

COVID-19 PROJECTIONS: MODEL PLAN

1. **Model Name:** COVID-19 Predictions Model

2. **Model Objective /Purpose (Description & Data):**

The model has been built for the purpose of predicting the number of people infected by the deadly Corona Virus for all the states throughout India. The model predicts the number of confirmed people infected by COVID-19 from Aug 08 to August 27, on daily basis for each state.

The model leverages the no. of confirmed cases, no. of casualties, no. of recoveries from the virus in the past few months (Mar'20 – Aug'20) on a daily basis, presence of Tier1/2/3 cities in the respective states and the spread of infection in these cities to identify the total number of COVID-19 positive cases in the coming **20** days for each state throughout India.

3. **Modelling Technique:**

The model is an ensemble of regression models created through the following machine learning algorithms:

- a. Prophet, a time series forecasting model
 - i. Different Models for Different States
- b. ARIMA, another time series forecasting model
 - i. Different Models for different States
- c. Multi Output XGBoost Regressor (Problem is transformed into supervised learning problem)
 - i. Single Multioutput model
- d. Multiple XGBoost models in Loops
 - i. 20 Models based on the prediction horizon (for predicting cases 1 day after the current situation, 2 day after the current situation, etc.)

4. **Model Details:**

- a. This model is built for Indian states only. The model development data is split into 80-20 split for evaluating the model. Model data includes all Indian States with history of COVID infected cases in the previous months.
- b. Models have been validated on **July 15 to Aug 6**
- c. **Dependent variable:** No of people infected with COVID-19(new confirmed cases) in the next 20 days for each state.

- d. The model is an ensemble of 4 different algorithms, a weighted average of the Prophet, ARIMA, Multi Output XGBoost model (varying weightages depending on the state) & Multiple XGBoost models varying on the future interval where prediction is required

5. Variables used in the Model:

- a. Periodicity/Seasonality take into consideration (On Weekly Basis)
- b. State-wise no. of confirmed COVID 19 cases in the past 3 months (May20-Jul20) on daily basis
- c. State-wise no. of casualties from COVID 19 in the past 3 months (May20-Jul20) on daily basis
- d. State-wise no. of recoveries from COVID 19 in the past 3 months (May20-Jul20) on daily basis
- e. State-wise no of COVID 19 cases in Tier1/2/3 cities in the past 3 months (May20-Jul20) on daily basis
- f. State-wise % of active cases as compared to cumulative cases
- g. State-wise % of deaths as compared to cumulative cases

6. Model Performance:

Model is evaluated on Mean Absolute Percentage Error (MAPE).

Weighted Average for each model (for each state):

$$W_i = (1/E_i) / (1/E_1 + 1/E_2 + 1/E_3)$$

Where E_i is the MAPE of each model, corresponding to a state

Model Prediction for each day for each state:

$$\text{\#confirmed cases} = \sum W_i P_i$$

Where W_i is the weight of each model and P_i is the Prediction of each model corresponding to a state for the day