CORONA CASE PREDICTOR

Sumer Sharma
Department of IT
Maharaja Agrasen Institute of Technology,
New Delhi, Delhi, India

Namita Goyal
Department of IT
Maharaja Agrasen Institute of Technology,
New Delhi, Delhi, India

Abstract— **What will be the further impact of the novel coronavirus (COVID-19) in India? To answer this question, we need an accurate analysis of the rate of death and recovery. At the same time, since the future does not usually repeat itself in the same way as in the past, so there is no certainty. The COVID-19 epidemic after spreading its roots to 206 countries around the world, has started again with more deadly waves than previous. Though vaccines are available now but still no one knows for how much time period certain vaccine can provide antibodies. So, the battle is still going on. Disease and death not only threaten people but also their economic impact. Even though if one got recovered from disease but post covid symptoms are the one which are haunting even more. Based on the official data model, diagnostic techniques are used to create a predictable but decisive prediction model for the spread of COVID-19 in India. The second wave of COVID-19 hit in the states of India during March and has since spread again to all other provinces with a great havoc and the situation is getting worse in countries with high global migration.**

I.   INTRODUCTION

The arrival and spread of the 2019 novel coronavirus (SARS-CoV-2) has created a global health crisis with a dramatic increase in the number of cases and deaths since it was first discovered in Wuhan, China, in December 2019. By the end of 2019, the new coronavirus (COVID-19) was widespread in China, and a large number of people were infected. During its first year of spread, domestic outbreaks were successfully controlled and many precautionary actions were taken while the new mutated coronavirus is rapidly spreading in many countries again. The mutated coronavirus has posed a serious threat to the health and safety of humans worldwide as a result of its amazing potential for transmission and potential harm. Research into domestic and international epidemics and future development trends has become a hot topic in current research. However, a different pattern of the spread of the epidemic was observed between countries and between countries as well. Different cultural and social diversity may have the same effect. Similarly, public health policies also play a role in the spread of the disease. As health is a matter for the state, countries within the country are free to take the necessary steps to curb the epidemic that spreads beyond the policy of the central government.
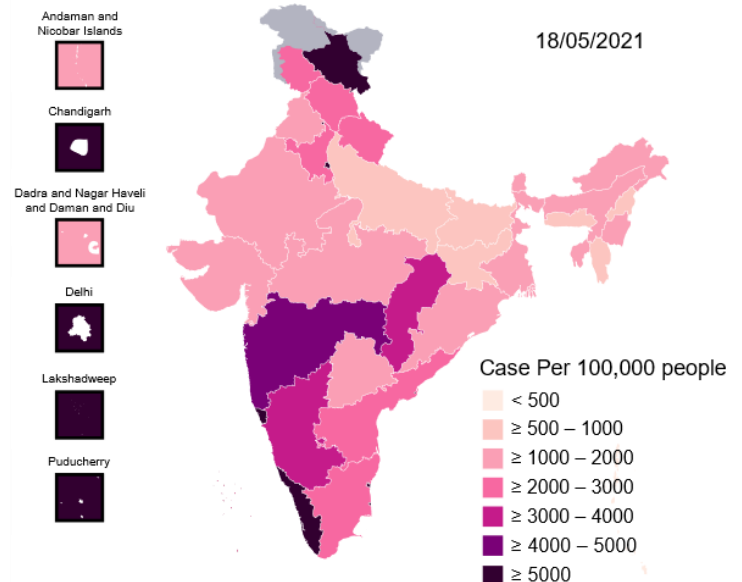
https://upload.wikimedia.org/wikipedia/commons/e/e4/India_COVID-19_cases_density_map.svg

Though vaccine research is at a great speed, scientists are giving their day and night for this. All our frontline soldiers are doing a great job and giving their maximum to prevent the spread of disease and saving as many people as the can. Currently, a key feature of the epidemiological risk factor for 2019 novel coronavirus disease involves close contact with people with an incubation period of 2–14 days. The mortality rate has expectedly rose from 2% to 3% and India with over 3.5 lakhs death till now according to our data and still counting. The various drugs are being tested in accordance with previous clinical studies of SARS and MERS, however, there is no such evidence for a 100% improved health effect. The apparent risk of contracting the disease has led many governments to implement various control procedures such as segregation and locking mechanisms. Despite drastic measures to prevent the world, the frequency of novel coronavirus continues to rise, with more than 17.4 crore cases confirmed and more than 37,00,000 deaths worldwide on June 01, 2021. Although countries around the world have improved laboratory capacity

building and response processes, and many laboratories have come up with various vaccines namely COVISHIELD and COVAXIN being most used in India. As of now India is behind vaccine drive, only 17% of population has been vaccinated with first dose and around 4% with both does. For controlling the spread of disease one of the major factors is via Herd Immunity which is achieved when 50-70 % of population has antibodies so that the spread of the disease can be under our control. Government is paying huge amount for its citizen to get free vaccine and as quickly as possible. Vaccine production in India is at par and export of other vaccine is also going parallelly for SPUTNIK-V, PFIZER, MODERNA so that as soon as possible at least 50% of the population get vaccinated.

There is a need for new solutions to develop, manage and analyse large data in a growing network of infected subjects, patient information, their social movements, and integration with clinical trials and health, medical and community data. Many data sources including, messaging, online communication, social media and web articles can be very helpful in analysing the growth of infection by social behaviour. To summarize this data with Machine Learning (ML) and Artificial Intelligence (AI), researchers can predict where and when the disease can spread, and inform those regions to match the necessary preparations. As the model predicts overall cases at any given time, it will help the government deal with the current situation and maintain the country's economy and health. Other measures of protection will be taken seriously and the public will be notified before any critical situation.

II. PROPOSED ALGORITHM

Random Forests:

A random forest is a collective learning method of sorting, reversing and other activities that work by building multiple decision-making trees during training. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data features can affect their performance. Random Forests are widely used as "black-box" models for businesses, as they make sensible predictions for a wide range of data while requiring minimal suspension in packages such as scikit-learn.

The RF algorithm always includes the following three steps:

(1) Choose a training set. Use a random sample bootstrap method to obtain K training sets from the actual database (M properties), with each training size set to the same as the actual training set.

(2) Create an RF model. Create a regression tree with each set of bootstrap training sets to produce K decision trees to form a "forest". When looking at the growth of each tree, this method

does not select the best features such as the internal areas of the branches but the process of making the branch random selection of $m \leqslant M$ for all the features.

(3) Create a simple vote. Since the training process for each decision tree is independent, the training of random forests can work similarly, which greatly improves efficiency. RF can be created by assembling K-decision trees trained in the same way.

Decision trees are a popular method for a variety of machine learning activities. The study of trees is always subject to measurement and other characteristic modifications, is strong in the inclusion of non-essential elements, and produces experimental models. However, they are seldom accurate. In particular, trees that are planted too deep tend to learn very unusual patterns: they override their training sets, which means they have a low bias, but very diverse. Random forests are a way of measuring multiple deciduous trees, trained in different parts of the same training set, with the aim of minimizing diversity. That comes as a result of a slight increase in bias and a certain loss of definition, but significantly increases performance in the final model. Random Forests are like a concerted effort of tree-algorithm decision-making efforts. Taking the interaction of multiple trees thus improves the performance of a single random tree. Although not exactly the same, the forests provide the results of K-cross verification results.
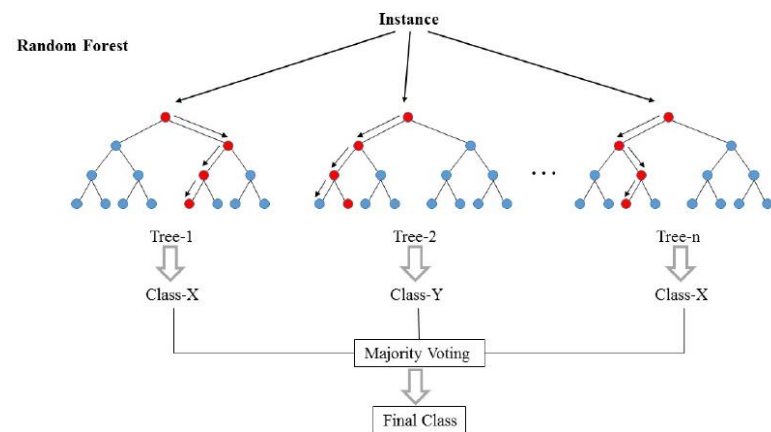
Firstly, updated dataset was downloaded from - https://github.com/owid/covid-19-data, dataset includes total cases, location, new cases, new deaths, and many more attributes for the second wave as Oxygen beds availability, population being vaccinated. For my research I only needed data for India so some sorting was done on the basis of location. Null values in the dataset was filled with average/mean of rest of the values which might create some inaccuracy.

df

|	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	total_cases_per_r
36785	IND	Asia	India	2020-01-30	1.0	1.0	NaN	NaN	NaN	NaN	
36786	IND	Asia	India	2020-01-31	1.0	0.0	NaN	NaN	NaN	NaN	
36787	IND	Asia	India	2020-02-01	1.0	0.0	NaN	NaN	NaN	NaN	
36788	IND	Asia	India	2020-02-02	2.0	1.0	NaN	NaN	NaN	NaN	
36789	IND	Asia	India	2020-02-03	3.0	1.0	NaN	NaN	NaN	NaN	
...	...	...	...	...	...	...	...	...	...	...	
37250	IND	Asia	India	2021-05-09	22662575.0	366494.0	391008.286	246116.0	3769.		
37251	IND	Asia	India	2021-05-10	22992517.0	329942.0	387097.714	249992.0	3876.		
37252	IND	Asia	India	2021-05-11	23340938.0	348421.0	382279.857	254197.0	4205.		
37253	IND	Asia	India	2021-05-12	23703665.0	362727.0	375179.286	258317.0	4120.		
37254	IND	Asia	India	2021-05-13	24046809.0	343144.0	365030.143	262317.0	4000.		

470 rows × 59 columns

Then comes the data analysing part in which many feature variables were tested with total cases as the main goal is to predict how many cases will be there on the particular time period. Features which showed no trend were dropped from the data frame at once and others were studied for some trend. Many features like new cases, new deaths, positive rate, new tests, stringency index, total deaths showed a similar trend in their respective graphs. Then dataset was broken into two sets-1: Training set. 2: Testing set.

```
plt.subplot(1,2,1)
plt.scatter(df['total_cases'],df['positive_rate'],alpha=0.3,c='green')
plt.xlabel('total_cases')
plt.ylabel('positive_rate')
plt.xticks([0,5000000,10000000,15000000,20000000,25000000],['0','5M','10M','15M
plt.title('6 Scatter Plot')

plt.subplot(1,2,2)
plt.plot(df['total_cases'],df['positive_rate'])
plt.xlabel('total_cases')
plt.ylabel('positive_rate')
plt.xticks([0,5000000,10000000,15000000,20000000,25000000],['0','5M','10M','15M
plt.title('6 Line Plot')

plt.tight_layout(pad=1.0)
plt.show()
```
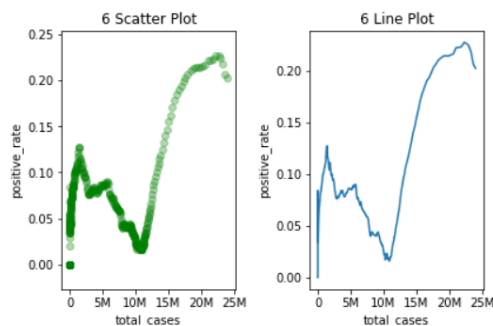

III. EXPERIMENT AND RESULT

Many data frames were made with a variety of feature variations and are trained in a training set and accuracy scores were observed. One with 99.81% accuracy was selected for prediction. The Pickle Library can convert complex material into a byte stream and can convert a byte stream into something with the same internal structure. Therefore, for our model to predict on any online platform, it has to be stored using the "dump" function. This file can then be converted to a random forest object with "load" function that can continue to generate the result by providing specific data values.

Splitting

```
from sklearn.model_selection import train_test_split
```

```
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.33,random_state=51)
```

Regression

```
from sklearn.linear_model import LinearRegression
```

```
reg=LinearRegression()
reg.fit(X_train,y_train)
print(reg.score(X_test,y_test))
```

0.9970033485621645

```
from sklearn.ensemble import RandomForestRegressor
```

```
forest=RandomForestRegressor(n_estimators=50)
forest.fit(X_train,y_train)
print(forest.score(X_test,y_test))
```

0.9981119442654744

At frontend, different input values have to be submitted according to this data were predicted.

Human behavior is the major factor. State and local governments, as well as individual people, differ in their response to the pandemic. Some follow COVID-19 precautions, such as physical distancing, hand-washing and mask-wearing. Others are not as prescriptive in requiring these measures or in restricting certain high risk activities. Also, places where people live or work closely together (multigenerational households, long term care facilities, prisons and some types of businesses) have also tended to see more spread of the coronavirus. Coronavirus

outbreaks at nursing homes and "superspreader" events — gatherings of people where one infected person or more transmits the virus to many others — continue to occur. There is a delay between a policy change such as reopening businesses or relaxing occupancy limits in a community and when the effects of this change show up in the COVID-19 data. An increase in the number of COVID-19 cases or hospitalizations will not be seen a week or even two weeks later. It seems to take much longer, perhaps as many as six to eight weeks, for effects of a policy or widespread behavior change to appear in the population-level data.

IV. CONCLUSION

The study concerns the spread of COVID-19 in India. A theoretical approach is used to develop an epidemic model to predict the disease. However, it's relevant to say that the long run estimation provided, is subject to certain system parameters and may vary based on the external inputs like lock-down measures, social-distancing, vaccine/drug development, rapid testing, etc.

Herd immunity is a public health term that refers to the fact that, when enough people in a community have immunity from a disease, the community is protected from outbreaks of that disease. Without a vaccine, most doctors and scientists agree that a herd immunity approach of letting the virus "take its course" is not acceptable. Letting the coronavirus circulate freely among the public would result in hundreds of thousands of deaths and millions more people left with lasting lung, heart, brain or kidney damage. As the majority of healthcare workers in India have already been completely vaccinated, mortality rates due to COVID in this high-risk group have plummeted to near zero. The focus of most state governments, therefore, is to mass vaccinate the entire population on a war footing. With high-risk population being prioritised, the mortality due to COVID would be expected to fall sharply. It would be desirable to authorise several other vaccines and procure as many doses as possible, for this purpose.

As the vaccine efficacy is around 80%, some people can still develop COVID post-vaccination but mortality is very rare. In Israel, where over 60% of the population has now received at least one dose of the vaccine, life has more or less returned to the pre-pandemic normal. It is heartening that less than 3% of ICU admissions for severe COVID in Israel were patients who had been fully vaccinated as against 75% who had not received even a single dose of the vaccine. There is considerable data to show that vaccination protects against mutant strains, although in variable degrees. The elimination of mortality due to COVID would effectively reduce this

virus to the status of a seasonal common cold. That is a good goal to have.

V. REFERENCE

[1] Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol. 2020 Apr;5(4):536-44.

[2] World Health Organization (WHO). WHO Statement regarding cluster of pneumonia cases in Wuhan, China 2020 [14 Januray 2020].

[3] National Institutes of Health. Coronavirus disease 2019 (COVID-19) treatment guidelines. 2021 [internet publication].

[4] World Health Organization. Prevention, identification and management of health worker infection in the context of COVID-19. 2020 [internet publication].

[5] Scientific Advisory Group for Emergencies. NERVTAG: update note on B.1.1.7 severity, 11 February 2021. 2021 [internet publication].

[6] *#IndiaFightsCorona COVID-19.* See www.mygov.in/covid-19/?cbps=1

[7] *India Scores High on Covid-19 Response Tracker Made by Oxford University.* See www.indiatoday.in/india/story/india-scores-high-on-covid-19-response-tracker-made-by-oxford-university-1665573-2020-04-10

[8] *Coronavirus Lockdown: How to Keep 130 Million Migrant Workers Afloat During COVID-19 Crisis.* See www.businesstoday.in/opinion/columns/coronavirus-lockdown-130-million-migrant-workers-afloat-covid-19-crisis-unorganised-sector/story/400806.html

[9] *How India Scaled Up PPE Production.* See https://timesofindia.indiatimes.com/india/0-to-12000/day-how-india-scaled-up-ppe-production/articleshow/75038389.cms.0

[10] Ministry of Health and Family Welfare, Government of India COVID-19 India dashboard. 2020. https://www.mohfw.gov.in/

[11] National Institute of Virology develops ELISA test to detect antibodies.https://www.thehindu.com/sci-tech/science/coronavirus-indias-national-institute-of-virology-develops-elisa-test-to-detect-antibodies/article31551219.ece

[12] DRDO *Covid-19 Sample Collection Kiosk (COVSACK).* Defence Research and Development Organisation - DRDO|GoI (2020). Available online at: https://drdo.gov.in/covid-19-sample-collection-kiosk-covsack

[13] asudha V. *Coronavirus Outbreak: Ayush Pushes "Traditional Cure," Med Council Backs Modern Drugs.* The Economic Times; (2020). Available online at: https://economictimes.indiatimes.com/news/politics-and-nation/ayush-pushes-traditional-cure-med-council-backs-modern-drugs/articleshow/74680699.cms

[14] https://clinmedjournals.org/articles/jide/journal-of-infectious-diseases-and-epidemiology-jide-6-147.php?jid=jide

[15] Vaccination drive in India https://www.nature.com/articles/s41541-021-00327-2