

Machine Learning algorithm for predicting number of COVID-19 cases

B. Suvarna

*Department of Computer Science and Engineering
VFSTR Deemed to be University, Guntur, Andhra Pradesh, India*

Dr T Maruthi Padmaja

*Department of Computer Science and Engineering
Vardhaman Engineering College, Hyderabad, Telangana, India*

Dr. Venkatesulu Dondeti

*Department of Computer Science and Engineering
VFSTR Deemed to be University, Guntur, Andhra Pradesh, India*

Haritha Telaprolu

*Department of Computer Science and Engineering
VFSTR Deemed to be University, Guntur, Andhra Pradesh, India*

Harshitha Pappula

*Department of Computer Science and Engineering
VFSTR Deemed to be University, Guntur, Andhra Pradesh, India*

Abstract- COVID-19 is the infectious disease caused by corona virus which is recently discovered. COVID-19 is now a pandemic and many countries are affected by this globally. To decrease the spread of virus and to make necessary arrangements in hospitals, to take the right decisions for the governments prediction the future cases are required. Hence the numbers of cases that are going to be happening in the upcoming weeks or months are fore casted using algorithms like PROPHET and growth rate . By using these algorithms, this work fore casted the number of cases for the next 365 days and can also be used for predicting next year's also.

Keywords – COVID-19, Prophet, Growth Factor

I. I. INTRODUCTION

Virus is an infectious agent. This virus resides in the living cells of the different life forms like humans, animals, plants and bacteria. It is a small infectious agent inside the living host cells, where the host cell shall produce thousands of original copies. This spread of viruses can be in many ways. Viral infections can cause disease like life-threatening infections in humans, animals and plants.

Corona viruses are family of virus's cause's saviour illness in animals and humans. COVID-19 is the infectious disease caused by a corona virus which is recently discovered. This virus mostly affects a person's respiratory system. It is a disease that was detected at the end of 2019 and announced a pandemic on March 11. In Wuhan, China, the first case was confirmed officially. Since then, there has been an exponential growth in the number of such cases around the globe.

As of June 12th 2020, the total reported cases reached 7,597,341 out of which 423,844 have died. How it spreads: This is mainly spreading through the droplets either through coughing or sneezing which is produced by the infected person. Many of the infections are coming through the direct contacts of a person which causes one way of spreading. It may spread the disease by touching some contaminated surface or fabric and then touching one's mouth, nose, or eyes. For people with the infection incubation period can be from one to fourteen days. And disease can also spread without any severe symptoms. So far there is no vaccine and medicine for preventing this virus but only we can limit the spread of virus by maintaining the distance between persons and maintaining hygiene. Some small work is already done in the analysis of COVID-19 [1-6].

Basically, it is a forecasting problem, several forecasting algorithms based on statistical theory [1] and ARIMA models are used to forecast future cases. This paper attempts to forecast the future cases using PROPHET and Growth rate models.

II. II. PROPOSED FRAMEWORK

A. 2.1 Prophet model–

Methodologies used for the prediction is Prophet and growth factor. Prophet is a model for predicting time series data mainly based on Gaussian Additive Model (GAM) where non-linear trends are fit with weekly, monthly and yearly. This model works very well with the time series data which contains historical data. Compared to a generalized additive model (GAM), Prophet suits many linear and nonlinear functions of time as components with time as a regressor. Simplest in form;

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

G(t)- non periodic changes

S(t)-seasonality

H(t)-holidays

E(t)- Covers idiosyncratic changes not suited to the model

This model frames the prediction problems as a curve fitting exercise rather than the time based observation [7].

This model provides two models for trend for g(t) and those are saturating growth model and piecewise linear model.

Saturating growth model is $g(t) = \frac{c}{1+\exp(-k(t-m))}$

Where c-carrying capacity

k-growth rate

m-offset parameter

B. 2.1 Growth factor –

Growth Factor: In this first find the growth factor and predict the values for future by multiplying the previous day cases with growth factor. Algorithm for finding growth difference is given below.

Algorithm

Input:

D={d1, d2, d3, ..., dn} //confirmed cases in each day

Procedure

for i in D do:

Growth_difference=D[i]/D[i-1]

end

growth_factor=average of Growth_difference

III. III. EXPERIMENT AND RESULT

For the experimentation the data set is taken from Kaggle. The data sets that are used in the experimentation are given in the table 1.

Table 1: Data Set description

S.No	Dataset name	Attributes	Size
1	covid_19_india.csv	date, time, state/union territory, confirmedIndianNational, confirmedForeignNational, cured, deaths, confirmed	24 14
2	Individual_details.csv	id,government_id,diagnosed_date,age,gender,detected_city,detected_district,detected_state,nationality,current_status,status_changed_date,notes	27 89 0
3	Confirmed_df.csv	province/state, country/region, latitude, longitude, dates from(22-01-2020 to 27-05-2020)	26 6
4	Deaths.csv	province/state, country/region, latitude ,longitude, dates from(22-01-2020 to 27-05-2020)	26 6
5	Recovered.csv	province/state, country/region, latitude, longitude, dates from(22-01-2020 to 28-05-2020)	25 3
6	Train.csv	Id, province/state, country/region, confirmed cases, fatalities	35 99 5
7	Test.csv	Forecast_id, province/state, country/region	13 45 9
8	Submission.csv	Forecast_id, confirmed cases, Fatalities	13 15 8

Flow of execution: For the prediction, take the dataset as input and then pre process the data. Once the dataset is ready then apply the model on dataset and make the predictions by using test dataset.

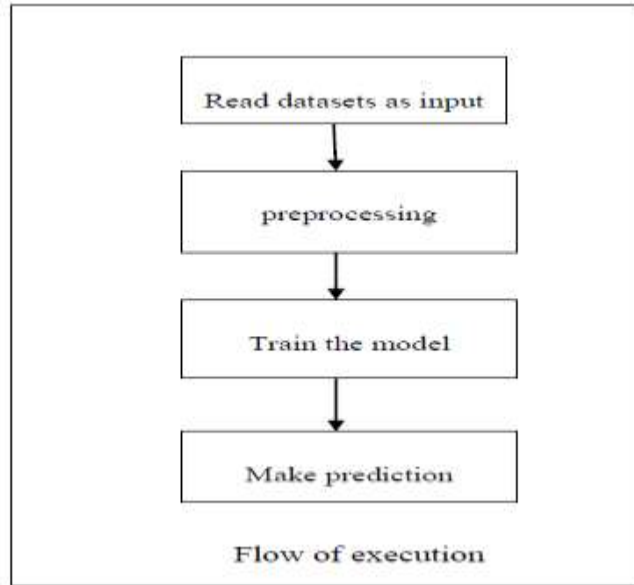


Figure 1: Flow of execution for prediction

We have extracted the data for India to find the confirmed cases and applied the model Prophet for predicting the future cases. By using this model we have predicted for the cases to the months of up to June, august, October, and December and up to March, 2021. Find the results up to 1st June 2020 in the figure2.

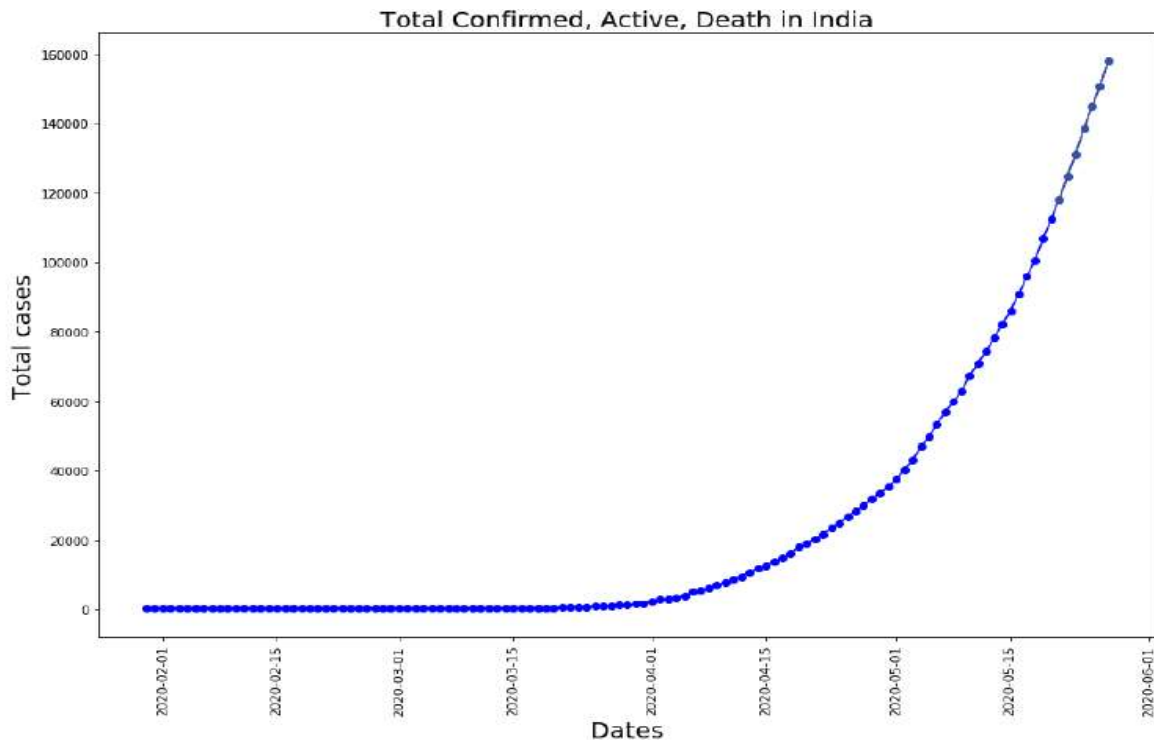


Figure 2. Number of cases confirmed for India up to 1st June, 2020

Results obtained by using Prophet Model: Results obtained with prophet model are described in figures from the figure numbers 4 to 8 and also in the table 2. The number of cases are going to be high From the obtained results.

Table 2. Results obtained by using Prophet Model

Month, Year	Predicted values
June, 2020	Above 2,50,000
August, 2020	Above 5,00,000
October, 2020	Above 8,00,000
December, 2020	Above 10,00,000
March, 2021	Above 15,00,000

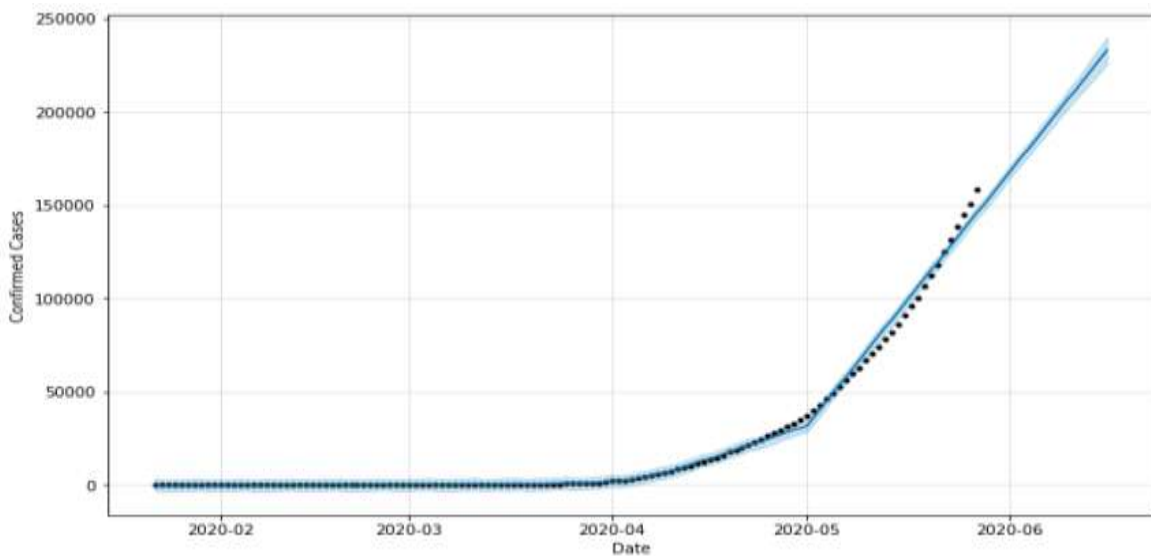


Figure 3: Prediction for the month end of June, 2020

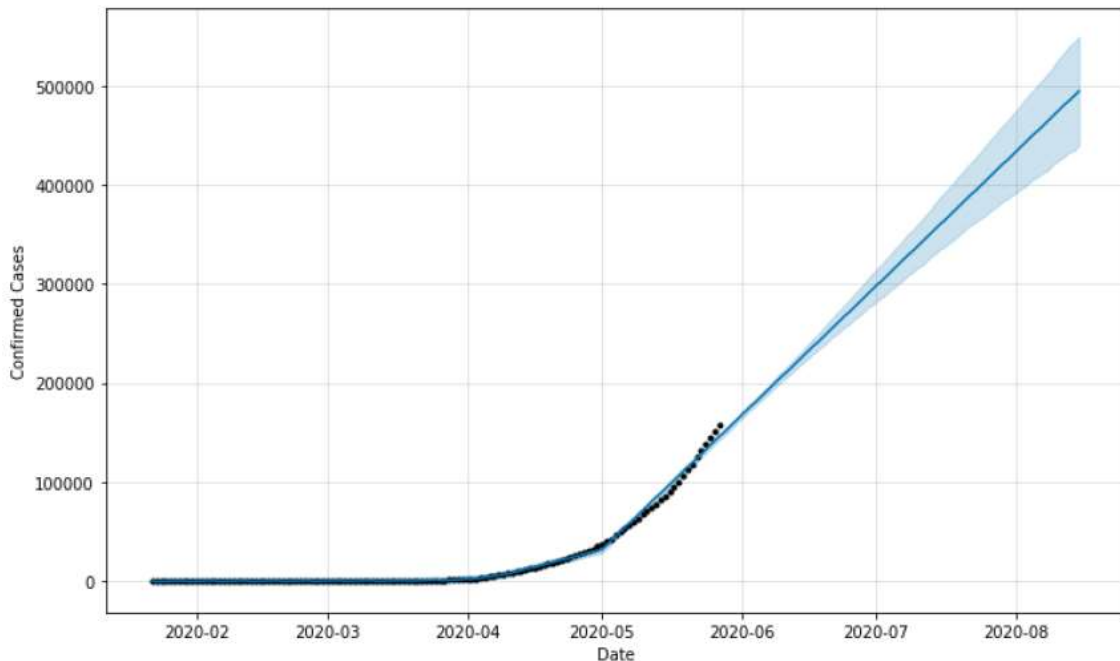


Figure 5. Predicting values for end of August, 2020

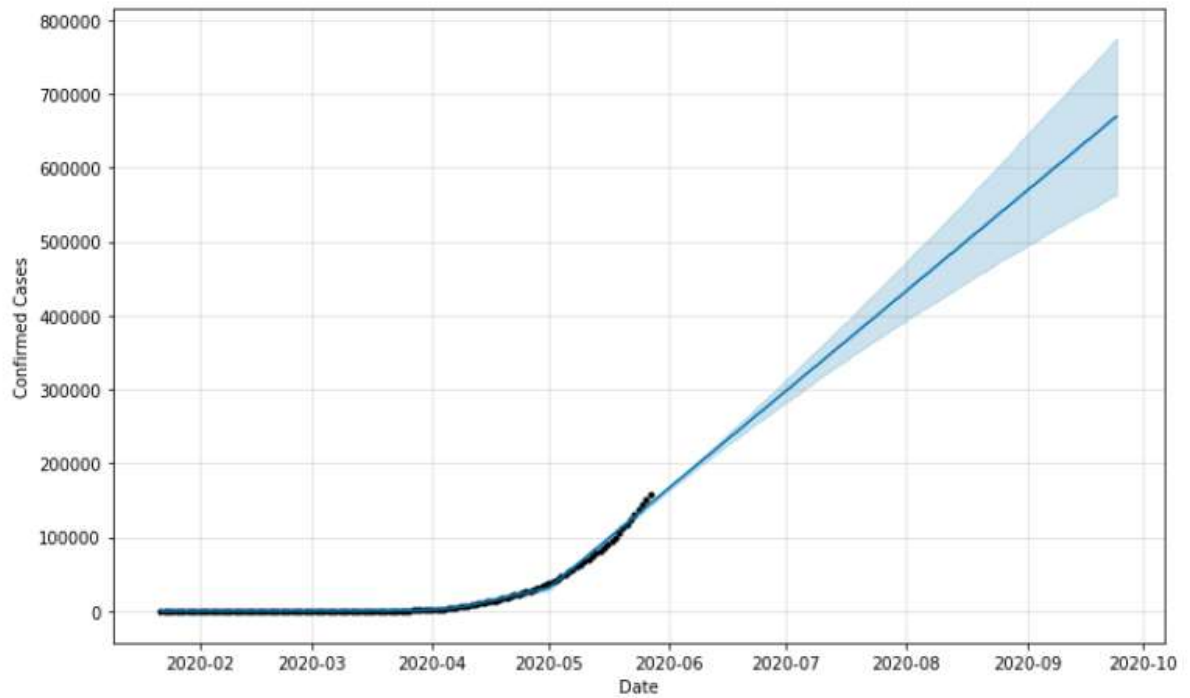


Figure 6. Predicting values for end of October, 2020

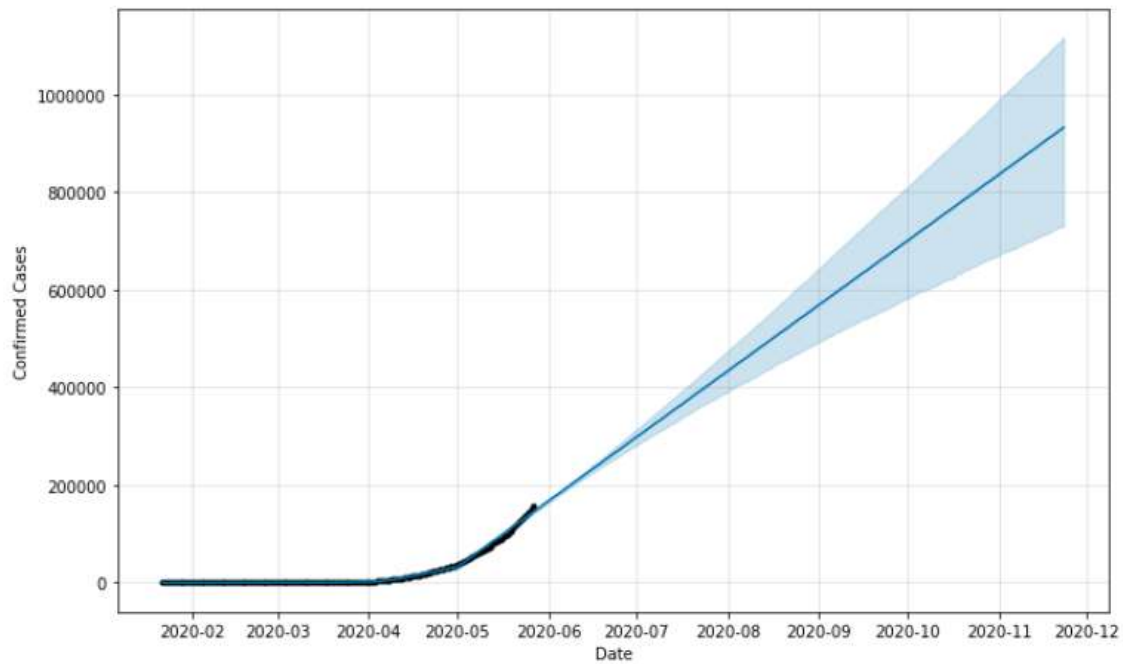


Figure 7. Predicting values for end of December, 2020

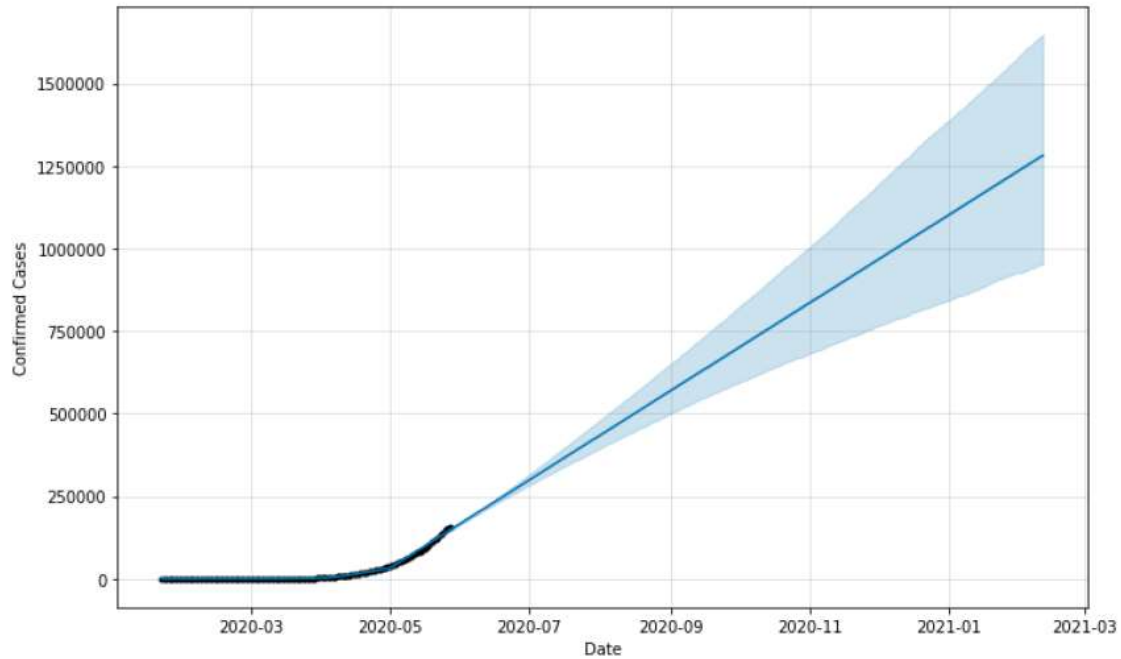


Figure 8: Predicting values for end of March, 2021

Results obtained by using growth factor: With this model we have predicted number of cases to the world wide. We have done the prediction starting from the end of May to the next 365 days i.e. up to May, 2021. And the predicted values are given in the table 3.

Table 3. Results obtained using growth factor

Days	Date	No of cases
15 days	11-06-2020	10,00,000
30 days	25-06-2020	60,00,000
60 days	22-07-2020	2.0*1e8
90 days	15-08-2020	8.0*1e9
180 days	12-30-2020	5.0*1e14
365 days	05-30-2020	3.0*1e24

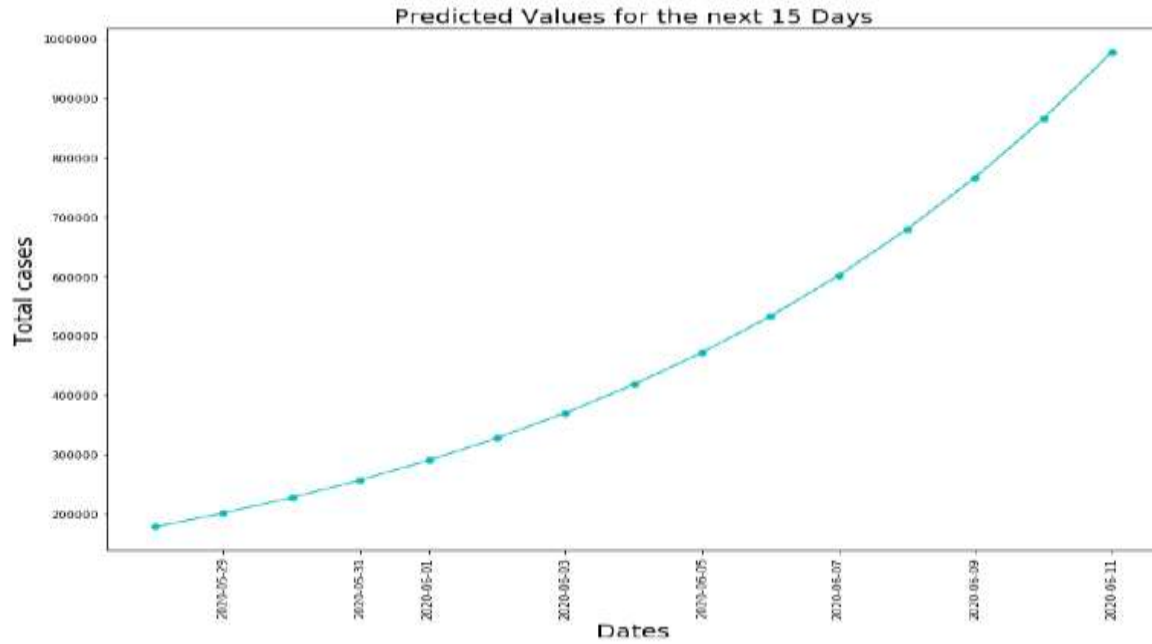


Figure 9. Predicted values for the next 15 days

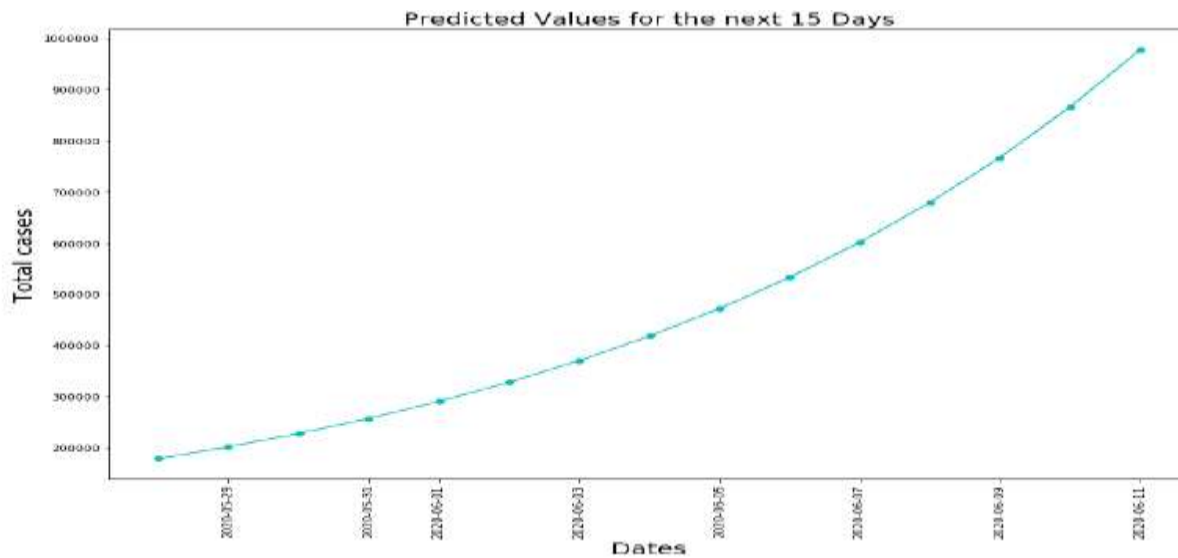


Figure 10: Predicted values for the next 30 days

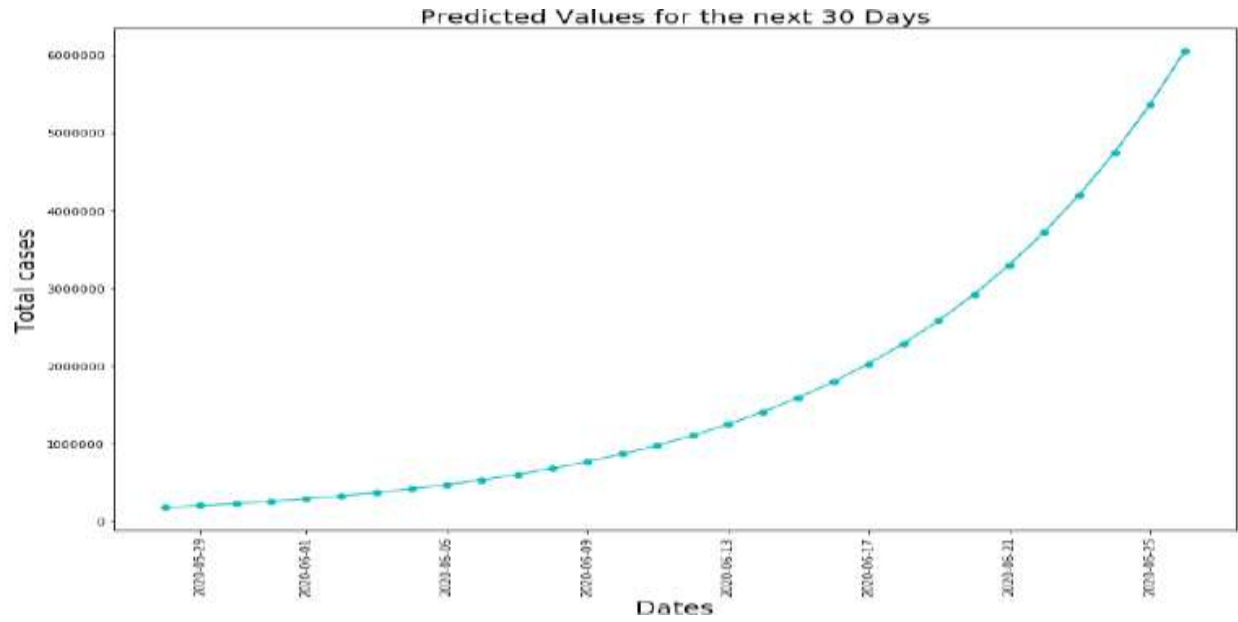


Figure 11. Predicted values for the next 30 days using growth factor

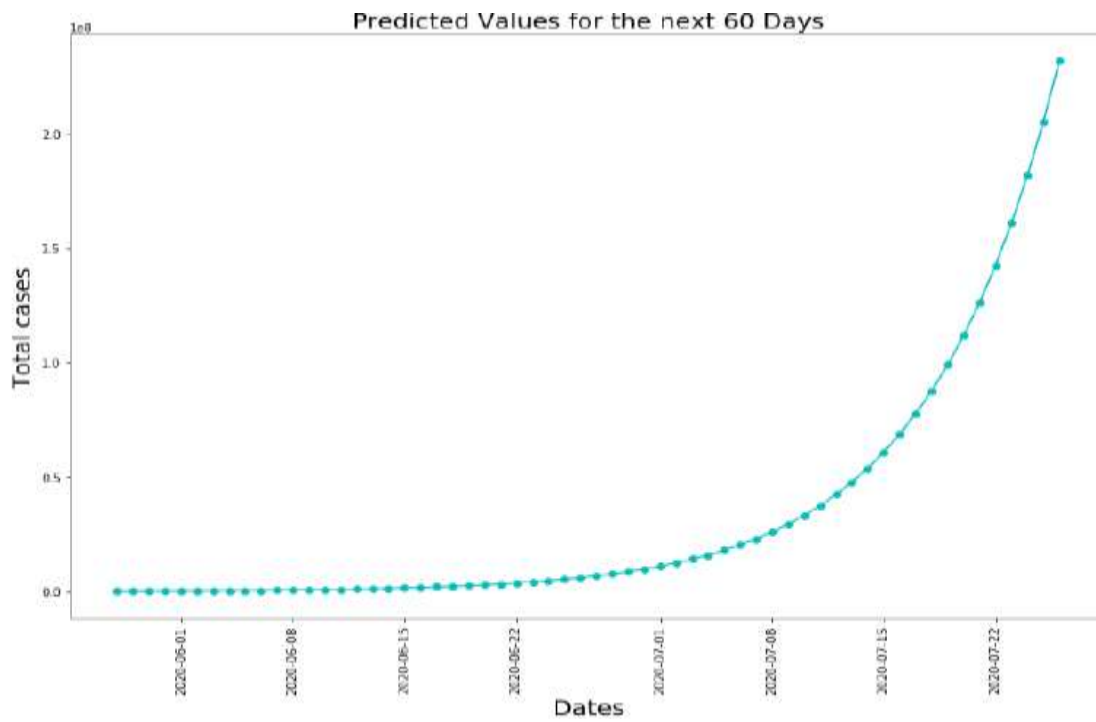


Figure 12. Predicted values for the next 60 days

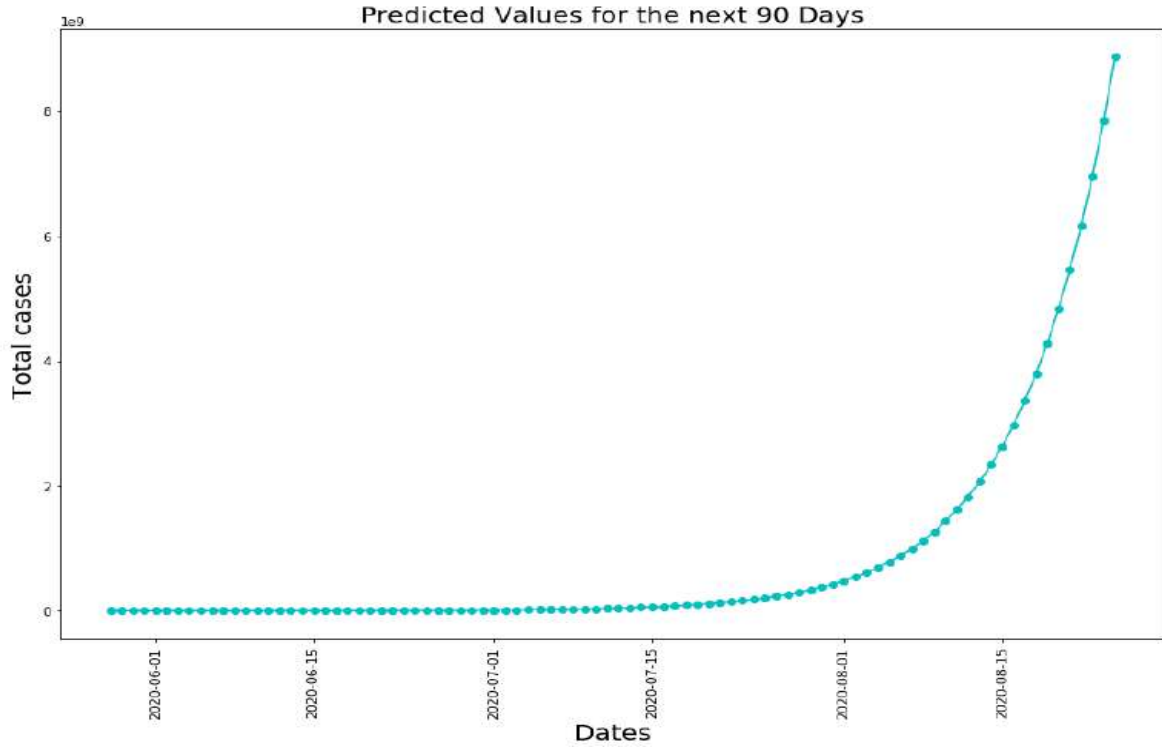


Figure 13: Predicted values for the next 90 days

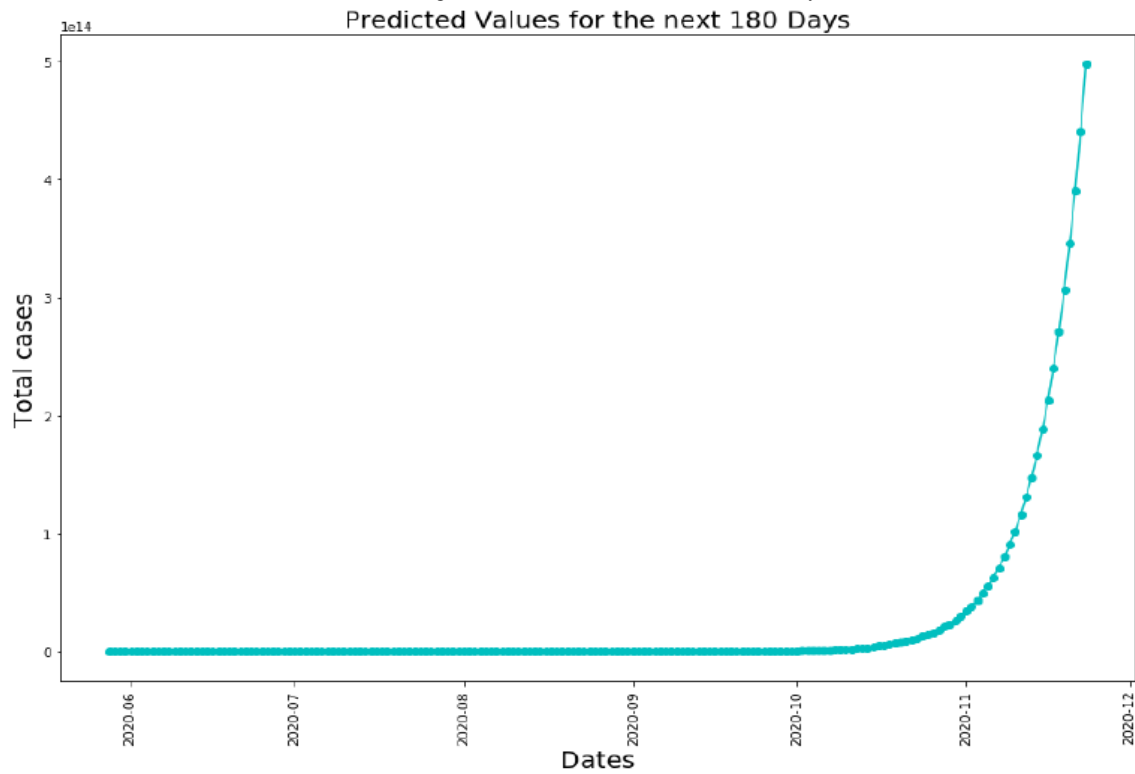


Figure 14: Predicted values for the next 180 days

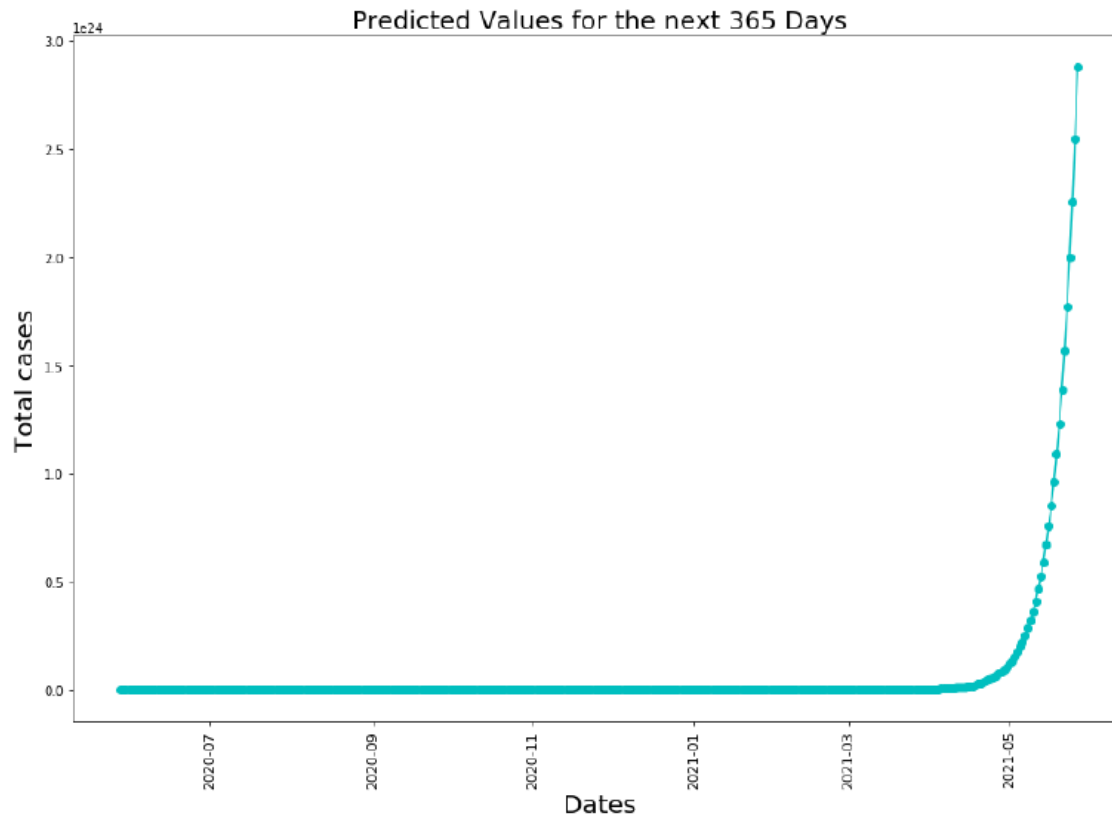


Figure 15: Predicted values for the next 365 days

IV. IV.CONCLUSION

To decrease the spread of virus and to make necessary arrangements in hospitals, to take the right decisions for the governments, predicting the future cases are required. . Hence for predicting the next numbers of cases that are going to be happening in the upcoming weeks or months, algorithms like PROPHET and growth rate are used. By using these algorithms we have predicted the no of cases for the next 365 days and can also be used for predicting next year's also. From the obtained results Prophet model is working very well compared with Growth rate.

V. REFERENCES

- [1] Lin Jia, Kewen Li, Yu Jiang, Xin Guo, Ting Zhao: Prediction and analysis of Coronavirus Disease 2019, arxiv. 2003.05447, 2020.
- [2] Qiang Li, Wei Feng: Trend and forecasting of the COVID-19 outbreak in China. Arxiv. 2002.05866, [10.1016/j.jinf.2020.02.014](https://arxiv.org/abs/2002.05866), 2020.
- [3] Bohdan M. Pavlyshenko: Machine Learning models for time series forecasting. MPDI, 2019, pp. 1-11. doi:10.3390/data401001, 2018.
- [4] Jagadish Kumar, K.P.S.S. Hembram: Epidemiological study of novel corona virus (COVID-19). arxiv.2003.11376, 2020.
- [5] Fattah J, Ezzine L, Aman Z, El Moussami H, Lachhab A: Forecasting of demand using ARIMA model. International Journal of Engineering Business Management; 10:1847979018808673, 2018.
- [6] Pavlyshenko, B.M. Linear: Mmachine learning and probabilistic approaches for time series analysis. In Proceedings of the IEEE First International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine, 23–27 August 2016; IEEE: Piscataway, NJ, USA, pp. 377–381. 2016
- [7] Robson, Winston A. "The Math of Prophet— Breaking down the Equation behind Facebook's open-source Time Series Forecasting procedure", *Medium*, Future Vision, June 17 2019, <https://medium.com/future-vision/the-math-of-prophet-46864fa9c55a>