-75026	-1225
Kerala	2545	4100	2132	1066	42648	-1733	-39259	-867
Madhya Pradesh	2805	2913	1582	792	31651	-1239	-28735	-620
Maharashtra	94689	153668	3861	1930	77215	11041	49447	5521
Odisha	2989	5360	812	406	16241	-311	-11983	-156
Puducherry	586	1207	158	79	3176	-51	-2261	-25
Punjab	1457	1995	1148	574	22959	-941	-21199	-470
Rajasthan	3832	4917	2344	1171	46868	-1819	-42406	-909
Tamil Nadu	51133	92079	3880	1940	77594	4713	-4556	2357
Telangana	13297	27842	1629	815	32574	840	-11593	420
Chandigarh	96	118	169	85	3380	-156	-3271	-78
J & K	3465	5023	299	149	5974	210	-1645	105
Uttar Pradesh	9140	14537	5102	2551	102036	-3681	-89961	-1841
Uttarakhand	835	1154	494	247	9876	-375	-8861	-187
West Bengal	7348	10801	3492	1746	69847	-2403	-60591	-1202
Bihar	3079	4283	658	330	13170	-217	-9415	-109
Manipur	831	1440	60	31	1214	76	-56	38
Mizoram	83	120	85	42	1697	-73	-1594	-36
Andaman/nicobar	56	112	45	23	904	-35	-818	-18
Goa	967	1656	137	68	2724	21	-1386	11
Assam	3901	6320	756	378	15121	-143	-9908	-72
Jharkhand	820	1173	588	295	11770	-469	-	-235

ISSN: 2233-7857 IJFGCN

							10753	
Arunachal	193	335	98	49	1967	-67	-1698	-33
Tripura	530	745	179	90	3591	-103	-2941	-52
Nagaland	244	371	82	41	1640	-45	-1327	-23
Meghalaya	18	24	186	93	3723	-184	-3702	-92
Daman and Diu	230	510	48	23	954	-3	-576	-1
Sikkim	70	126	66	33	1326	-55	-1226	-27

Table 4. Forecasted number of Health Facilities

	ARIMA	Prophet
Mean Squared Error	1600610122.17	15338371964.15
Root Mean Squared Error	40007.63	123848.18
Mean Absolute Deviation	28278.92	109359.41
Mean Absolute Percentage Error	2.38%	9.77%

Table 5. Model performance

4. Conclusion

In this project, the prediction provides a simple way to understand the spread of COVID-19 and to predict new cases or infection rates by states to enable hospitals/health officials to better plan resourcing and lower the mortality rate. In this study, two different machine learning-inspired time series methods were developed to predict the infected cases. The prediction of confirmed cases was carried out by using ARIMA & Prophet model and results indicates that ARIMA gives better accuracy as compare to prophet model for forecasting the future cases [9]. The complexity of Prophet Model while working state wise is much lower than that of ARIMA Model. As in ARIMA Model the calculation of the p, d, q values changes state wise due to change in the trend of the data. So we used Prophet Model to predict state wise hospital beds, ICU and Ventilator count.

Prediction has been made up to 15th Aug, which shows the exponential growth of confirmed cases. As the number of cases will be increasing exponentially, we would definitely be able to see the shortage of healthcare facilities [10]. Andhra Pradesh, Delhi, Maharashtra, Tamil Nadu, and Telangana needs more health facility requirement in future as per prediction in figure (3.1). Especially Maharashtra is at high risk due to highest number of COVID-19 cases. As per updated news, Non-government health care providers in Maharashtra have been asked to accommodate the maximum number of patients with 80 per cent of the operational bed capacity & also Delhi opens makeshift 10,000-bed hospital amid coronavirus surge, link is given in reference. As government of two major states in India is arranging beds for treating patients of their respective regions, hopefully there would be significant decrease in the infection rate. There might be a possibility of shortage of hospital facilities as per the current growth rate of infected cases and some preventive measures are needed to be taken for the same [11].

ISSN: 2233-7857 IJFGCN

5. References

[1] “covid19-in-india,” https://www.kaggle.com/sudalairajkumar/covid19-in-india.

[2] “India Lockdown news: India to be under complete lockdown for 21 days starting midnight: Narendra Modi,” Indiatimes.com. [Online]. Available: https://economictimes.indiatimes.com/news/politics-and-nation/india-will-be-under-complete-lockdown-starting-midnight-narendra-modi/articleshow/74796908.cms. [Accessed: 29-Aug-2020].

[3] G. Kapoor et al., “State-wise estimates of current hospital beds, intensive care unit (ICU) beds and ventilators in India: Are we prepared for a surge in COVID-19 hospitalizations?,” medRxiv, p. 2020.06.16.20132787, 2020.

[4] L. Bayyurt and B. Bayyurt, “Forecasting of COVID-19 cases and deaths using ARIMA models,” medRxiv, p. 2020.04.17.20069237, 2020.

[5] H. Tandon, P. Ranjan, T. Chakraborty, and V. Suhag, “Coronavirus (COVID-19): ARIMA based time-series analysis to forecast near future,” arXiv [q-bio.PE], 2020.

[6] N. Phutela, A. G. Bakshi, S. Gupta, and G. Gabrani, “Forecasting the stability of COVID-19 on Indian dataset with prophet logistic growth model,” Research Square, 2020.

[7] S. Rajagopalan and A. Choutagunta, “Assessing healthcare capacity in India,” SSRN Electron. J., 2020.

[8] V. Papastefanopoulos, P. Linardatos, and S. Kotsiantis, “COVID-19: A comparison of time series methods to forecast percentage of active cases per population,” Appl. Sci. (Basel), vol. 10, no. 11, p. 3880, 2020.

[9] P. Wang, X. Zheng, J. Li, and B. Zhu, “Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics,” Chaos Solitons Fractals, vol. 139, p. 110058, 2020.

[10] M. L. Ranney, V. Griffeth, and A. K. Jha, “Critical supply shortages - the need for ventilators and personal protective equipment during the covid-19 pandemic,” N. Engl. J. Med., vol. 382, no. 18, p. e41, 2020.

[11] CDC, “Coronavirus Disease 2019 (COVID-19),” Cdc.gov, 22-Jul-2020. [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/hcp/mitigating-staff-shortages.html. [Accessed: 29-Aug-2020].

ISSN: 2233-7857 IJFGCN