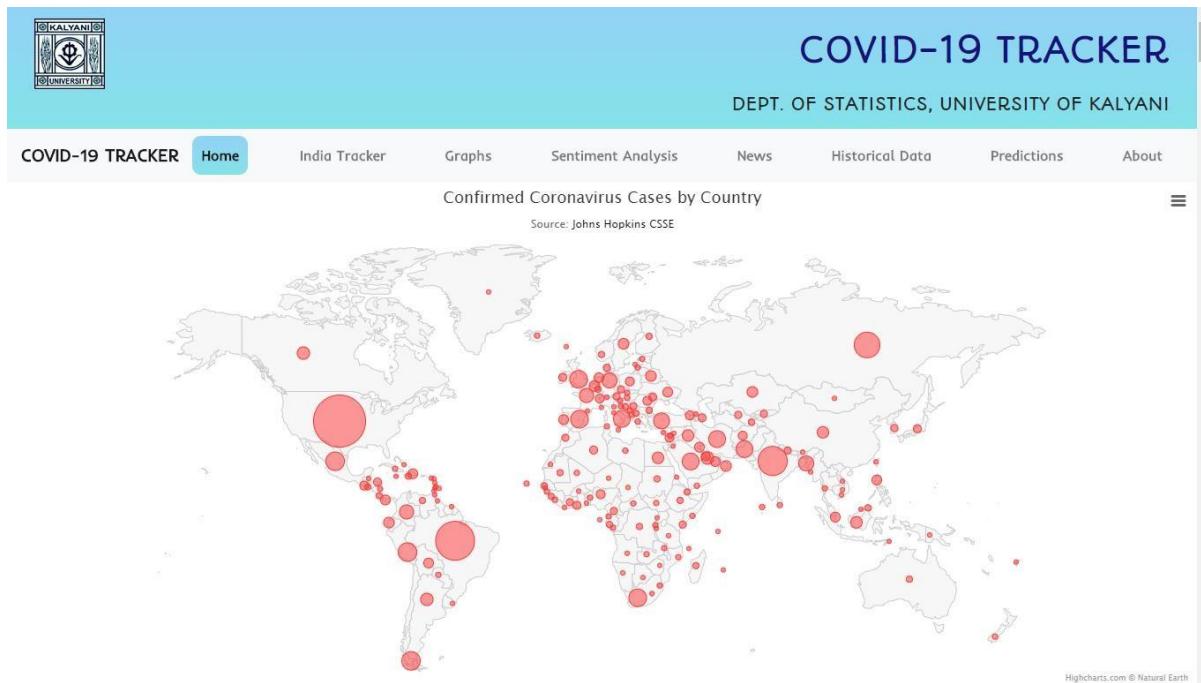


COVID-19 Tracker Documentation

Access our project at the following link:

<https://sites.google.com/view/kucovid19stat/> or, <http://3.6.38.122:5000/>



Group Members

Name: ABHIJIT SHOW (2 years)

Roll: 96/STA NO. 180001

Registration No.: 100543 of 2018-2019

Name: SAYAN GUHA ROY (2 years)

Roll: 96/STA NO. 180036

Registration No.: 100542 of 2018-2019

Name: SOURAV GHOSH (2 years)

Roll: 96/STA NO. 180041

Registration No.: 100587 of 2018-2019

**Department of Statistics
University of Kalyani**

Contents

Topics	Page
1. Abstract	2
2. Pages for Global and India Tracker	3
3. Home Page	4
4. India Tracker	5
6. Graphs (World/India)	5
7. Sentiment Analysis (World/India)	8
8. News (World/India)	10
9. Historical Data	11
10. Predictions (World/India)	12
11. Description of the Facebook Prophet Model	14
• Introduction	14
• What's new in Prophet	14
• The Prophet Forecasting Model	15
○ Trend	15
○ Seasonality	17
○ Holidays and Events	17
12. Outputs of the Facebook Prophet Model	18
• 7-day Predictions for World	18
• 7-day Prediction for India	19
• Forecast Plot and Components for World	20
• Forecast Plot and Components for India	23
13. Python Packages Used in our Project	26
14. Acknowledgement	28
15. Declaration	29

Abstract

Covid-19 has put the world to a standstill. Doctors, healthcare workers and personnel of many other essential services are fighting at the frontline to tackle this global pandemic. Although we are not fighting the battle at the frontline, as students of statistics this is our humble attempt at partaking in the struggle.

We have created a website to track COVID-19 where we have displayed the data from the world as a whole and also country wise. The data is categorised into three components: **confirmed cases, deaths and recovered**. The values are given for both daily and cumulative type. We have tried our best to keep the display simple yet visually appealing. We have used line charts and pie charts and also an exquisite race chart for display. All the above-mentioned charts are interactive and are customized to give the user a clear idea of the intended meaning of the values as all the categories are separated by different colours, this not only made the graphs more appealing to the eyes but also helped in distinguishing different aspects. There are two sections dedicated to graphs, one for India and the other for the entire world, graphs for India are under **Graphs** under **India Tracker** and those for the world are under **Graphs** under **Home**. The entire site is designed to make it as user friendly as possible. Our source for global data is Johns Hopkins University, to access the data [click here](#) and to access the data for India [click here](#).

We didn't just stop at displaying the data but went ahead and made predictions for values of **total confirmed cases, deaths and recovered** for both India and the world. We obtained our predictions by using a Facebook Prophet model. Facebook Prophet is an excellent time series model developed by Facebook. It is fast and simple it also automatically takes into account the crucial time points in the data provided. Using this model, we have obtained appreciable accuracy in India predictions and excellent accuracy in world predictions. The predictions are displayed under **Predictions**.

Our joys would know no bounds if our humble effort comes of any use to anyone battling this global pandemic. It is our utmost desire that we overcome this situation with unity and fraternity.

Pages for Global Tracker

- + [Home page](#)
- + [Graphs](#)
- + [Sentiment Analysis](#)
- + [News](#)
- + [Historical Data](#)
- + [Prediction](#)
- + [About](#)

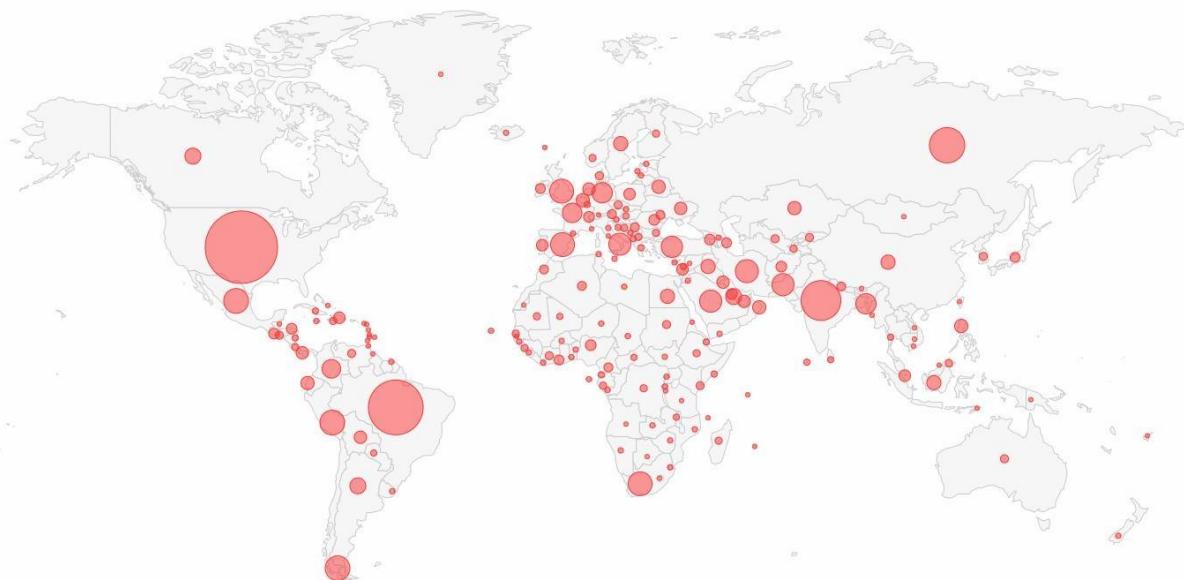
Pages for India Tracker

- + [India Tracker](#)
- + [India Graphs](#)
- + [India Sentiment Analysis](#)
- + [India News](#)
- + [India Prediction](#)
- + [About](#)

Home Page

Confirmed Coronavirus Cases by Country

Source: Johns Hopkins CSSE



- Bubble chart showing the Covid-19 hotspots around the world
- Live total confirmed Covid-19 cases, deaths, recovered and affected countries.

Country / Other	Total Cases	New Cases	Active Cases	Cases/1M pop	Total Deaths	New Deaths	Deaths/1M pop	Total Recovered	Recovered/1M pop	Serious/Critical	Tests	Tests /1M pop
World 🌎	13,520,403	71,852	5,035,384	1,735	582,784	2,435	74.8	7,902,235	1,018.19	59,508	279,844,436	36,057.32
USA 🇺🇸	3,549,519	4,442	1,809,175	10,721	139,314	171	421	1,601,030	4,835.83	16,337	44,124,272	133,275
Brazil 🇧🇷	1,933,655	2,451	645,807	9,095	74,336	74	350	1,213,512	5,707.53	8,318	4,572,796	21,507
India 🇮🇳	956,992	19,505	327,665	693	24,703	388	18	604,624	437.98	8,944	12,412,664	8,991
Russia 🇷🇺	746,369	6,422	211,350	5,114	11,770	156	81	523,249	3,585.44	2,300	23,754,645	162,773
Peru 🇵🇪	333,867		98,377	10,121	12,229		371	223,261	6,768.06	1,325	1,963,921	59,535
Chile 🇲🇽	319,493		23,204	16,708	7,069		370	289,220	15,124.82	1,915	1,322,503	69,161
Mexico 🇲🇽	311,486	7,051	81,183	2,415	36,327	836	282	193,976	1,503.91	378	756,137	5,862
Spain 🇪🇸	303,699		275,290	6,495	28,409		608			617	6,026,446	128,893
South Africa 🇿🇦	298,292		147,667	5,027	4,346		73	146,279	2,465.32	539	2,232,738	37,630
UK 🇬🇧	291,373		246,405	4,291	44,968		662			162	12,270,317	180,712

- A table showing live total confirmed cases, new cases, active cases, cases per one million population, total deaths, new deaths, deaths per one million, total recovered, recovered per one million population, serious/critical cases, total tests, tests per one million population for each affected country in descending order.

India Tracker

🌟 Coronavirus Cases: 958,044

🔴 Deaths: 24,713

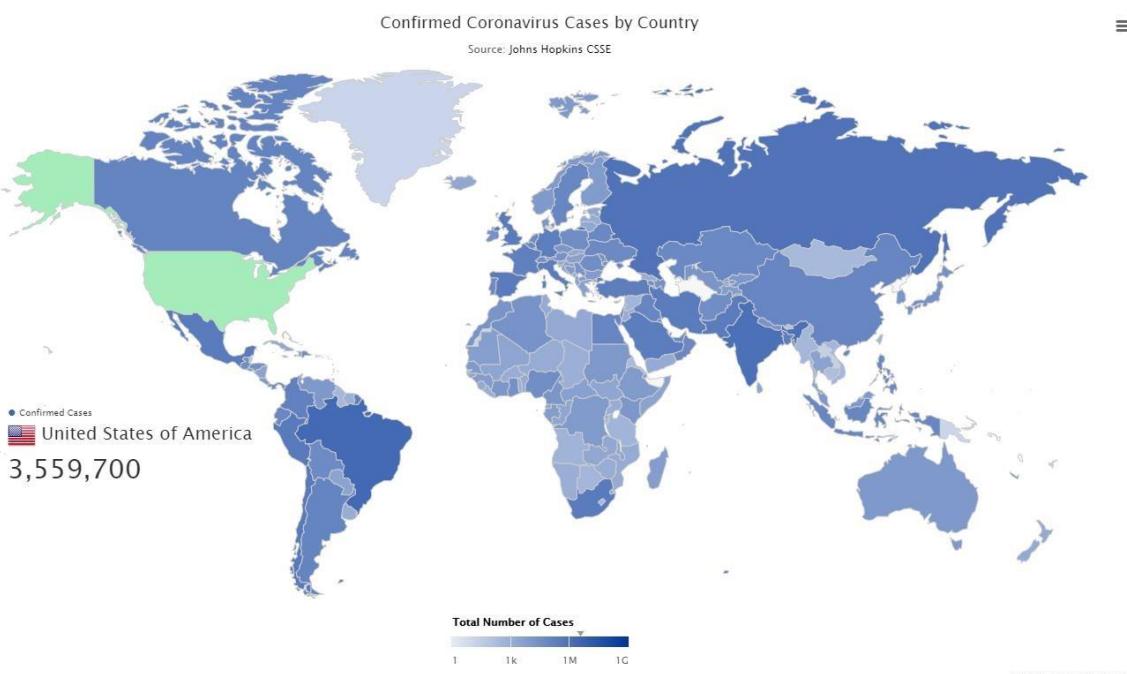
🟢 Recovered: 605,556

🌐 Affected States: 35

State / Other	Total Cases	New Cases	Total Deaths	New Deaths	Total Recovered	New Recovered	Active Cases
Total	958,044	20,482	24,713	398	605,556	12,467	327,386
Maharashtra	275,640	7,975	10,928	233	152,613	3,606	111,801
Tamil Nadu	151,820	4,496	2,167	68	102,310	5,000	47,343
Delhi	115,346		3,446		93,236		18,664
Gujarat	44,648	925	2,080	10	31,346	791	11,222

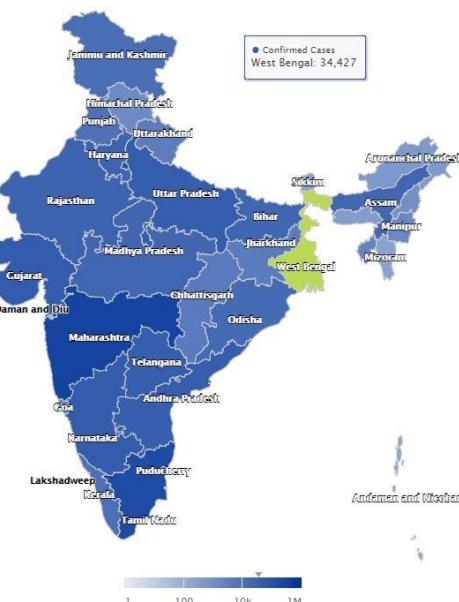
- Live total confirmed Covid-19 cases, deaths, recovered in affected states.
- A table showing live total confirmed cases, new cases, total confirmed deaths, new deaths, total confirmed recovered, new recovered, active cases.

Graphs (World/India)



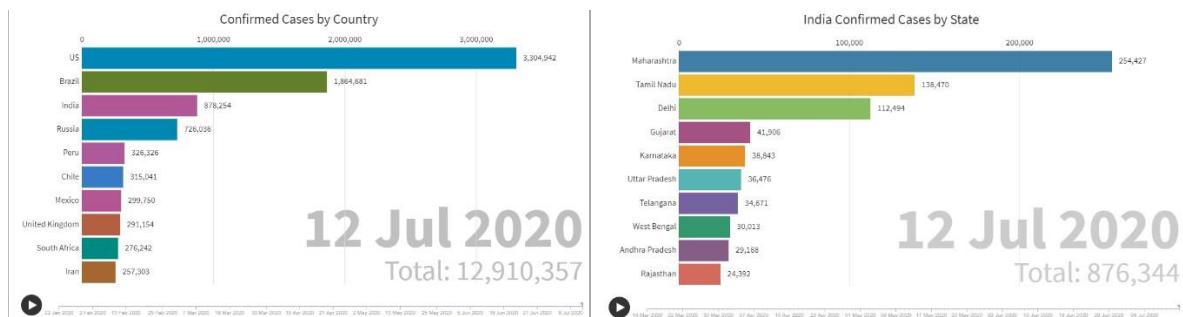
Confirmed Coronavirus Cases of India by State

Source: Covid19India

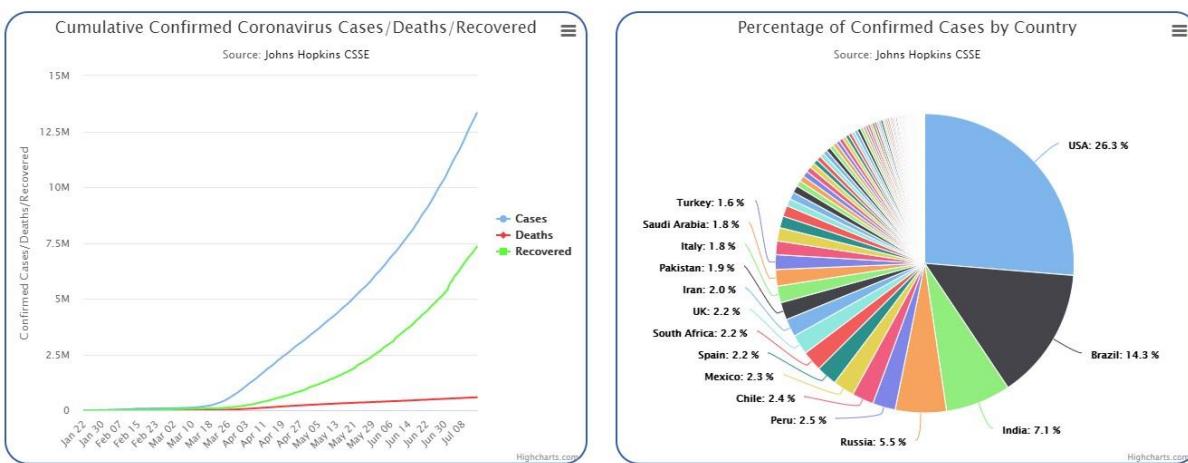


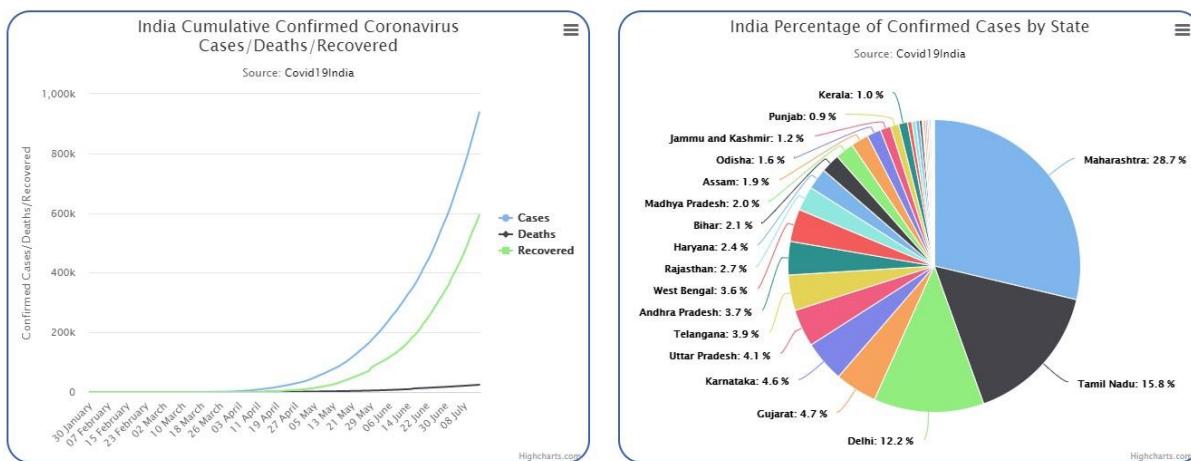
Highcharts.com © projects.datameet.org

- Hover able responsive world map showing total confirmed cases for each affected country/state.

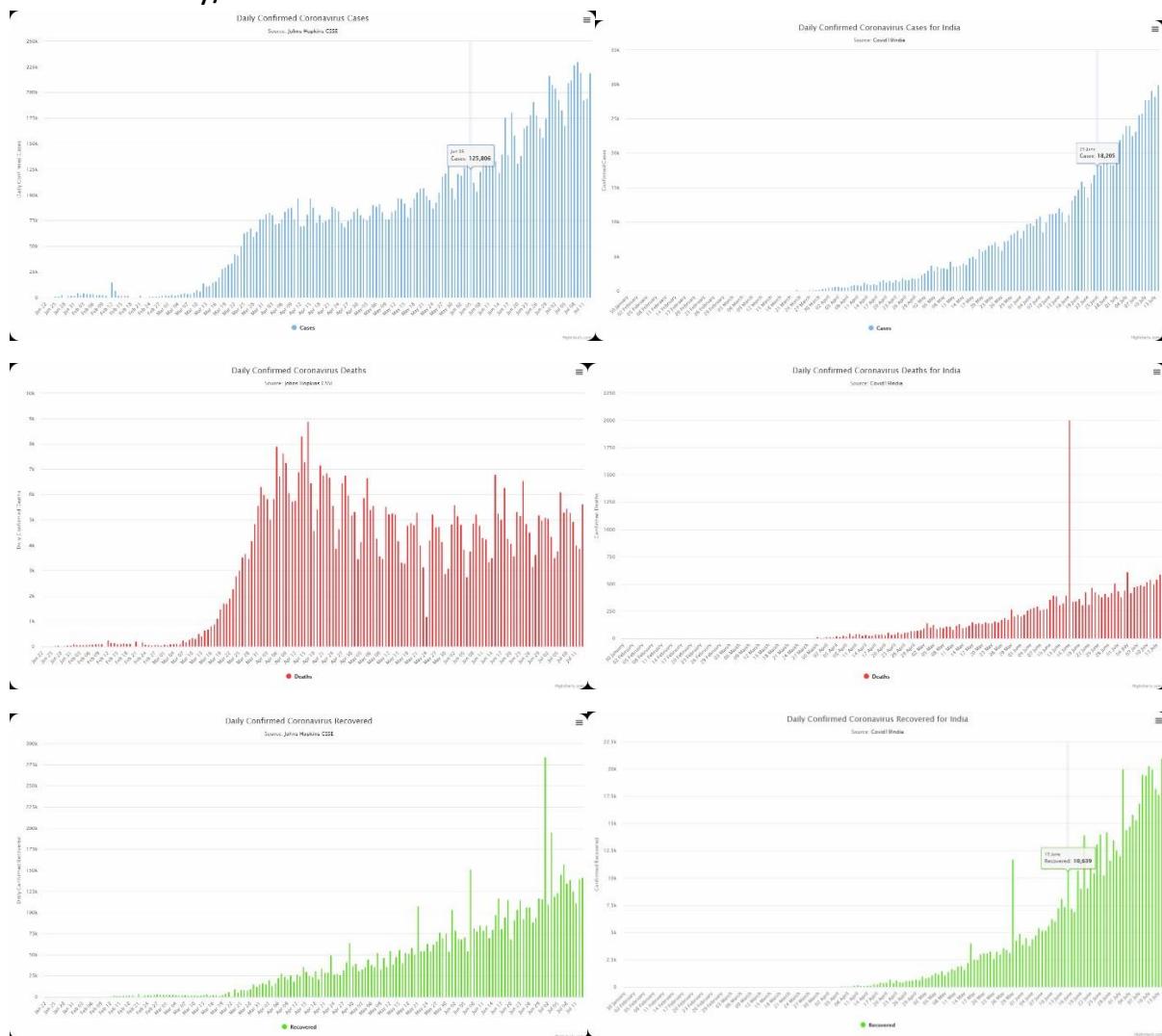


- A bar chart race showing top ten countries/states with highest confirmed cases at a given point of time.





- A line chart showing global/India cumulative confirmed Covid-19 cases, deaths and recoveries.
- A multi-coloured pie chart showing percentage of confirmed cases by country/state.



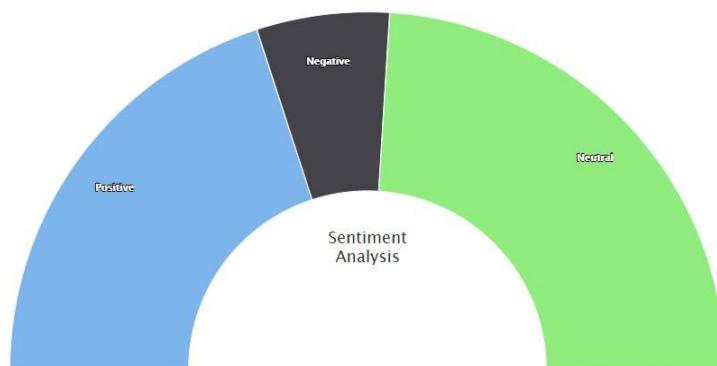
- Bar charts showing daily confirmed cases, deaths and recoveries in a time series format.

Sentiment Analysis (World/India)

Sentiment Analysis is the process of determining whether a piece of writing is positive, negative, or neutral. A sentiment analysis system for text analysis combines natural language processing (NLP) and machine learning techniques to assign weighted sentiment scores to the entities, topics, themes and categories within a sentence or phrase.

- Here we used the twitter API to fetch 100 recent tweets on the hashtags #coronavirus and #covid19.
- We analysed the fetched tweets and categorised them into three categories namely positive, negative, and neutral.
- We used a semi-donut chart to display the result from the sentiment analysis.

≡

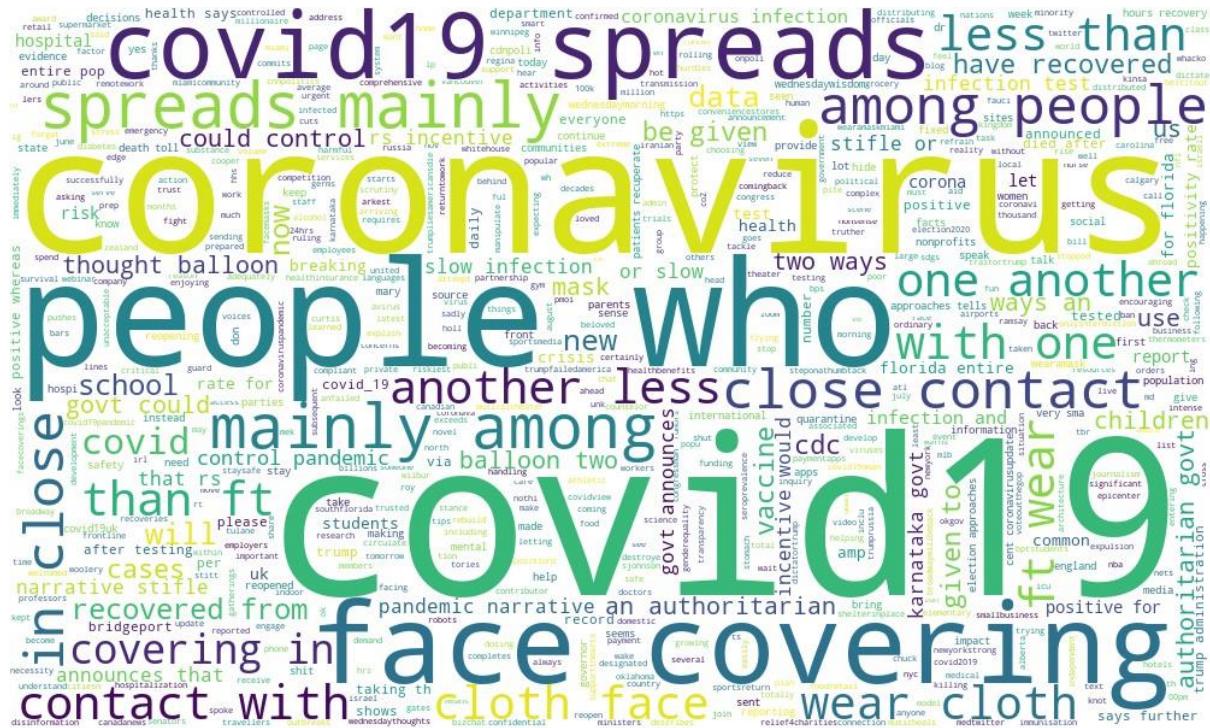


Highcharts.com

Word Cloud Visualization

A word cloud (wordle or weighted list in visual design) is a novelty visual representation of text data, typically used to depict keyword metadata (tags) on websites, or to visualize free form text. Tags are usually single words, and

the importance of each tag is shown with font size or colour. This format is useful for quickly perceiving the most prominent terms to determine its relative prominence. When used as website navigation aids, the terms are hyperlinked to items associated with the tag.



- Here we used the fetched tweets to generate a word cloud visualization.

#	Name	Tweet	Sentiment
1	Faith	RT : Thought balloon : Two ways an authoritarian govt could control a #pandemic narrative: 1.) Stifle or slow infection test...	Negative
2	Ashish kumar	RT : Karnataka govt announces that Rs 5,000 incentive would be given to people who have recovered from coronavirus infection and a...	Neutral
3	Maurice Wilson	RT : #COVID19 spreads mainly among people who are in close contact with one another (less than 6 ft). Wear a cloth face covering in...	Neutral
4	General Bethlehem	RT : BREAKING: The Department of Health says a further 85 people in the UK have died after testing positive for #COVID19, taking th...	Positive
5	COVID-19 HQ	RT : Extreme. Give me party or give me expulsion? Tulane will ban all parties or large gatherings of more than 15 people, incl...	Positive
6	StrudelHundt	RT : Thought balloon : Two ways an authoritarian govt could control a #pandemic narrative: 1.) Stifle or slow infection test...	Negative
7	pradeip nanda	RT : Will someone please wake them up and explain them that Covid19 race is not an athletic event where being No 3 is an award...	Neutral

- A table showing the name of the person, the tweet, and the sentiment of the tweet.

News (World/India)

Trump administration orders hospitals to send coronavirus data to Washington, not the CDC



The Trump administration has ordered hospitals to bypass the CDC and send all COVID-19 patient information to Washington starting Wednesday. - Google News

Tap Here



India Sees Highest Recorded Spike of 29,429 Cases Taking India's Tally Over 9.3 Lakh, Death Toll at 24,30... - News18



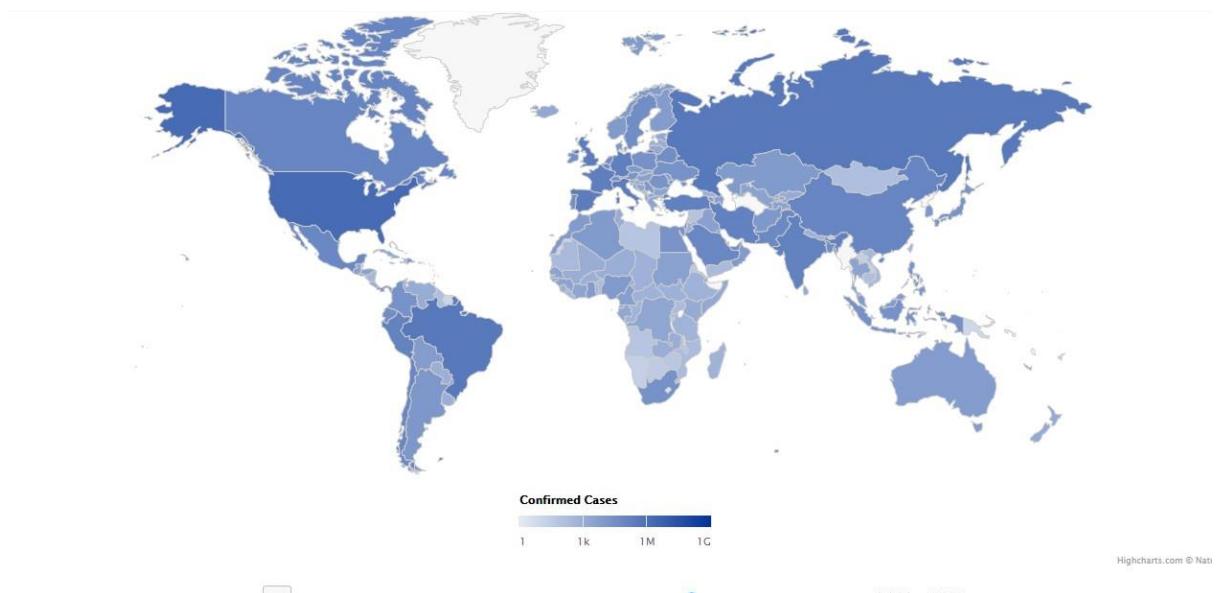
The number of recoveries stands at 5,92,031, while there are 3,19,840 active cases of coronavirus infection presently in the country, the updated data at 8 am showed. - News18

Tap Here



- We have showcased the trending news regarding Covid-19.
- Each and every news can be listened to in an audio format by tapping the audio button.
- Every news showcased contains the link to the complete news article.

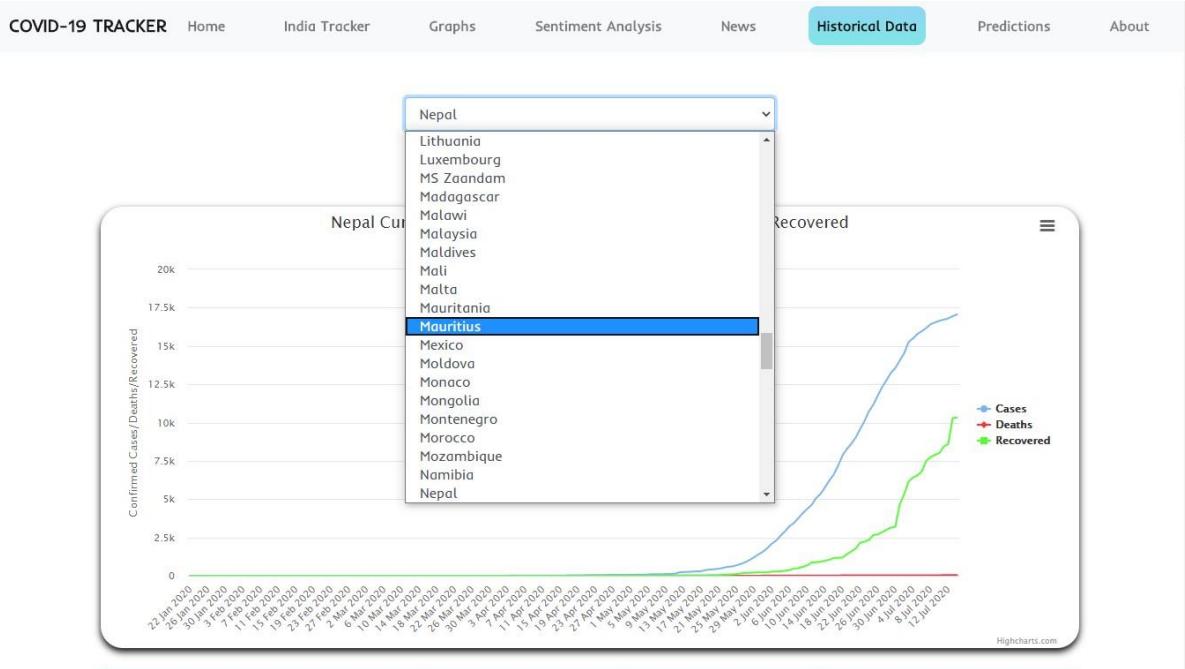
Historical Data



Click the Play/Pause Button

Hover Over any Country to check the number of Cases at that Date

- There is an interactive animation portraying the spread of Covid-19 confirmed cases worldwide.



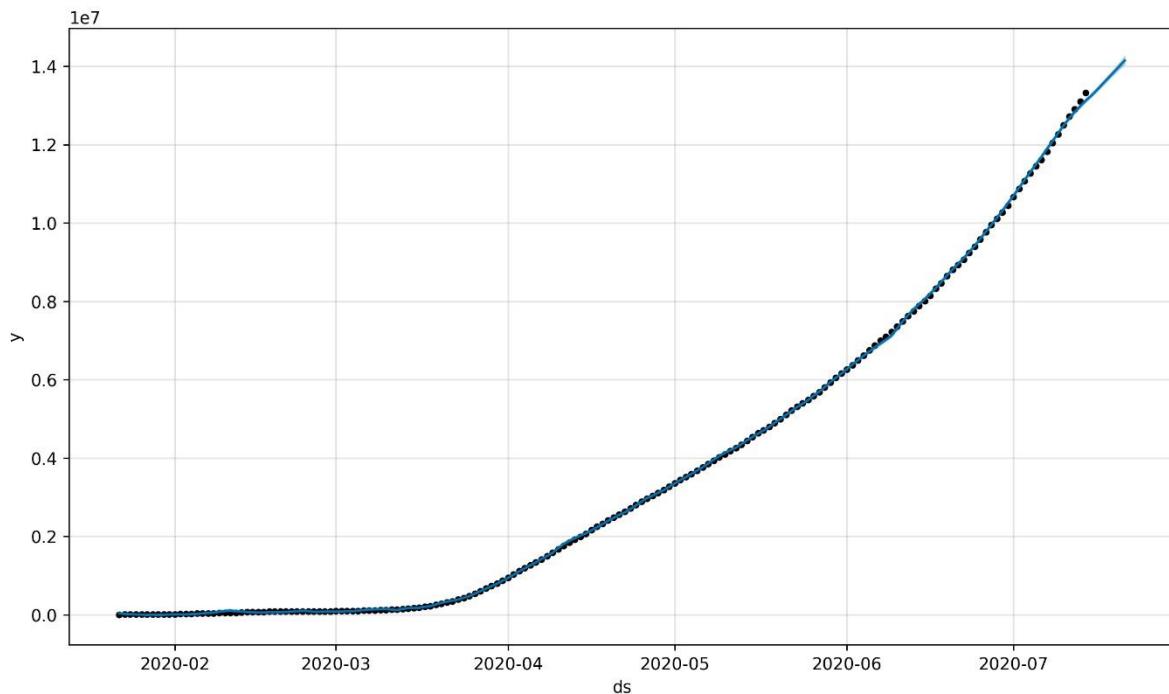
- There is a dropdown menu consisting of countries and the world as a whole, upon selection of a specific country timeseries of cumulative confirmed cases, deaths and recovered for that specific country is displayed in a table format. Same thing is displayed upon selecting world from the dropdown menu.

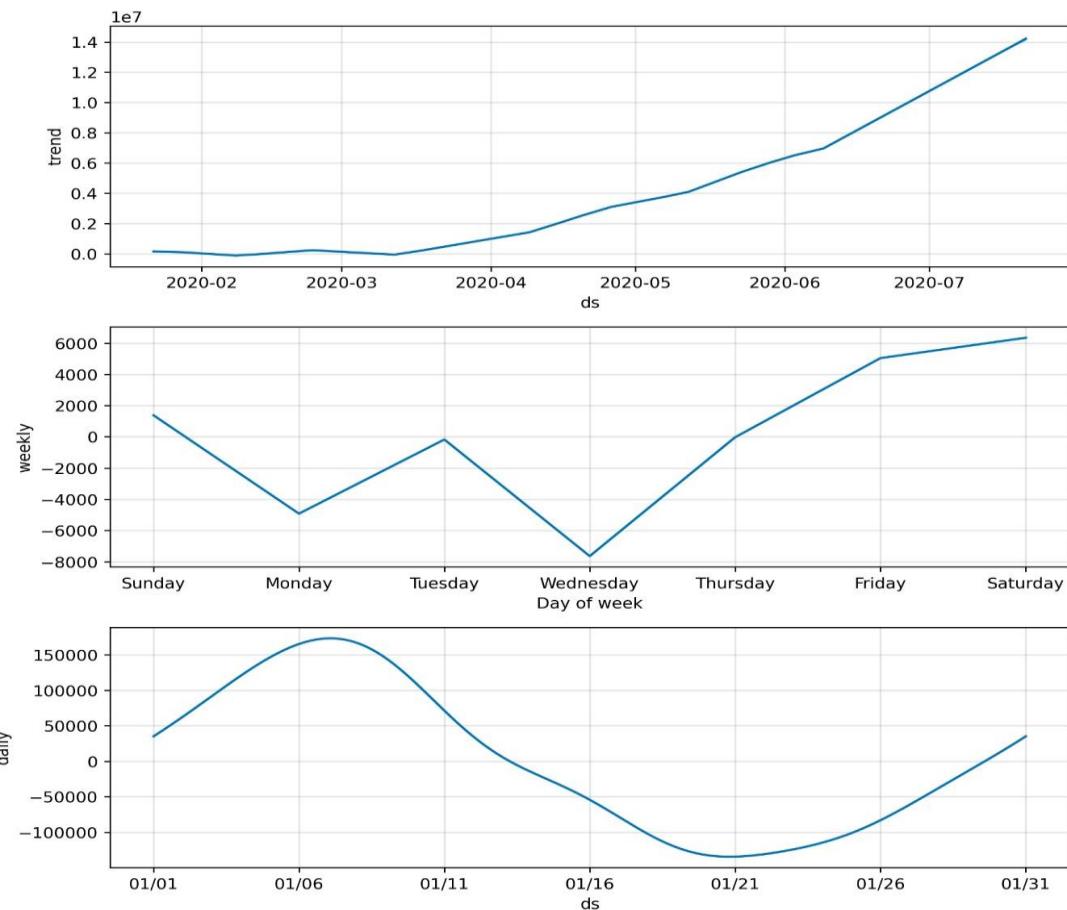
Predictions (World/India)

7 Day Forecast for Total Confirmed Cases, Deaths & Recovered

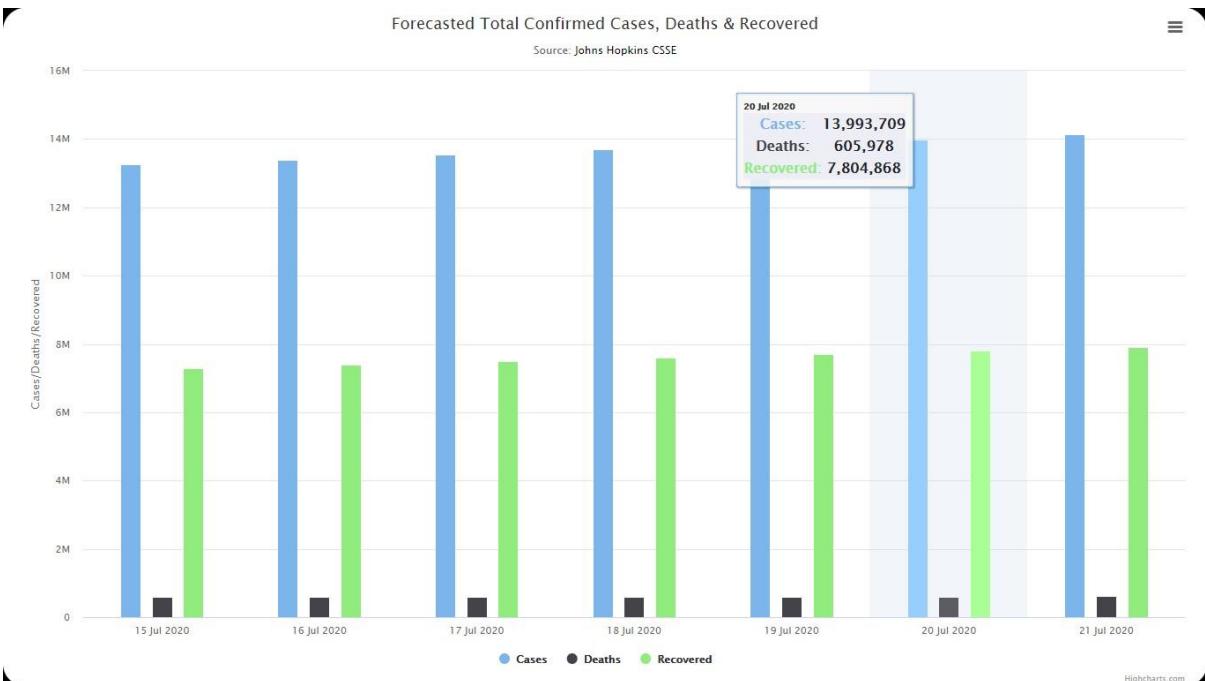
Date	Cases	Deaths	Recovered
15 Jul 2020	13,254,123	582,686	7,286,979
16 Jul 2020	13,397,404	588,163	7,387,096
17 Jul 2020	13,545,252	593,565	7,486,623
18 Jul 2020	13,697,128	598,182	7,593,485
19 Jul 2020	13,846,657	602,019	7,694,828
20 Jul 2020	13,993,709	605,978	7,804,868
21 Jul 2020	14,148,339	611,292	7,917,347

- We have used the Facebook Prophet model to predict the outcome for confirmed cases, deaths and recoveries for a seven-day period.
- This prediction has been done for both India and the entire World.





- The predictions are displayed using graphs generated by the Fb Prophet model showing the predictions as well as trends.



- They are displayed using multiple bar charts.

Description of the Facebook Prophet Model



Introduction

Understanding time based patterns is critical for any business. Questions like how much inventory to maintain, how much footfall do you expect in your store to how many people will travel by an airline – all of these are important time series problems to solve.

This is why time series forecasting is one of the must-know techniques for any data scientist. From predicting the weather to the sales of a product, it is integrated into the data science ecosystem and that makes it a mandatory addition to a data scientist's skillset.

If you are a beginner, time series also provides a good way to start working on real life projects. You can relate to time series very easily and they help you enter the larger world of machine learning.

Prophet is an open source library published by Facebook that is based on **decomposable (trend + seasonality + holidays) models**. It provides us with the ability to make time series predictions with good accuracy using simple intuitive parameters and has support for including impact of custom seasonality and holidays!

In this article, we shall cover some background on how Prophet fills the existing gaps in generating fast reliable forecasts followed by a demonstration using Python. The final results will surprise you!

What's new in Prophet?

When a forecasting model doesn't run as planned, we want to be able to tune the parameters of the method with regards to the specific problem at hand. Tuning these methods requires a thorough understanding of how the underlying time series models work. The first input parameters to automated ARIMA, for instance, are the maximum orders of the differencing, the auto-regressive components, and the moving average components. A typical analyst will not know how to adjust these orders to avoid the behaviour and this is the type of expertise that is hard to acquire and scale.

The Prophet Forecasting Model

We use a decomposable time series model with three main model components: trend, seasonality, and holidays. They are combined in the following equation:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

- **$g(t)$** : piecewise linear or logistic growth curve for modelling non-periodic changes in time series
- **$s(t)$** : periodic changes (e.g. weekly/yearly seasonality)
- **$h(t)$** : effects of holidays (user provided) with irregular schedules
- **ϵ_t** : error term accounts for any unusual changes not accommodated by the model

Using time as a regressor, Prophet is trying to fit several linear and non-linear functions of time as components. Modeling seasonality as an additive component is the same approach taken by exponential smoothing in Holt-Winters technique . We are, in effect, framing the forecasting problem as a curve-fitting exercise rather than looking explicitly at the time based dependence of each observation within a time series.

Trend

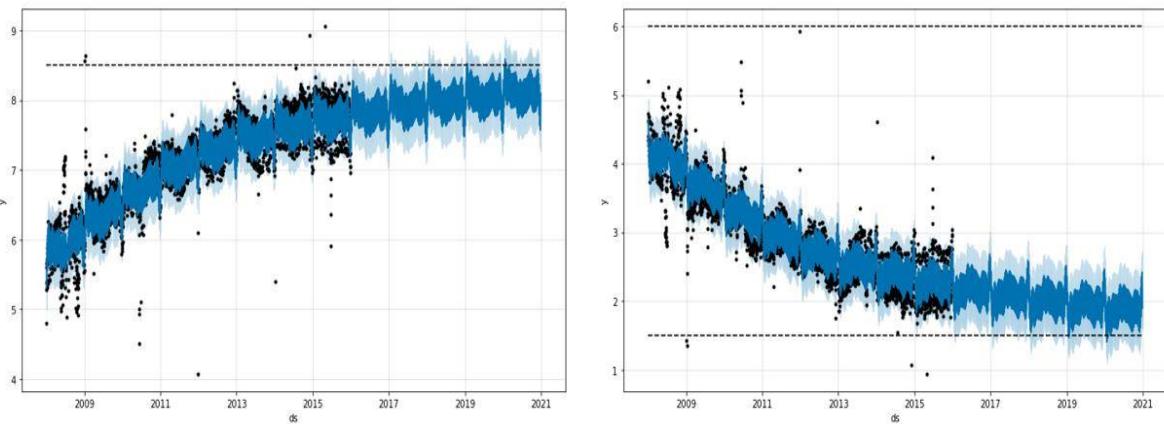
Trend is modelled by fitting a piece wise linear curve over the trend or the non-periodic part of the time series. The linear fitting exercise ensures that it is least affected by spikes/missing data.

Saturating growth

An important question to ask here is – Do we expect the target to keep growing/falling for the entire forecast interval?

More often than not, there are cases with non-linear growth with a running maximum capacity. I will illustrate this with an example below.

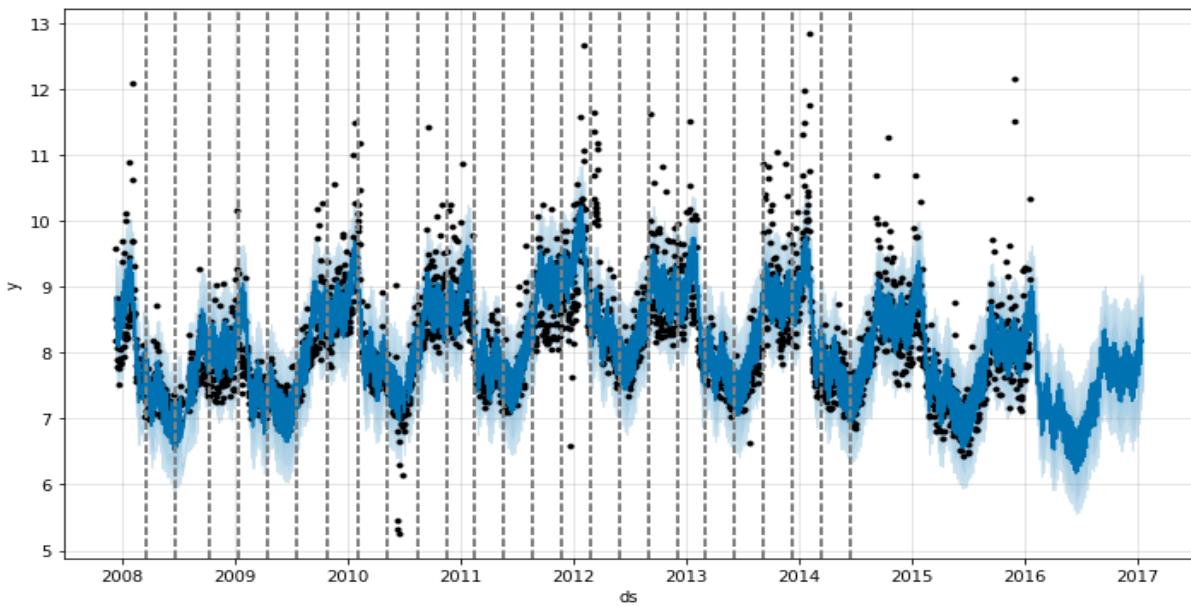
Let's say we are trying to forecast number of downloads of an app in a region for the next 12 months. The maximum downloads is always capped by the total number of smartphone users in the region. The number of smartphone users will also, however, increase with time.



With domain knowledge at his/her disposal, an analyst can then define a varying **capacity $C(t)$** for the time series forecasts he/she is trying to make.

Changepoints

Another question to answer is whether my time series encounters any underlying changes in the phenomena e.g. a new product launch, unforeseen calamity etc. At such points, the growth rate is allowed to change. These changepoints are automatically selected. However, a user can also feed the changepoints manually if it is required. In the below plot, the dotted lines represent the changepoints for the given time series.



As the number of changepoints allowed is increased the fit becomes more flexible. There are basically 2 problems an analyst might face while working with the trend component:

- Overfitting
- Underfitting

A parameter called `changepoint_prior_scale` could be used to adjust the trend flexibility and tackle the above 2 problems. Higher value will fit a more flexible curve to the time series.

Seasonality

To fit and forecast the effects of seasonality, prophet relies on fourier series to provide a flexible model. Seasonal effects $s(t)$ are approximated by the following function:

$$s(t) = \sum_{n=1}^N \left(a_n \cos \left(\frac{2\pi nt}{P} \right) + b_n \sin \left(\frac{2\pi nt}{P} \right) \right)$$

P is the period (365.25 for yearly data and 7 for weekly data)

Parameters $[a_1, b_1, \dots, a_N, b_N]$ need to be estimated for a given N to model seasonality.

The fourier order N that defines whether high frequency changes are allowed to be modelled is an important parameter to set here. For a time series, if the user believes the high frequency components are just noise and should not be considered for modelling, he/she could set the values of N from to a lower value. If not, N can be tuned to a higher value and set using the forecast accuracy.

Holidays and events

Holidays and events incur predictable shocks to a time series. For instance, Diwali in India occurs on a different day each year and a large portion of the population buy a lot of new items during this period.

Prophet allows the analyst to provide a custom list of past and future events. A window around such days are considered separately and additional parameters are fitted to model the effect of holidays and events.

Outputs of the Facebook Prophet Model

Here we have used the past COVID-19 cases, deaths and recovered to make a 7-day prediction of the cases, deaths and recoveries. We made the prediction for both World and India. A sample output of the 7-day predictions is provided below.

7-day Predictions for World Cases

ds	y_hat	yhat_lower	yhat_upper
29-09-2020	33353111.4	33281350.51	33423063.87
30-09-2020	33618495.92	33549813.64	33692705.78
01-10-2020	33893952.08	33825058	33972489.78
02-10-2020	34176274.58	34097487.06	34248503.6
03-10-2020	34450135.34	34375081.45	34526075.89
04-10-2020	34706574.56	34627482.4	34782468.84
05-10-2020	34963838.72	34873951.53	35048461.11

7-day Predictions for World Deaths

ds	yhat	yhat_lower	yhat_upper
29-09-2020	1010607.495	1009197.381	1011955.426
30-09-2020	1016651.833	1015261.761	1018070.985
01-10-2020	1022568.976	1021099.813	1024061.276
02-10-2020	1028438.503	1026780.845	1030049.053
03-10-2020	1033618.959	1031878.805	1035338.785
04-10-2020	1037974.924	1035876.341	1039954.204
05-10-2020	1042744.731	1040433.619	1044900.487

7-day Predictions for World Recovered

ds	yhat	yhat_lower	yhat_upper
29-09-2020	23188801.28	23147368.81	23228523.49
30-09-2020	23407262.11	23365986.06	23446509.24
01-10-2020	23627142.99	23583156.72	23671523.74
02-10-2020	23851502.63	23806330.01	23897445.85
03-10-2020	24073342.31	24027181.23	24121326.45
04-10-2020	24278401.98	24225449.93	24327115.15
05-10-2020	24488355.34	24434116.73	24544553.66

7-day Predictions for India Cases

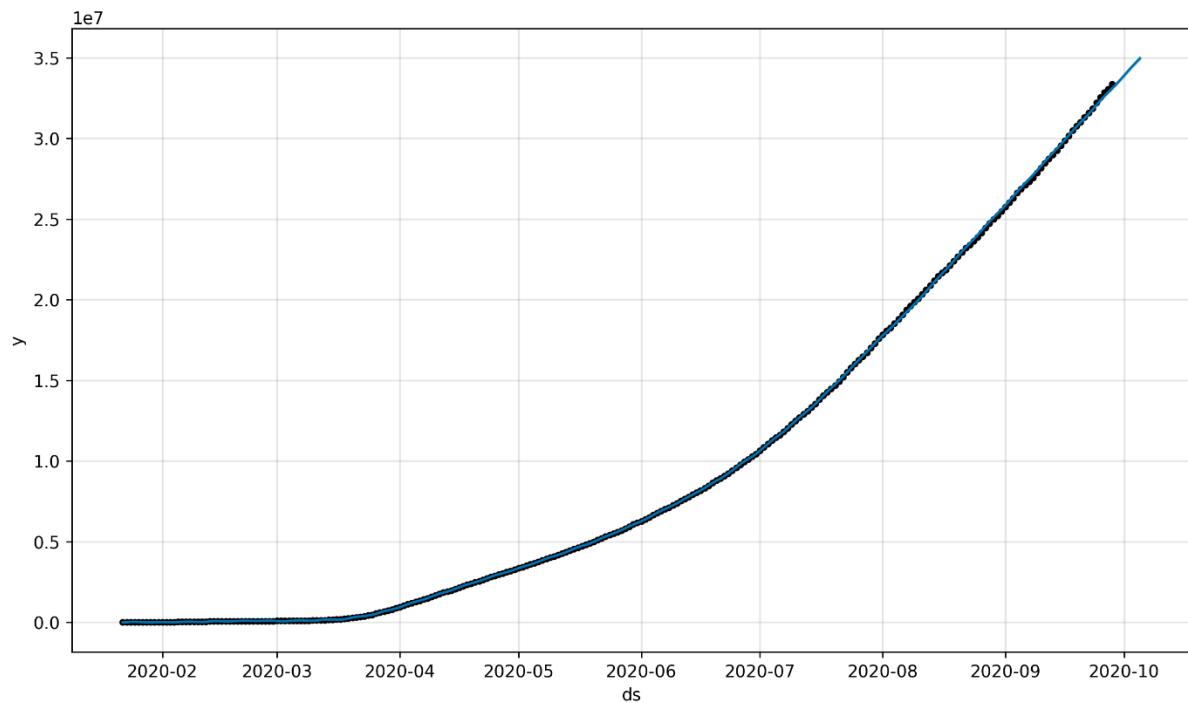
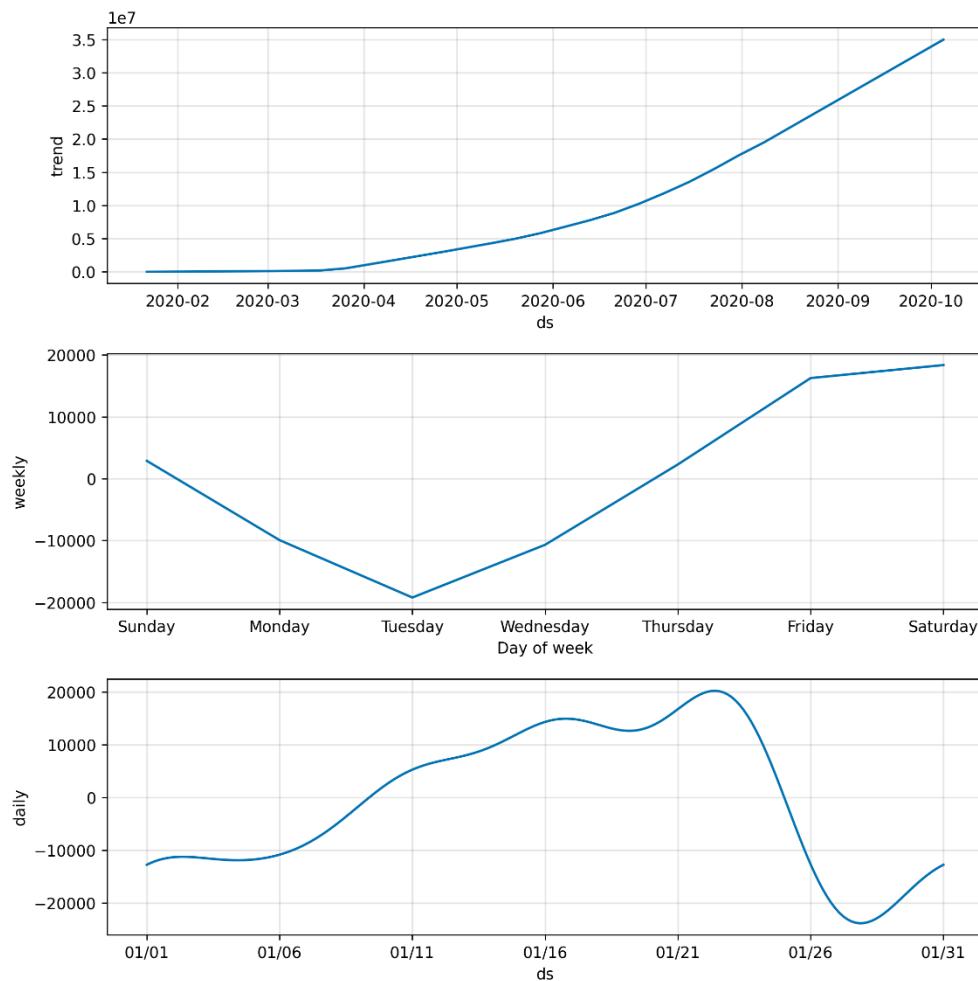
ds	yhat	yhat_lower	yhat_upper
29-09-2020	6692569.262	5352854.685	8278755.307
30-09-2020	7300894.483	5870337.967	9041470.411
01-10-2020	7685454.708	6177289.7	9536012.902
02-10-2020	8041634.921	6420597.471	9975003.118
03-10-2020	8385148.517	6561086.764	10453043.45
04-10-2020	8673844.136	6925460.229	10918302.79
05-10-2020	8897600.435	7129567.472	11361541.66

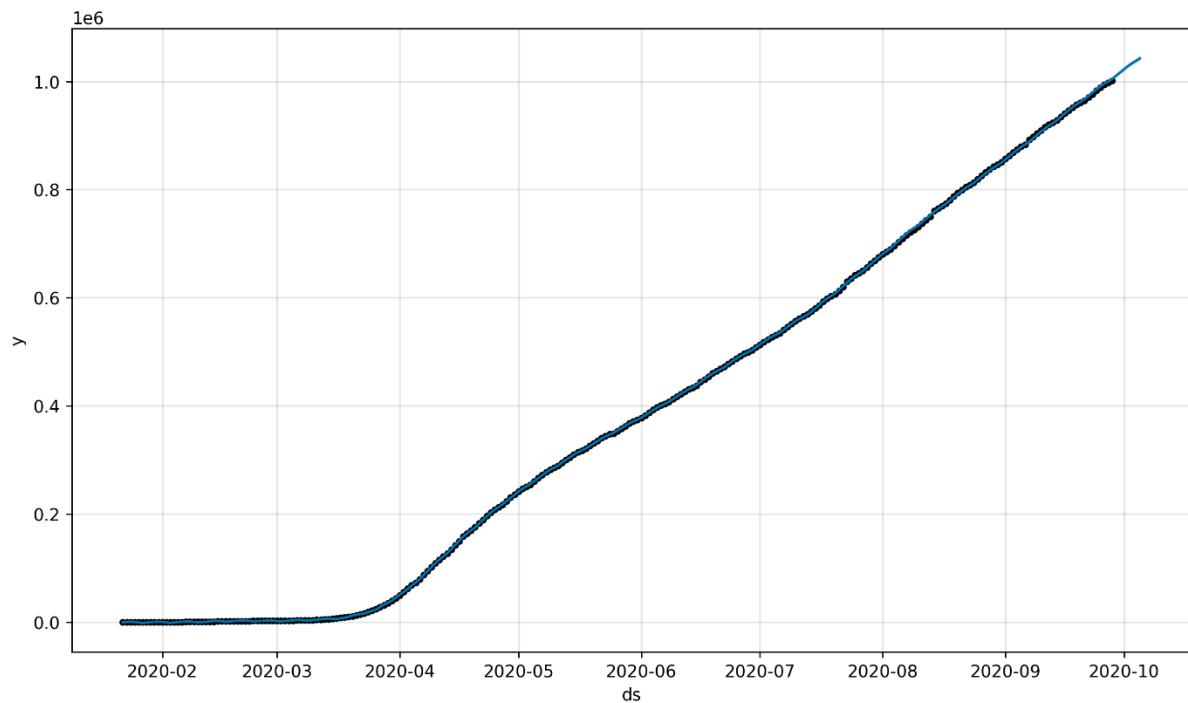
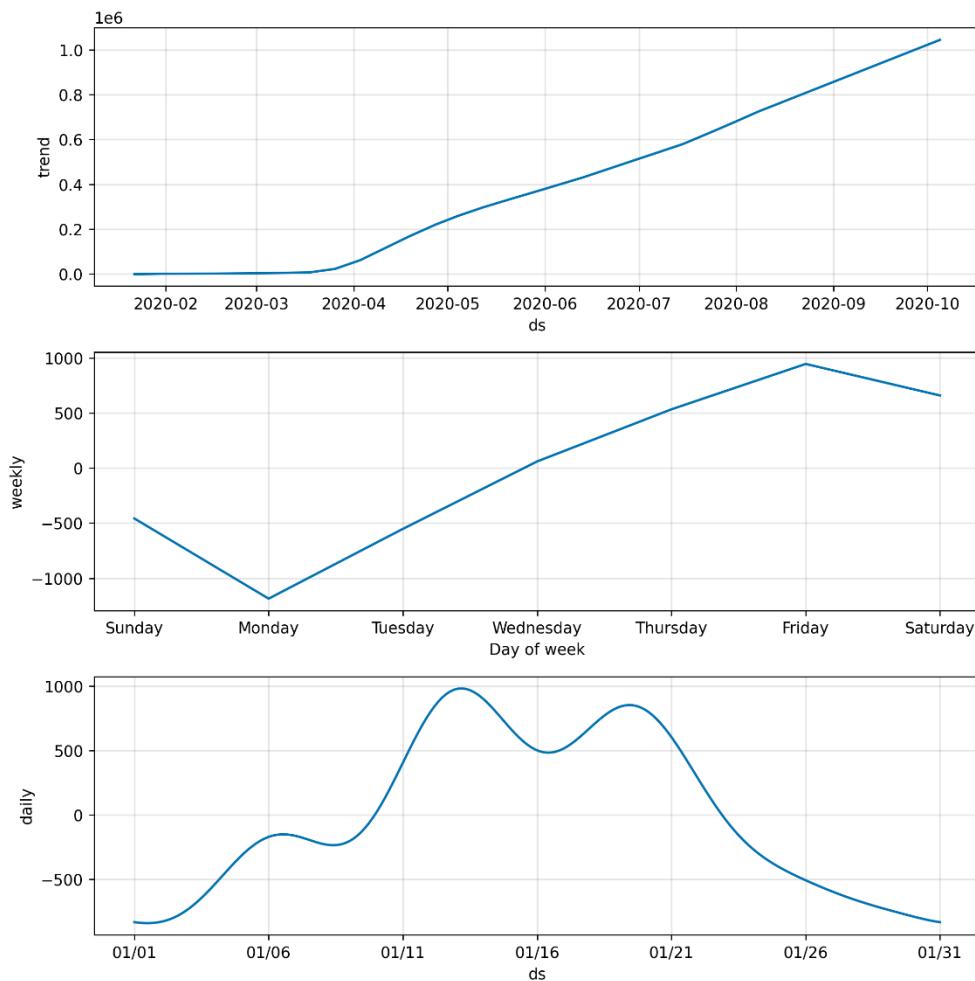
7-day Predictions for India Deaths

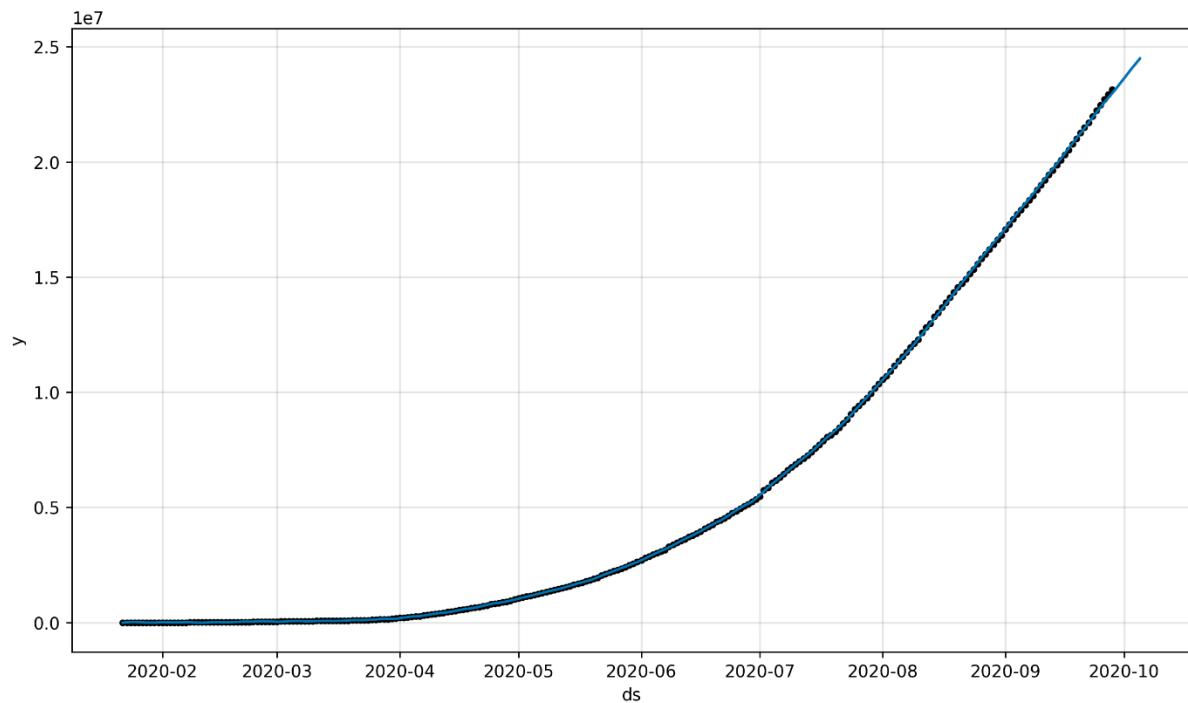
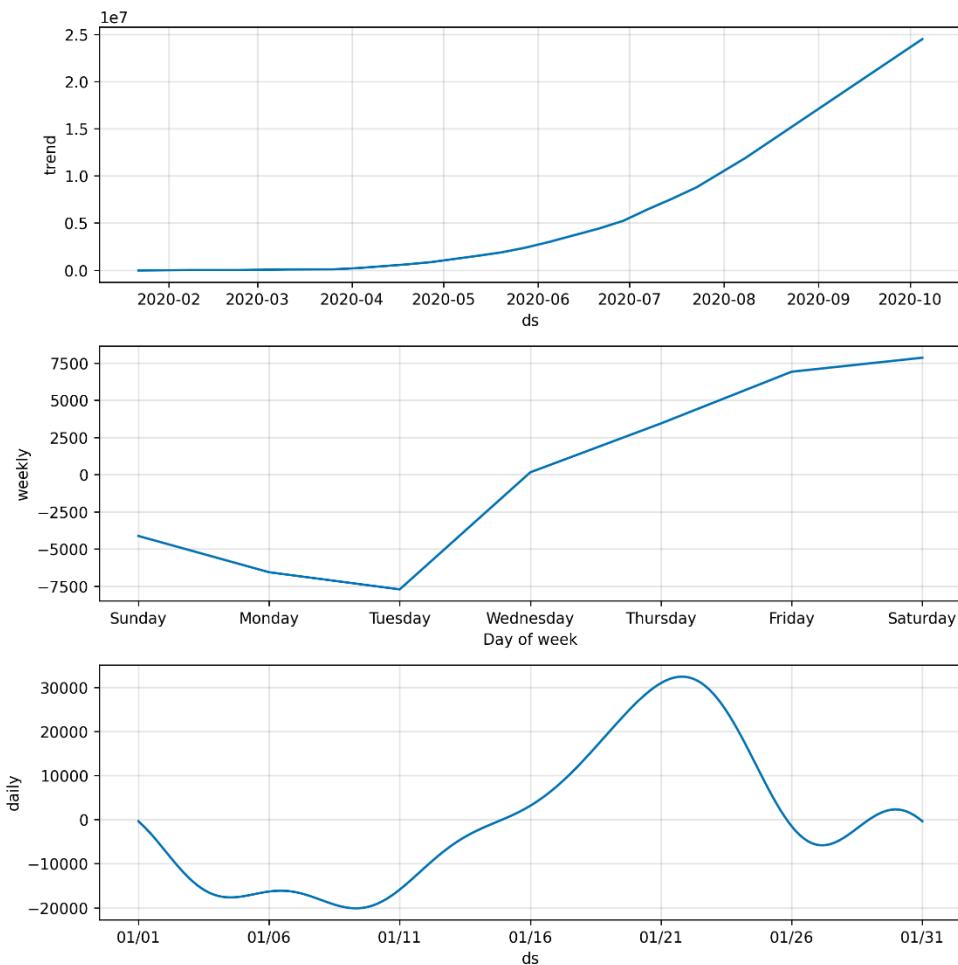
ds	yhat	yhat_lower	yhat_upper
29-09-2020	100613.3051	94568.75417	107414.4068
30-09-2020	102195.0105	95298.83201	108784.0725
01-10-2020	105182.6764	98573.95269	112607.5429
02-10-2020	107689.9141	100334.9915	115893.6874
03-10-2020	108081.3694	100133.3133	116409.3849
04-10-2020	110489.4106	101505.3059	119708.7721
05-10-2020	112524.9778	103392.515	123677.0863

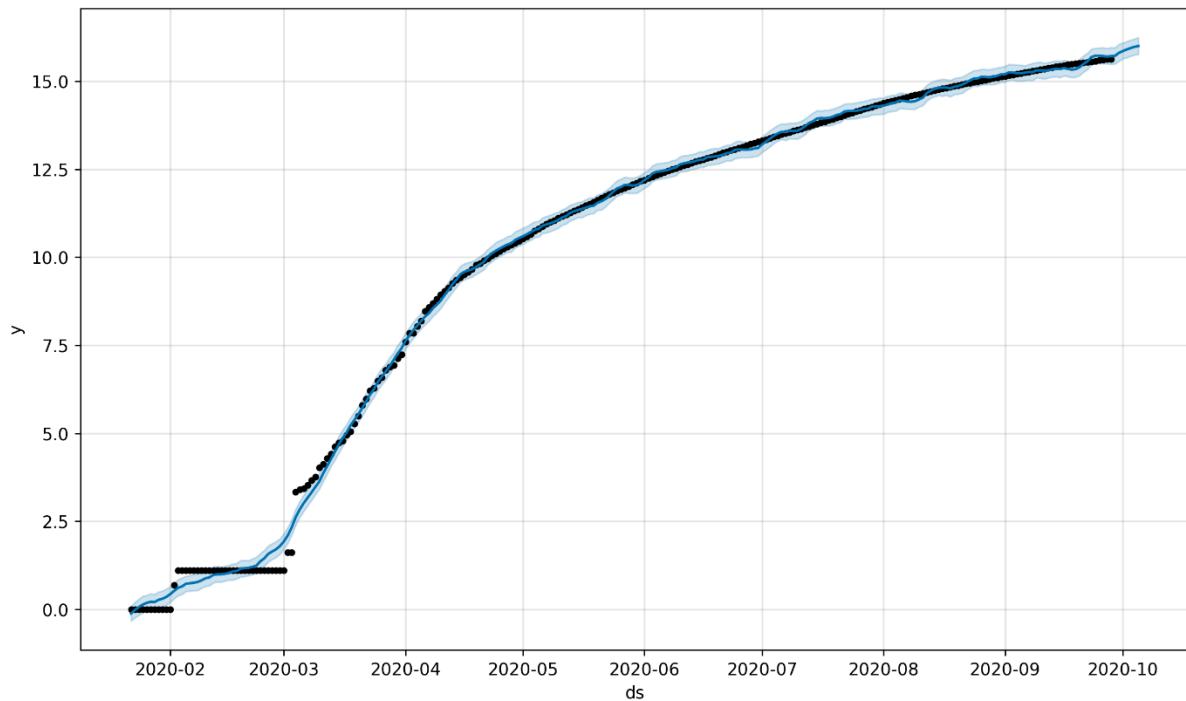
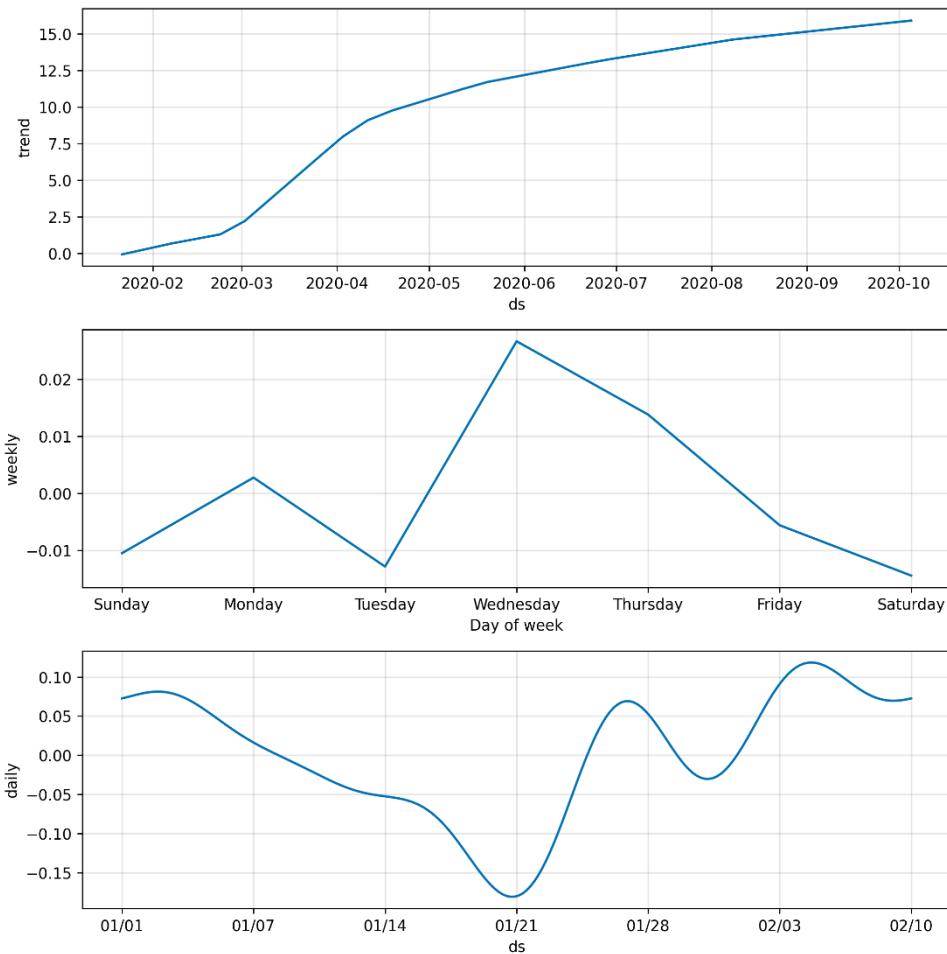
7-day Predictions for India Recovered

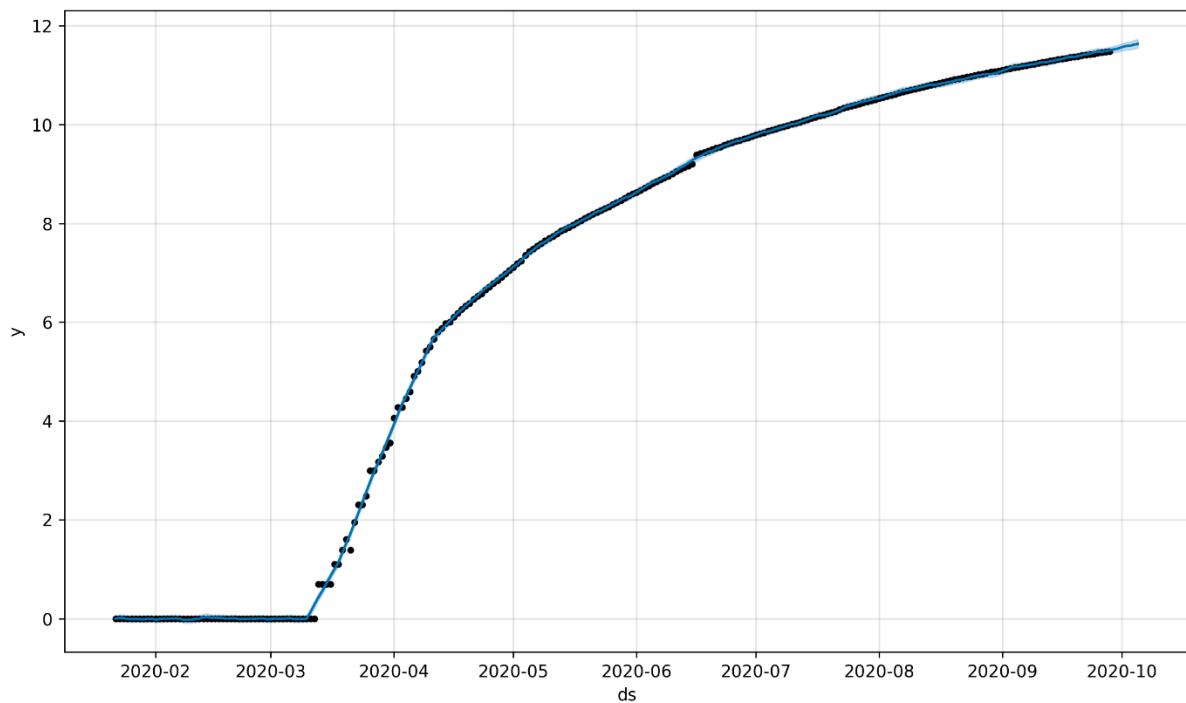
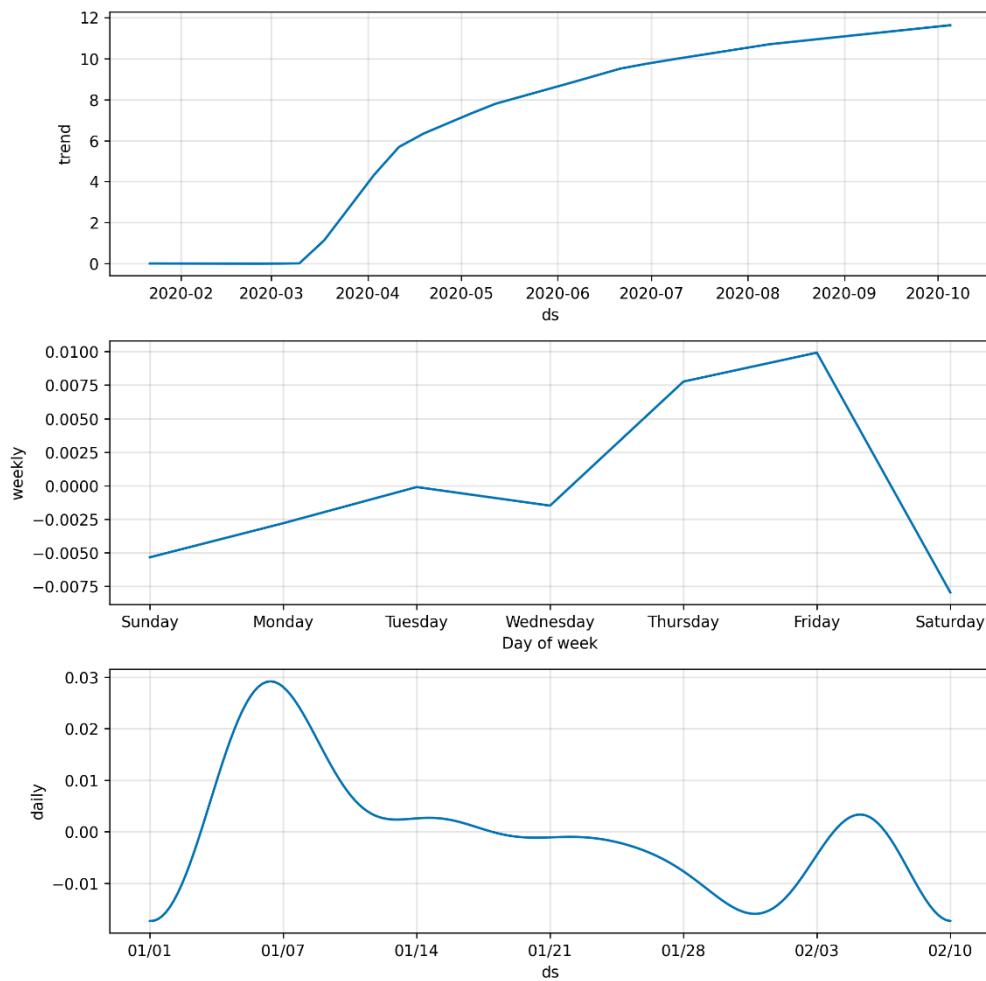
ds	yhat	yhat_lower	yhat_upper
29-09-2020	5563626.025	4768199.405	6468409.972
30-09-2020	5590971.012	4861367.424	6448978.515
01-10-2020	5681598.127	4951442.762	6707582.942
02-10-2020	5845708.936	5016919.957	6717086.113
03-10-2020	5972320	5140113.959	6861408.807
04-10-2020	6444282.754	5534073.54	7532722.444
05-10-2020	6580414.324	5498200.869	7606469.397

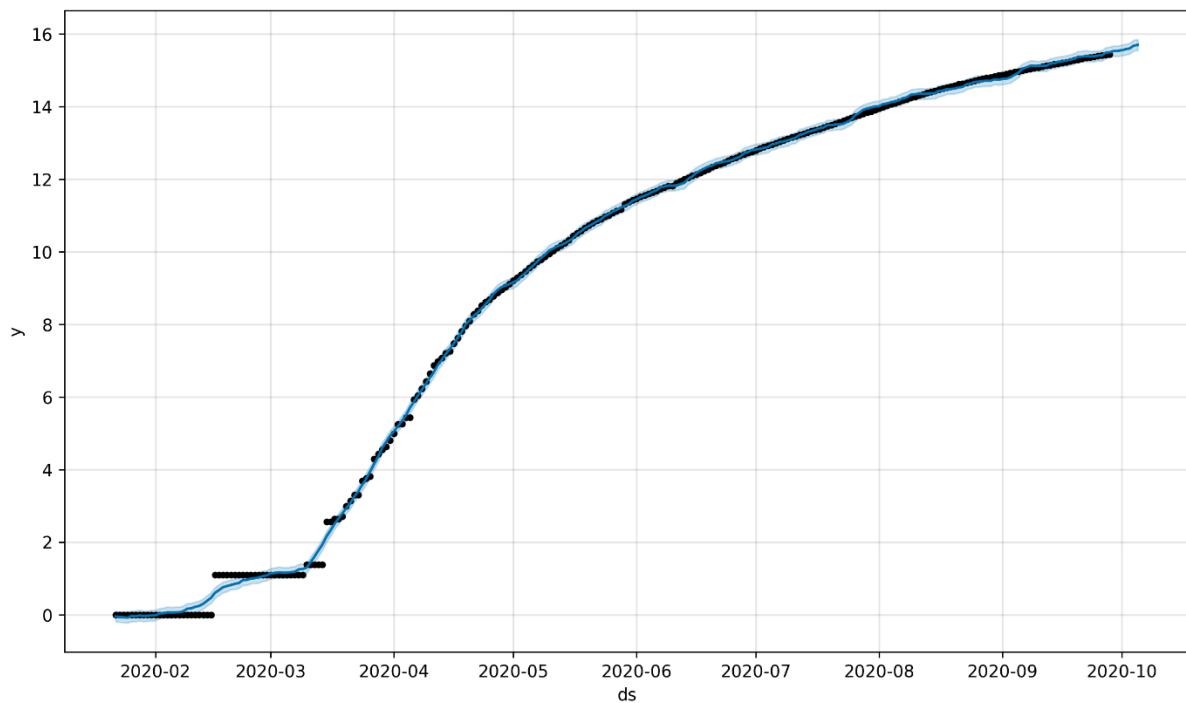
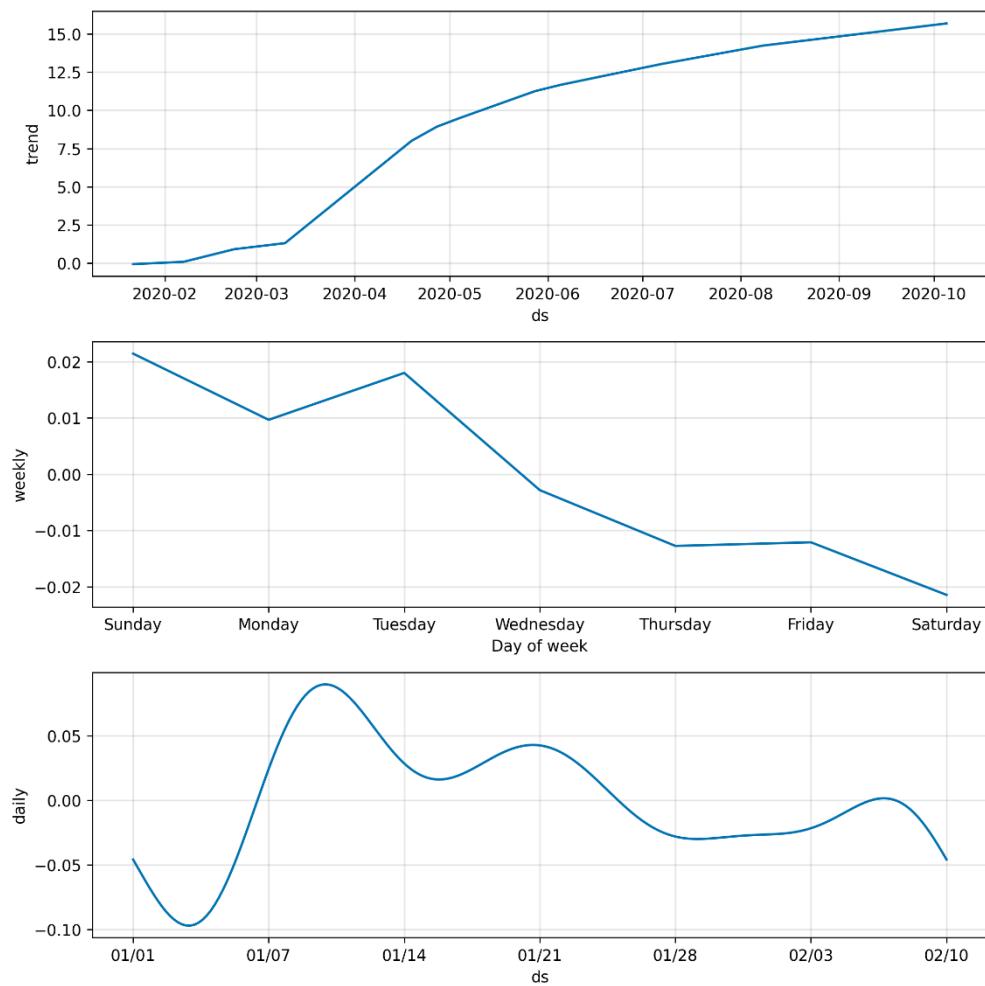
Forecast Plot for World Cases*Forecast Component Plot for World Cases*

Forecast Plot for World Deaths*Forecast Component Plot for World Deaths*

Forecast Plot for World Recovered*Forecast Component Plot for World Recovered*

Forecast Plot for India Cases*Forecast Component Plot for India Cases*

Forecast Plot for India Deaths*Forecast Component Plot for India Deaths*

Forecast Plot for India Recovered*Forecast Component Plot for India Recovered*

Python Packages Used in Our Project



In our project we have used some very well-known packages of Python which are given below.

- **flask** - Flask is a lightweight WSGI web application framework. It is designed to make getting started quick and easy, with the ability to scale up to complex applications. It began as a simple wrapper around Werkzeug and Jinja and has become one of the most popular Python web application frameworks.
- **flask-cors** - A Flask extension for handling Cross Origin Resource Sharing (CORS), making cross-origin AJAX possible.
- **pandas** - Pandas is an open-source, BSD-licensed Python library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc. In this tutorial, we will learn the various features of Python Pandas and how to use them in practice.
- **numpy** - NumPy is a python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices. NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely. NumPy stands for Numerical Python.
- **wordcloud** – package in python used to create word clouds.
- **textblob** - *TextBlob* is a Python (2 and 3) library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.

- **newsapi-python** - News API is a simple HTTP REST API for searching and retrieving live news articles from all over the web. Using this, one can fetch the top stories running on any news website or can search top news on a specific topic (or keyword).
- **tweepy** - An easy to use python library for accessing Twitter API.
- **requests** - Requests is an elegant and simple HTTP library for Python, built for human beings.
- **json** - JSON (JavaScript Object Notation), specified by **RFC 7159** (which obsoletes **RFC 4627**) and by ECMA-404, is a lightweight data interchange format inspired by JavaScript object literal syntax (although it is not a strict subset of JavaScript 1).
- **calendar** - This module allows you to output calendars like the Unix **cal** program, and provides additional useful functions related to the calendar. By default, these calendars have Monday as the first day of the week, and Sunday as the last (the European convention). Use `setfirstweekday()` to set the first day of the week to Sunday (6) or to any other weekday. Parameters that specify dates are given as integers. For related functionality, see also the `datetime` and `time` modules.
- **configparser** - This module defines the class **ConfigParser**. The **ConfigParser** class implements a basic configuration file parser language which provides a structure similar to what you would find on Microsoft Windows INI files. You can use this to write Python programs which can be customized by end users easily.
- **matplotlib** - Matplotlib is one of the most popular Python packages used for data visualization. It is a cross-platform library for making 2D plots from data in arrays. It provides an object-oriented API that helps in embedding plots in applications using Python GUI toolkits such as PyQt, WxPython or Tkinter. It can be used in Python and IPython shells, Jupyter notebook and web application servers also.
- **re** - This module provides regular expression matching operations similar to those found in Perl.
- **fbprophet** - Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well.

Acknowledgement

We would love to convey our deepest appreciation to all those people whose open source codes inspired and helped us to complete this ambitious project of ours. We are greatly indebted to John Hopkins University for their fabulous accumulation and structuring of the data, which they made available for all. A big shout out to the data science team at Facebook for creating such a beautiful model which anyone would be able to use any way they like.

In addition, we would like to thank Dr. (Prof.) Sisir Kumar Samanta, Dr. (Prof.) Chandranath Pal and Dr. (Prof.) Ganesh Dutta who provided us with the opportunity and scope for our project. We thank the Department of Statistics, University of Kalyani for allowing me to continue this project beyond the stipulated class schedule. I am also grateful to different journals and sites for the subject references.

It will be disgraceful if we do not mention about the moral support provided by our family and the invaluable suggestions given by our friends.

Declaration

I hereby declare that all the data included in this project are true to the best of my knowledge and shall not be circulated outside the realms of this project.

Date: 26th August, 2020

Abhijit Show

Place: Thakurpukur