

PREDICTIVE MEASURES FOR MITIGATION OF COVID-19 COMMUNITY TRANSMISSION

Manik Kapil

April 26, 2020

1. INTRODUCTION

1.1 BACKGROUND

Coronaviruses are family of viruses that cause diseases generally in birds and animals, which in some cases get transmitted to humans, like the one the whole world is currently facing. COVID-19 or Coronavirus Disease 19 is the name WHO (World Health Organization) coined for this new disease which as of 26 April 2020, consumed at least 203,705 lives and more than 2,935,876 confirmed cases. It causes respiratory tract infections that can range from mild to lethal. India being world's second most populous country with its high population density was at high risk if the contagion spread. A pre-emptive 21 days lockdown and then extended 19 days lockdown indeed helped in containing the rate of spread of the virus. IMA (Indian Medical Association) & ICMR (Indian Council of Medical Research) have been on their feet even before the first case was recorded in India, still as we speak there are at least 27.3k active cases and as many as 825 deaths and we are only in 2nd Stage of Transmission. To tackle this unprecedented situation, there were multiple novel techniques devised and used, therefore, it is imperative to estimate the existing cluster of cases in the country, investigate the treatment, recovery and deaths in those areas and to explore if there are any predictions can be made for new Quarantine and isolation facilities. A large number of daily wage workers suddenly boggled by this situation find them without food and shelter. We'll also try and predict the areas where migrant workers might take food and shelter while maintaining social distancing.

1.2 PROBLEM

The world has come to a halt amid COVID-19. All economic machineries have come to a standstill and the liquidity in market has dried out. Hence, the breakage of transmission chain and release in economic activity is one following the other. Hence, this project aims at predicting probable locations that can be used as quarantine facilities, food shelters and even testing facilities.

1.3 INTEREST

Every government in the world, every CEO of every company, every worker either daily-wage or salaried, every front-line COVID worker is interested in finding some way that can break the existing transmission chain and slowly the world returns to normalcy. Hence, a large number of people are stake holders in this.

2. DATA ACQUISITION AND CLEANING

2.1. DATA SOURCES

There were multiple datasets used in this notebook. First, the datasets were scrapped, cleaned, merged in some cases and then were used for analysis.

Datasets included:

2.1.1 Daily COVID-19 patient database across Indian states

2.1.2 Case time series database, which included patient data attributed by date

2.1.3 Covid-19 Daily tests dataset

2.1.4 Some other supporting datasets which included, Indian coordinates.

2.2. DATA PRE-PROCESSING

2.2.1 Extensive data cleaning was done as these datasets had a large number of missing values, values misplaced and improper and non-formatted values.

2.2.2 There were a large number of patients that had no information pertaining to their addresses apart from just the state name. Hence, it became difficult in keeping a track of these patients. There were some columns that had all the missing data except a few and we had to alter those columns to make our analysis easier. The only way to plot through folium is by the means of coordinates of a location. Hence, it became imperative to append the coordinates for each state.

2.2.3 The Data format was Timestamp in case time series dataset and there was no way to move forward with this format, hence, I had to alter it down to just the date part to move further. This was achieved with the help of a self-defined function. There were some outliers and some missing values in this dataset. For e.g. there were missing values for 'Daily Confirmed' and 'Daily Recovered' for some dates. Hence, we had to get rid of those to find our desired result.

2.2.4 The daily tests dataset was the most messed-up. With duplicate columns and very few countable values in those had to be cleaned. The Time column in this dataset was again given in timestamp format and had to be formatted again for further use. The datatypes of some columns were misplaced and were creating problems in mathematical analysis. Hence, I had to assign proper datatype pertaining to that column. Multiple missing values had to be dealt with.

2.2.5 Finally after these datasets, I used Foursquare data for specific locations to get better insights about current scenario and then made recommendations. Even in this data, the proper address attributes like name of city, state were missing. But since, the coordinates were provided, I was able to draw a map with those locations as markers.