Predicting the Spread of the Corona Virus Disease Requires Analyzing Data from Cases across Multiple States in India

Bande Ganesh[1*], V Ravinder naik[2], Banothu Ramji[3], Veerender Aerranagula[4], B. Sankara Babu[5] and Sandeep Sharma[6]
[1]Dept. of CSE, KG Reddy College of Engineering & Technology, Hyderabad, Telangana, India.

[2] Dept. of CSE-AI&ML, CMR Technical Campus, Kandlakoya, Hyderabad, Telangana, India
[3,4] Dept. of CSE-Data Science, CMR Technical Campus, Kandlakoya, Hyderabad, Telangana, India
[5]Professor, Department of Computer Science and Engineering, GRIET, Bachupally, Hyderabad, Telangana
[6]Uttaranchal Instiiute of Technology, Uttaranchal University, Dehradun, 248007

Abstract. Data analysis is very sophisticated tool in recent corona virus pandemic to find the trend of spreading pattern for controlling the infection. In this perspective, predictive analytics can be useful for data analysis to forecast the corona virus pandemic. This paper presents the infection pattern of corona virus disease, termed as COVID-19 in top seven states in India. Prophet Algorithm forecasting model was used to analyze state-wise spreading pattern of corona virus disease with respect to confirmed, deaths and cured cases. This predictive model can be very helpful to government and health care communities to combat this deadly virus by initiating suitable actions to control its spread.

Keyword: COVID-19, Predictive analytics, Forecasting model, Prophet Algorithm.

1 INTRODUCTION

The world is experiencing a significant challenge due to the pandemic outbreak of corona virus disease causing global health crisis. The corona virus is termed as COVID-19 by the World Health Organization (WHO).This pandemic infection is caused by the extreme critical respiratory disorder corona virus 2 (SARS-CoV-2) [1]. WHO had confirmed COVID-19 as pandemic worldwide on March 11, 2020 [2]. Corona viruses generally target the upper and lower parts of human respiratory systems, causing a range of illness symptoms such as fever, tiredness, dry cough, fatigue, headache, nasal congestion, loss of taste and smell, difficulty in breathing, muscle and joint pains, sore throat, sputum

*Corresponding author: gani1255@gmail.com

production, chills, vomiting and diarrheal [3]. The COVID-19 outbreak was first originated from Wuhan city, China in the month of December, 2019 [4, 5].Since its emergence, the virus has spread at an alarming pace worldwide causing an unprecedented socio-economic crisis. Until 2September 2020, corona virus disease has affected 26468031victims in 215nations and has become an utmost worldwide health concern [6]. The first confirmed case of COVID-19 was noticed in India on January 30, 2020 in the state of Kerala and afterward COVID-19 cases were rising at a rapid pace in India with a total number of 4023179 cases until5th September 2020 [7, 23]. Therefore, it is very important to have better understanding of its epidemic nature for further prevention and control. In this regard, epidemiological information's need to be examined so that the investigative data analysis will help the situational consciousness among the people in the society to prevent corona virus spread [8].This situation emphasizes the need for a reliable and accurate model that can not only fit the existing data well but can also predict with considerable accuracy [9, 13].Here, we present predictive model by analyzing the data of Indian COVID-19 scenario from 26 February 2020 till 2September 2020.This study signifies state-wise spreading pattern of corona virus disease in India by analyzing each state-wise confirmed, cured and death cases with the help of time series based forecasting model like Prophet algorithm. This predictive model can be of great help for the policy-makers, researchers, scientists, and medical professionals to combat this deadly virus and control its spread in future.

2 MATERIALS AND METHODS

The dataset available in the Ministry of Health & Family Welfare, India [14], COVID-19 tracker [15, 21] and open dataset of COVID-19 provided by Kaggle site [16] have been used for analysis. The dataset contains state-wise infected cases of patients in India. Top seven infected states such as Maharashtra (MH), Andhra Pradesh (AP), Tamil Nadu (TN), Karnataka (KR), Delhi (DL), Uttar Pradesh (UP), West Bengal (WB) were chosen to study the infected cases from 26th February 2020 till 2September 2020. This dataset consists of reported numbers for confirmed, recovered and death cases. The active cases are calculated by subtracting the recovery and death numbers from the confirmed numbers. The insight of dataset with corresponding column description is presented in Table 1. The statistical models are used in Python language in Jupiter platform to analyze the cases for the purpose of prediction. By using Prophet Model, the data is analyzed to forecast for next 90 days. It is interesting to note that Prophet is an open library data-source developed by the core team in data-science of facebook, which is based on trend, seasonality and holiday's effects using the time series forecasting model [17]. The Prophet algorithm is very fast which fits the model within seconds and gives reliable output with great accuracy [18, 20]. The package of Prophet offers instinctive parameters which are very simple to use, even the non-expert in predictive models can operate this for important predictions. Prophet is implemented by facebook in Python and R to generate the forecasting results [19].The three primary sections of Prophet Forecasting model are given by (a) trend, (b) seasonality and (c) holidays [20].

The Following equations used:

$$Y(t) = g(t) + s(t) + h(t) + \varepsilon t$$

Where,
g(t)- represents the model trend,

s(t)- represents the seasonality model (periodic changes such as weekly/yearly seasonal components).
h(t)- represents the holidays/events model (the irregular schedules of effects of holidays or events),
(εt)- represents the error term (for any unusual changes error term occurs which is not fit by the model).
The outcome of predict function is represented by the Data Frame that consists of several columns. Out of these columns, the forecasted value (‘yhat‘), the forecast date time (‘ds‘) and the lower and upper bounds on the predicted value (‘yhat_lower‘ and ‘yhat_upper‘) are the most essential columns.

3 RESULTS AND DISCUSSION

We have analyzed the data provided in Table 1to visualize the occurrence of COVID-19 across different states in India. The dataset provides the information on confirmed, death, cured and active cases which were reported from 26February 2020 till 2September 2020.It is obvious from the data in Table 1 that Maharashtra (MH)registered the maximum confirmed cases followed by Andhra Pradesh(AP), Tamil Nadu(TN), Karnataka(KR), Uttar Pradesh(UP), Delhi(DL) and West Bengal(WB). Interestingly, Maharashtra, Karnataka, Tamil Nadu, Andhra Pradesh are neighbor states and Uttar Pradesh, Delhi are also neighbor states. However, more cases of recovery were reported than death in every state. For example, there was recovery of 1104426 patients in Maharashtra as compared to 37056 deaths.The highest number of deaths (37056) was reported in Maharashtra state followed by Tamil Nadu (9586), Karnataka (8994), Andhra Pradesh (5889), Uttar Pradesh (5864), Delhi (5401) and West Bengal (5017). Figure 1 shows the heat map of the confirmed cases all over India. This map tracks the total number of confirmed corona virus patients across different states in India. It is clearly seen that the COVID-19 cases have propagated rapidly almost all the states in India. Figure 2 shows the different classes of data for confirmed, death and recovery cases in India during 26February to 2 September, 2020. Further, top seven infected states have been chosen for comparative analysis to comprehend the state-wise specific numbers of confirmed, death and recovered cases of patients reported during a specific time (Figure 3 and Table 1).The recovery and mortality rates are calculated as,

$$\text{Recovery rate(RR)} = \frac{\text{No. of recovered cases}}{\text{No. of confirmed cases}} \times 100$$

$$\text{Mortality rate(RR)} = \frac{\text{No. of death cases}}{\text{No. of confirmed cases}} \times 100$$

According to Figure 3 and Table 1,Maharashtra registered the highest cases of confirmed, recovered and death patients in India until September 2. The next higher confirmed cases were found in Andhra Pradesh (AP) followed by Karnataka (KR) and Tamil Nadu (TN).The recovery rates for Andhra Pradesh (AP) and Tamil Nadu (TN) were found around 91%.In case of Karnataka state the recovery rate was 80.4 %. Interestingly, Tamil Nadu (TN), Karnataka (KR) and Andhra Pradesh (AP), are neighboring states. The two neighbor states viz. Uttar Pradesh (UP) and Delhi (DL) situated in the northern parts of India, registered fifth and sixth highest number of confirmed cases. The recovery rate for Uttar

Pradesh was 86 % and that for Delhi was 88.6%.The state in the eastern part of India, West Bengal (WB) declared the seventh highest figure of confirmed cases of patients and the recovery rate in this state was found 87.8%. The pattern of the curves of confirmed and recovery cases initially overlapped and then jumped at increasing rate from the month of July. The mortality rates were observed in the following states as Maharashtra (2.6 %), Delhi (1.9 %), West Bengal (1.9 %), Tamil Nadu (1.6 %), Uttar Pradesh (1.5 %), Karnataka (1.5 %), Andhra Pradesh (0.8 %).

Additionally, FB Prophet [19] algorithm was used as the paradigm for forecasting number of contaminated patients in next three months, if the respective state governments do not take adequate safety measures like social distancing and lockdown. Prediction of spread and mortality rates are very crucial in epidemic syndrome like COVID-19, because preventive actions can be implemented by the government, public, and healthcare systems based on this prediction. FB Prophet Algorithm works accurately for data series of time, which can be used for forecasting. Figures 4-6 show the states-wise prediction of confirmed cases (Figure 4), recovered cases (Figure 5) and death cases (Figure 6) from September 2020 to January 2021. This prediction results would be useful for the states governments to instigate proper control measures in time to curb this epidemic. In all these figures (Figures 4-6), X-coordinate denotes the time span and Y-coordinate shows the respective cases of confirmed, cured and death patients. The predicted values of confirmed cases would be: Maharashtra (2.18 million), Andhra Pradesh (1.59 million), Tamil Nadu (1.13 million), Karnataka (1.17 million), West Bengal (0.55 million), Uttar Pradesh (0.75 million), Delhi (0.31 million).The predicted values of cured cases would be: Maharashtra (1.8 million), Andhra Pradesh (1.36 million), Tamil Nadu (1.1 million), Karnataka (0.97 million),West Bengal (0.49 million), Uttar Pradesh (0.58 million), Delhi (0.28 million). The predicted values of death cases would be: Maharashtra (62396), Tamil Nadu (20501), Karnataka (18448), Andhra Pradesh (14384), Uttar Pradesh (10006), West Bengal (10114), Delhi (6258). The analysis shows that the deaths would increase in large number. Hence, state governments should take proper control measures on time to curb this COVID-19 epidemic.

Table 1: State-wise tabular lists of confirmed, recovered, death and active cases in India till 2nd September, 2020

S. No.	State Name	Confirmed	Recovered	Deaths	Active
1	Maharashtra	1400922	1104426	37056	259006
2	Andhra Pradesh	700235	636508	5869	57858
3	Tamil Nadu	603290	547335	9586	46369
4	Karnataka	611837	492412	8994	110412
5	Uttar Pradesh	403101	346859	5864	50378
6	Delhi	282752	250613	5401	26738
7	West Bengal	260324	228755	5017	26552

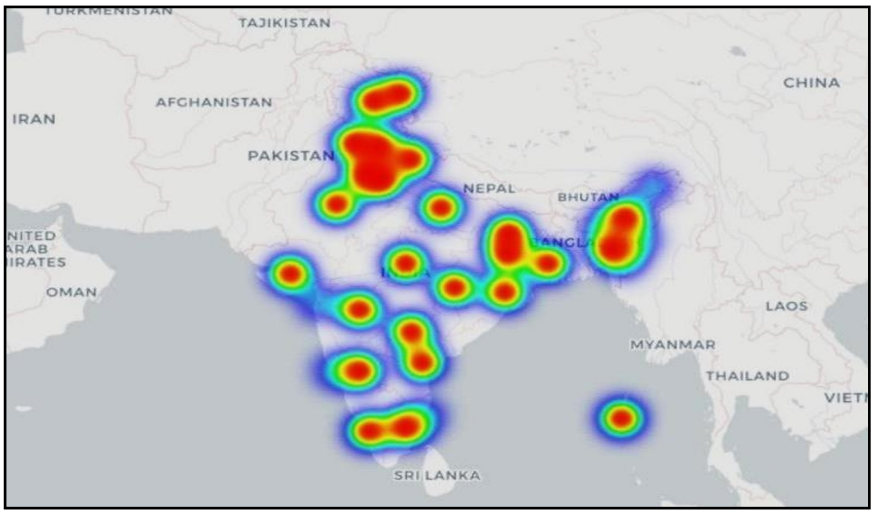


Fig 1: The confirmed COVID-19 cases across various states in India represented by heat map

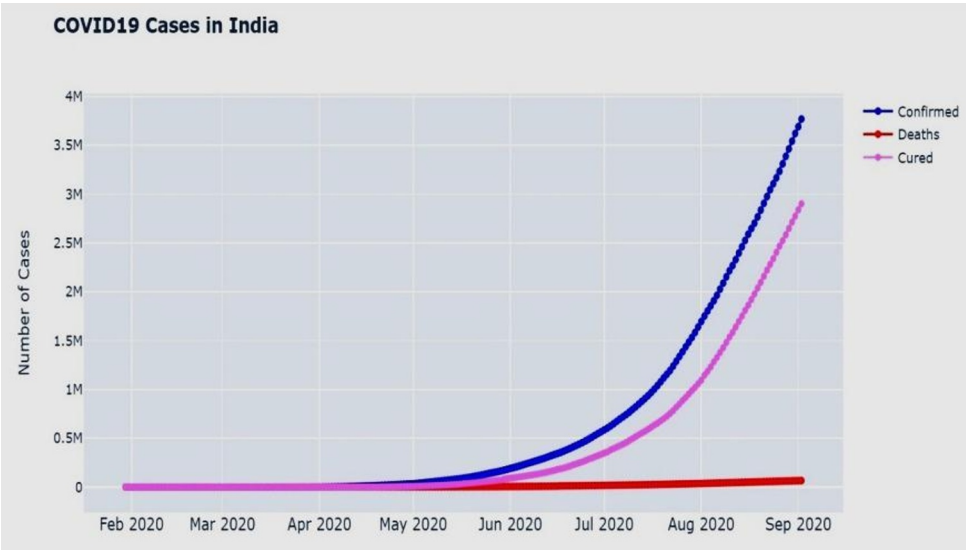


Fig 2: The COVID-19 scenario in India represented date-wise from 26th February to 2nd September, 2020.

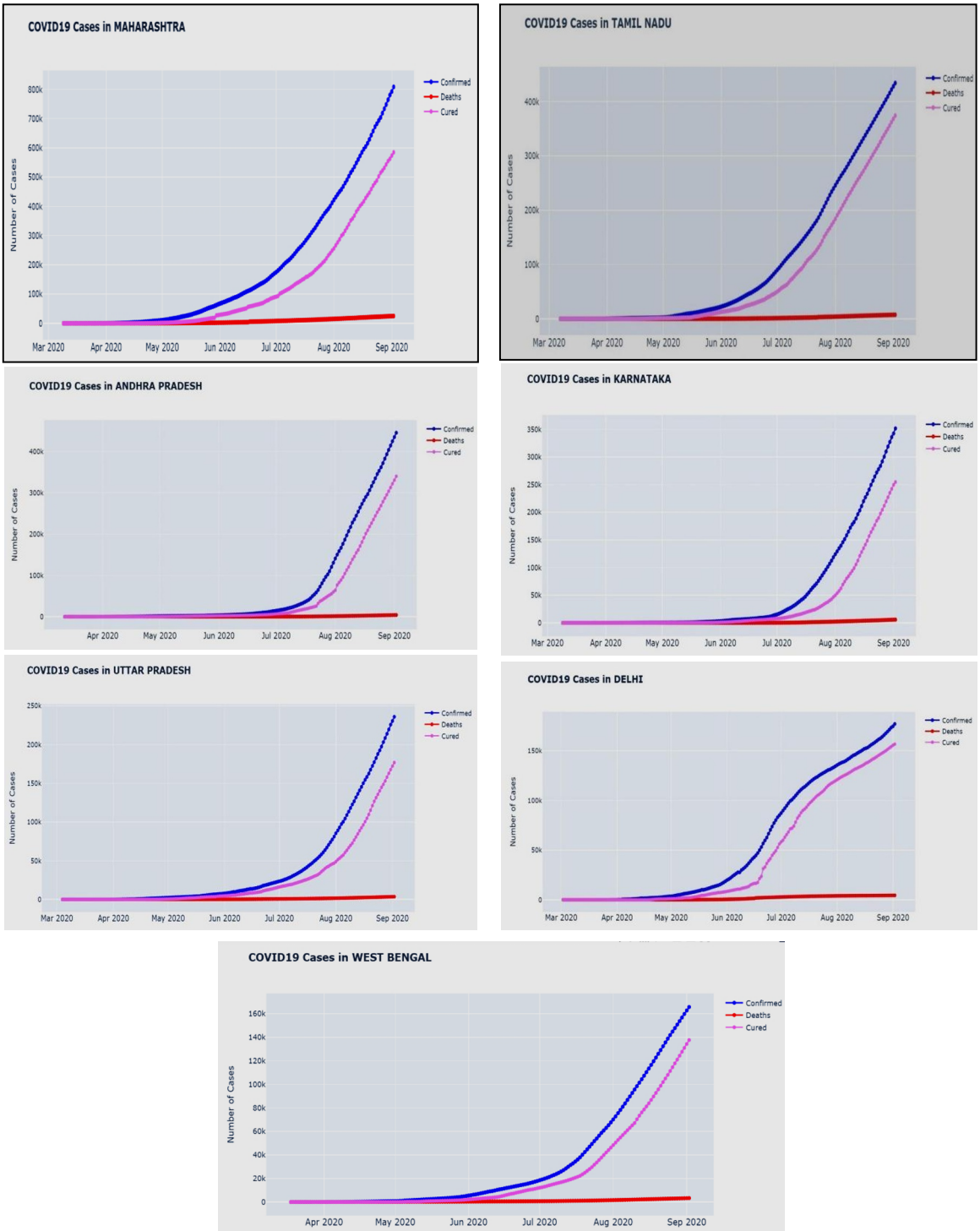
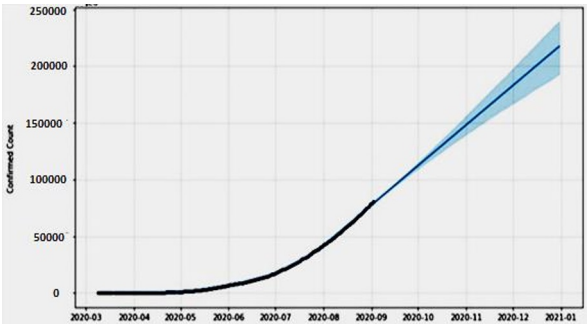
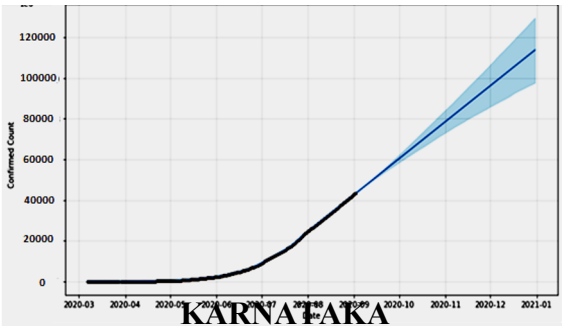


Fig 3: The confirmed, death and cured COVID-19 cases from top seven state-wise data until 2nd September 2020

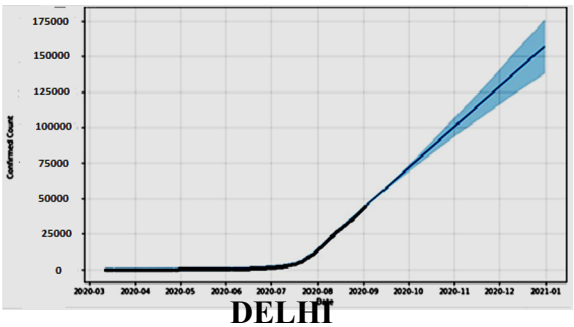
MAHARASHTRA



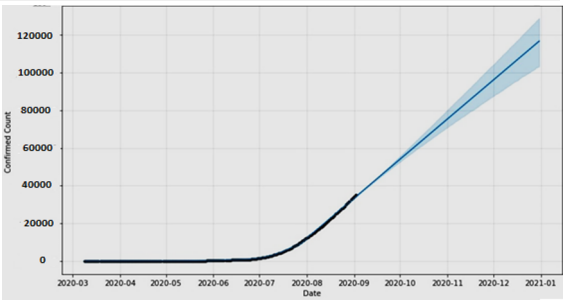
TAMIL NADU



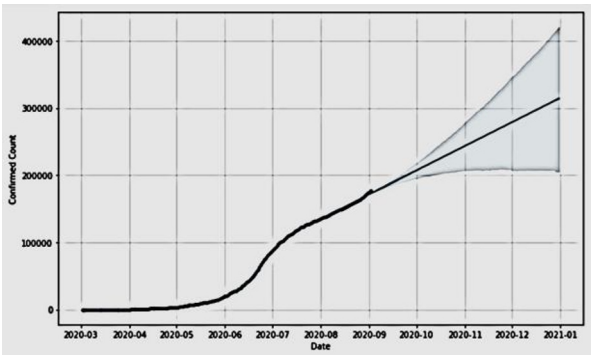
ANDHRA PRADESH



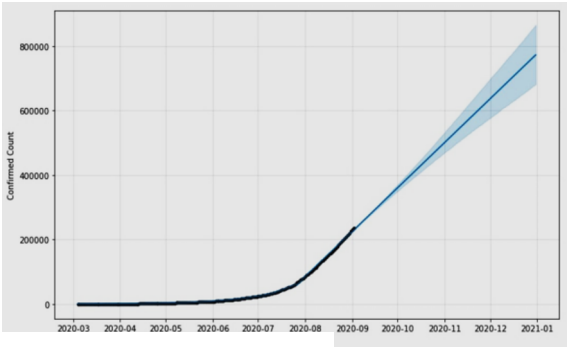
KARNATAKA



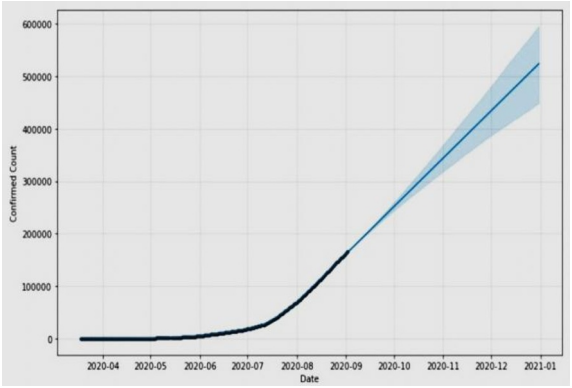
DELHI



UTTAR PRADESH

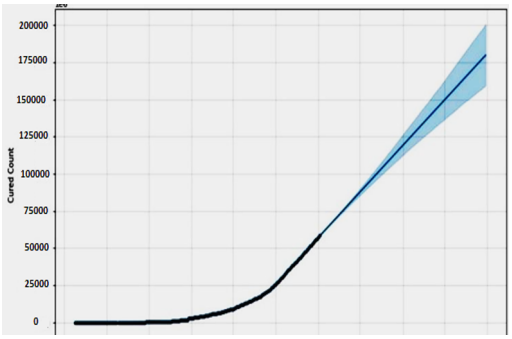


WEST BENGAL



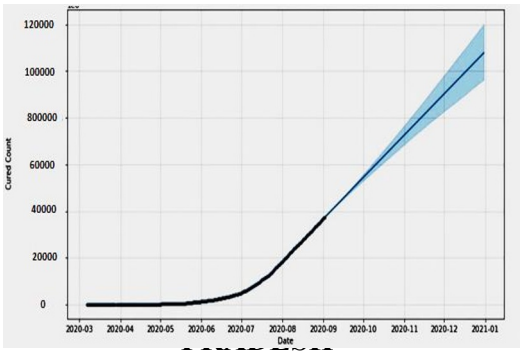
ılues of confirmed cases up to January 2021

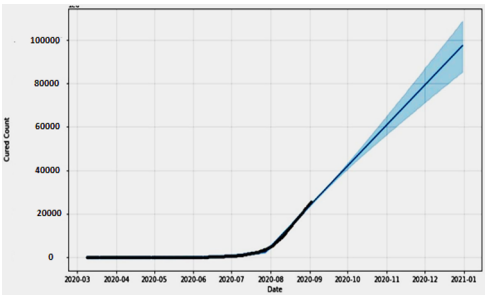
MAHARASHTRA

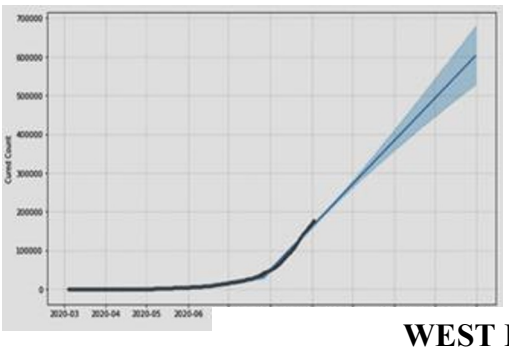


ANDHRA PRADESH

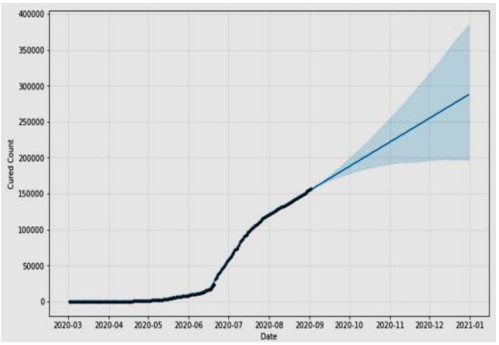
TAMIL NADU



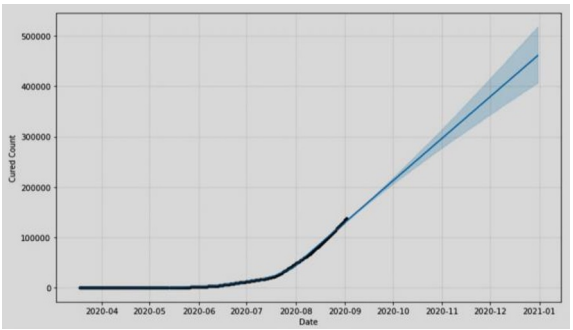




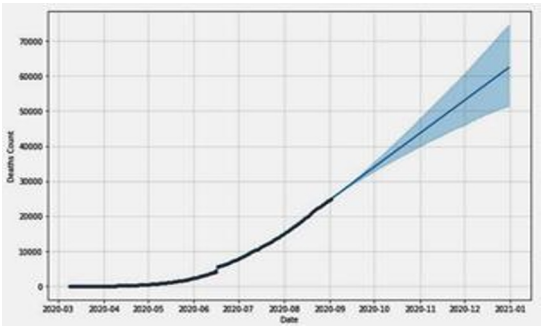
DELHI



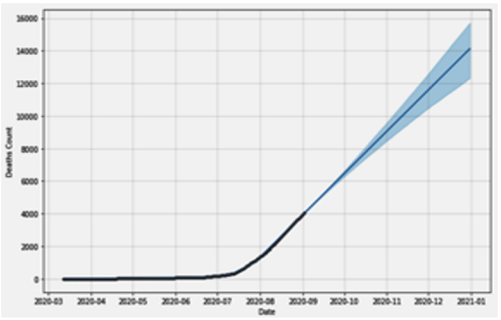
WEST BENGAL



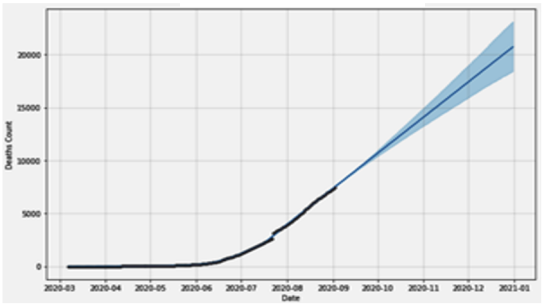
MAHARASHTRA



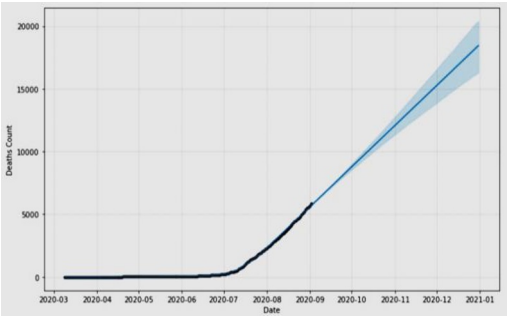
ANDHRA PRADESH



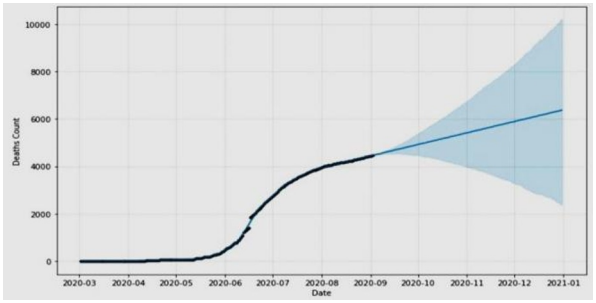
TAMIL



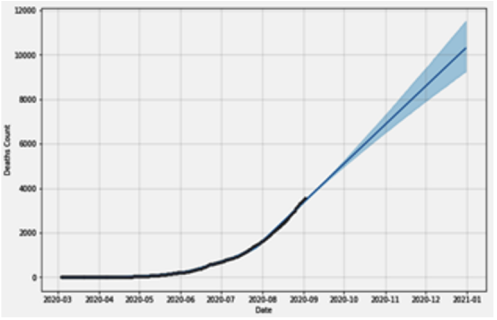
KARNATAKA



UTTAR PRADESH



DELHI



WEST BENGAL

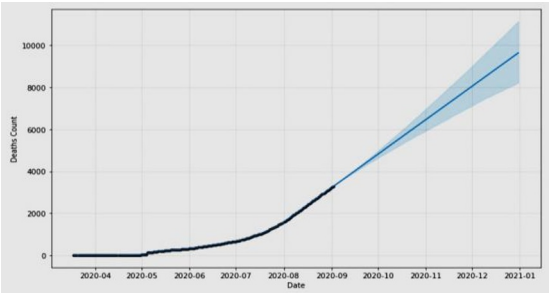


Fig 6: The top seven state-Wise predicted values of death cases up to January, 2021

CONCLUSION

It is imperative to adopt mitigation strategy for controlling the pandemic spread of corona virus disease using suitable models. This work is focused to study and comprehend the view point of COVID-19 epidemic with the help of predictive analytics. Here, Prophet Algorithm was employed for predictive analysis by using the Kaggle dataset of number of cases for confirmed, recovery and death patients. The present study analyzes the effect of COVID-19 on the top seven infected states in India. This analysis can help the government agencies for better decision-making.

REFERENCES

1. C. Drosten et al., "Identification of a novel corona virus associated with severe acute respiratory syndrome". N EnglJ Med. 348, (2003)1967-1976.
2. https://www.ecdc.europa.eu/
3. Y, Chen,Q. Liu,D. Guo,"Emerging corona virus es: genome structure, replication, and pathogenesis". J Med Virol.92, (2020) 418-423.
4. N. Zhu et al.,"A novel corona virus from patientswith pneumonia in China, 2019". N Engl J Med.382, (2020) 727-733.
5. J. F. W. Chanet al.,"A familial cluster of pneumonia associated with the 2019 novel corona virus indicating person-to-person transmission: a study of a family cluster". Lancet. 395, (2020)514–523.
6. https://covid19.who.int/ [online dataaccessed on 5 September 2020 (2:45 PM CEST)].
7. Madhu, Bhukya, and M. Venu Gopalachari. "Classification of the Severity of Attacks on Internet of Things Networks." In Sentiment Analysis and Deep Learning: Proceedings of ICSADL 2022, pp. 411-424. Singapore: Springer Nature Singapore, 2023.
8. https://www.mohfw.gov.in/[online data accessed on 5 September 2020 (8:00AM IST)].
9. S. K,Dey,M. M. Rahman, U. R. Siddiqi, A. Howlader, "Analyzing the epidemiological outbreak of COVID-19: Avisual exploratory data analysis approach". J Med Virol.92, (2020) 632-638.
10. V. Volpert,M. Banerjee,S. Petrovskii,"On a quarantine model of corona virus infection and data analysis".*Math Model* Nat Pheno.1*,* (2020) 24.
11. Madhu, Bhukya, M. Venu Gopala Chari, Ramdas Vankdothu, Arun Kumar Silivery, and Veerender Aerranagula. "Intrusion detection models for IOT networks via deep learning approaches." Measurement: Sensors 25 (2023): 100641.
12. Z. Liu,P. Magal, O. Seydi, G. Webb,"Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data".Math Biosci Eng. 17, (2020) 3040-3051.

13. M. Chen,Y. Hao,K. Hwang,L. Wang, "Disease prediction by machine learning over big data from healthcare communities". IEEE Access. 5, (2017) 8869-8879.
14. P. Kumar et al., "Forecasting the dynamics of COVID-19 Pandemic in Top 15 countries in April 2020: ARIMA Model with Machine Learning Approach, medRxiv. (2020).
15. Sivakumar, S. A., Tegil J. John, G. Thamarai Selvi, Bhukya Madhu, C. Udhaya Shankar, and K. P. Arjun. "IoT based Intelligent Attendance Monitoring with Face Recognition Scheme." In 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), pp. 349-353. IEEE, 2021.
16. L. Jia, K. Li, Y. Jiang, X. Guo, T. Zhao,"Prediction and analysis of corona virus disease" arXiv:2003, (2019) 05447.
17. https://www.mohfw.gov.in/
18. https://www.covid19india.org/
19. https://www.kaggle.com/datasets?
20. https://mlcourse.ai/articles/topic9-part2-prophet/
21. https://machinelearningmastery.com/time-series-forecasting-with-prophet-in-python/
22. https://facebook.github.io/prophet/docs/quick_start.html
23. https://www.analyticsvidhya.com/blog/2018/05/generate-accurate-forecasts-facebook-prophet-python-r/