

COVID-19 Predictions in India

Anuj Tanwar

Bellevue University

Business Problem

COVID-19 outbreaks don't only impact lives of people but disrupts economy and healthcare infrastructure of the country. So, it is important to study the pandemic and predict potential future infectious disease outbreaks. A comprehensive understanding of the spread and past count of cases will help in predicting recent future and enable administration by giving a heads-up to stay alert and can provide sufficient time to facilitate support interventions.

Keywords: Covid-19, Predictions

Background/History

India has seen an increase in COVID-19 case again during the start of year 2022. A simple google search can tell us that on April 4th 2022, there were only 795 new cases, on Apr 14th 2022 there were 949 and in last week count of new cases have been on constant rise and going above 2000 per day now. This trend shows the cases have been rising. As per Ministry of Health and Family Welfare Government of India (<https://www.mohfw.gov.in/>) there had already been more than half a million documented deaths due to COVID and active cases are on rise. Government of India has also issued new guidelines and restrictions. Referring to the below google graph, we can see that COVID spread had increased in India during the summer months in last 2 years. So, the concerning questions that we have are, are we going to have another wave of COVID? If so, then how severe it can be? When will we see the peak? How long did the previous waves last? What was the trend in death toll every day? What was the cured trend?

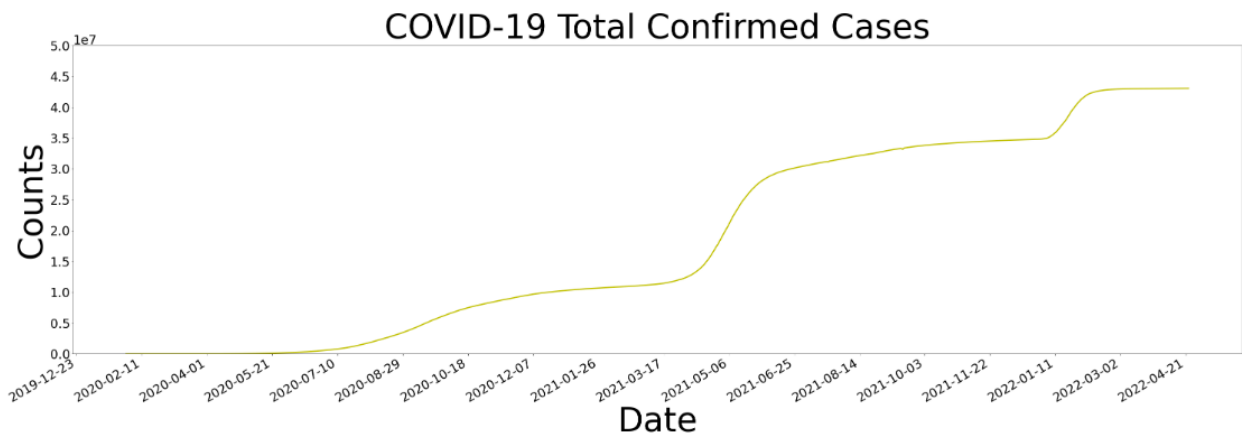
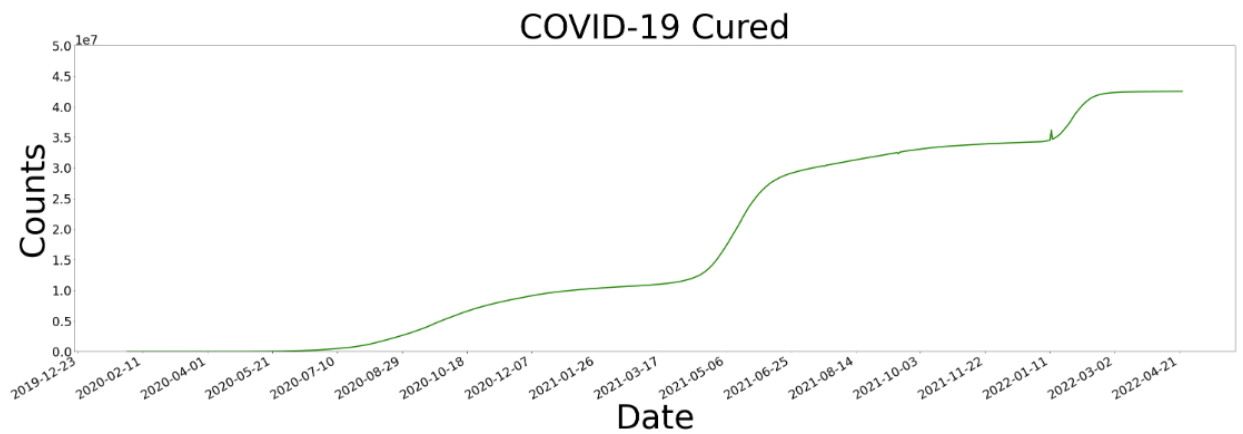
Data Explanation

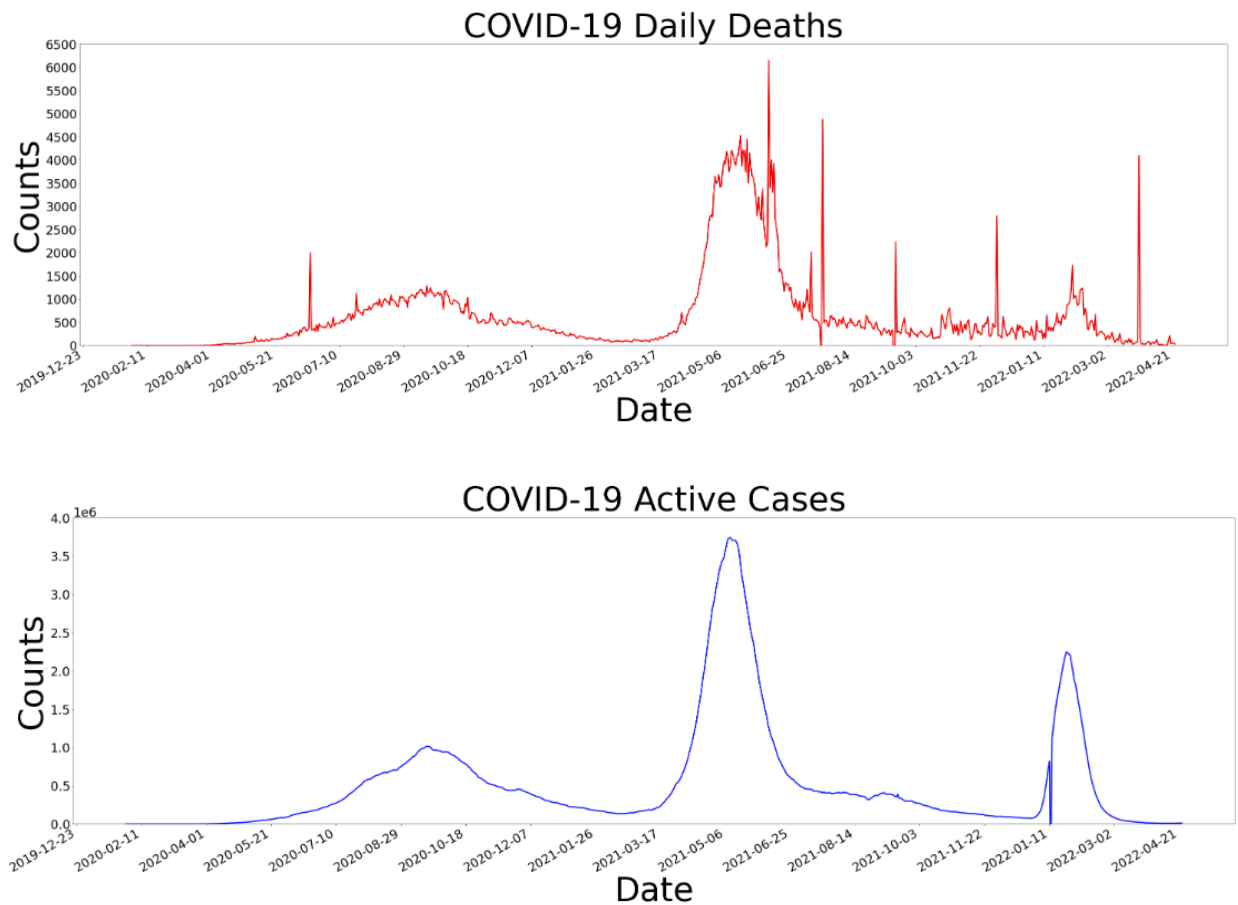
COVID-19 Data is collected mainly from **Datameet** website and Ministry of Health and Family Welfare Government of India website.

Data Cleaning: Keys and values were read separately from JSON dataset and then combined into a dataframe. At this point, dataframe has multiple lines for each date, one row for each type(active cases, deaths, cured, total confirmed cases). This need dataframe to be pivoted on date field to convert types to individual columns as shown below.

	dt	type	value		dt	active_cases	cured	death	total_confirmed_cases
0	2020-01-30	active_cases	1	➔	2020-01-30	1.0	0.0	0.0	1.0
1	2020-01-30	cured	0		2020-02-02	2.0	0.0	0.0	2.0
2	2020-01-30	death	0		2020-02-03	3.0	0.0	0.0	3.0
3	2020-01-30	total_confirmed_cases	1		2020-03-02	5.0	0.0	0.0	5.0
4	2020-02-02	active_cases	2		2020-03-03	6.0	0.0	0.0	6.0
...
3347	2022-04-22	total_confirmed_cases	43052425		2022-04-19	11860.0	42511701.0	521966.0	43045527.0
3348	2022-04-23	active_cases	15079		2022-04-20	12340.0	42513248.0	522006.0	43047594.0
3349	2022-04-23	cured	42517724		2022-04-21	13433.0	42514479.0	522062.0	43049974.0
3350	2022-04-23	death	522149		2022-04-22	14241.0	42516068.0	522116.0	43052425.0
3351	2022-04-23	total_confirmed_cases	43054952		2022-04-23	15079.0	42517724.0	522149.0	43054952.0

Univariate Time Series: Each variable was plotted in time-series to study the trend.





Data Preparation

Joined case dataset with vaccine data(mohfw_vaccination_status.json) to get additional information. Dropped features not useful for analysis. Transformed features such as report_date and used it as key to join case dataset with vaccine dataset. Dropped records with invalid values such as counts cannot be negative. Used imputer to fill in NaNs with mean. Used MinMaxScaler() to rescale each column. Filled missing data with mode of respective column.

Method and Analysis: Model building and evaluation

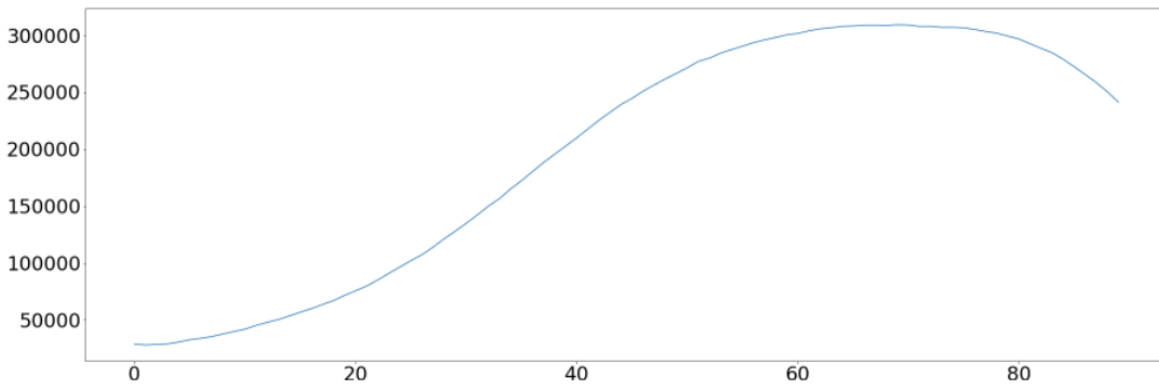
Model Used: Long short-term memory (LSTM). It is an artificial neural network used in the fields of artificial intelligence and deep learning. I have used sequential model from tensorflow.keras package. Added a long short-term memory layer with 100 memory units. Used rectified linear activation function (RELU) which will output the input directly if it is positive, otherwise, it will

output zero. Used 20% dropout. Compiled the model with adam optimizer. Data Fitting used EarlyStopping class from keras callbacks. This enables model to stop training when a metric has stopped improving.

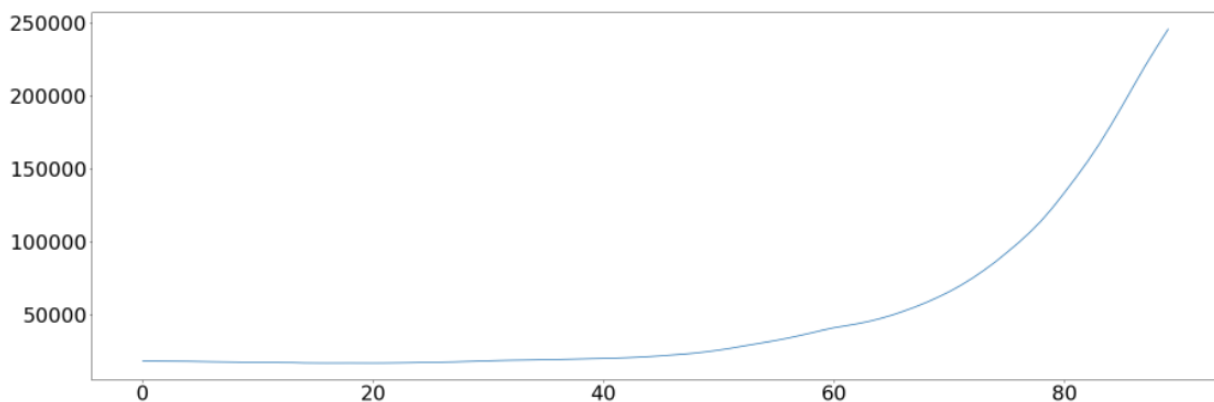
Conclusion

Model clearly suggest that vaccine reduces covid counts. Without first dose the 90 days total case prediction is 17.7 million but with just first dose it comes down to 4.5 million.

Predictions Without Vaccine Data



Predictions With Vaccine Data



Assumptions

One of the assumption is a uniform mixing of the infected and vulnerable populations along with the assumption that the total population is constant in time. Another assumption is that there was no biasing during data collection and mentioned website reflect actual data.

Recommendations

Case estimations are purely on past data, future numbers depend on lot of factors such as restrictions (mask mandate, social distancing, etc.) and vaccine dosage and boosters. Those factors can affect the numbers. Model considered only first vaccine dose and based on that we can recommend encouraging public to take vaccine and booster shots whichever is applicable.

Challenges/New Opportunities

Model can be enhanced in future predictions considering factors like subsequent vaccine dosage, Enforced Restrictions and guideline such as mask mandates, lock downs, number of people allows in public transportation such as buses and trains, weather conditions such as summer or winter counts.

Implementation Plan

Plan to start with data collection by reading data from both the websites. Followed by data cleaning step as explained in data explanation step above. Next step would be to prepare data by handling NaNs, dropping unnecessary features, transforming and joining wherever required. At this stage data would be ready to build and evaluate the model. Final stage would be conclude the findings.

Ethical Considerations

Uneasiness around the sudden loss of personal liberties that had been taken for granted by citizens in many countries. Regulation and protocols must be in place to avoid panic and frustration in public. Privacy is at risk with digital surveillance like tracing apps and enforcing people to report

covid results. Personal biasing due to scare of COVID or otherwise. Personal biasing during the analysis can influence the result drawn from analysis.

References

- Economic Impact of COVID-19 (2022, Feb 07) KFF. <https://www.kff.org/global-health-policy/issue-brief/economic-impact-of-covid-19-on-pepfar-countries/>
- Datameet Covid 19 (2022, Oct 21) Github.
<https://github.com/datameet/covid19/tree/master/data>
- Covid 19 Positivity Rate (2023, Jan 27) MOHFW. <https://www.mohfw.gov.in/>