RESEARCH ARTICLE   **OPEN ACCESS**

COVID-19 Predictions outcome, Forecasting and Analysis

Mriga Khanna[1], Anmol Kukreja[2], Richa Singh[3]Dr.Pavan Kumar Shukla[4], Kunal Lala[5]

[1,2,3]*Student,*[4]*Professor,*[5]*Assistant Professor*
[1,2,3,4,5]*Department of Electronics & Communication EngineeringR.K.G.I.T, Ghaziabad, (U.P) India*

Abstract: Coronavirus (COVID19), first detected in December 2019 in Wuhan, China. The first case was Followed back on 17 November 2019 which was again declared an Epidemic in March 2020. This is an infectious disease caused by SARS-CoV-2. In almost every affected country, the number of infected patients and those who have died is high they have been increasing at an alarming rate. Since earlier forecasts can reduce the spread of the virus, it is highly desirable to have intelligent guessing and testing tools. The use of effective forecasting models can help government in implementing better strategies to prevent the spread of the virus. The proposed project uses 'Fbprophet' to predict the total number of deaths, cases found, aggregate number of validated cases and number of daily cases. Model made in Anaconda Distribution to get predicted numbers of cases to date.
Keywords: Machine Learning, COVID 19, visualization, pandas.

I. INTRODUCTION

COVID-19 has become a major global problem since World War II and the world's largest epidemic since the Spanish influenza of 1918-19. The epidemic has had a profound effect on people's lives and the country's economy [4]. Among the many questions related to infection, governments and individuals are most concerned about (i) when will COVID19 infection rate reach its peak? (ii) How long will the epidemic stop and (iii) What will be the total number of people who will eventually become infected? (iv) What will be the death toll? [4] These questions are of great concern to

India, a country with a large population and economic differences. The spread of the disease in India is much lower than in China, the USA and other European countries. India is under total closure from March 21, 2020 with the expert's belief that this could be detrimental to reducing the spread of Covid19 among its citizens [4]. As of April, 30 the number of COVID19 cases in India was 36669 and she died in 1229 as a result of Severe Acute Respiratory Syndrome (SARS). The total number of people found COVID19 in India is 140980 as of 4 June Imprisonment affects the poor and the migrant workers. Staying at home may not be possible immediately because many people may die of starvation and other diseases [4]. Media reports from around the world report on the problem and how it affects people's lives. A lot of research is being done at all levels to quickly gather data, develop mitigation tools and similar methods and applications. Therefore, policymakers and authorities want to have an overview of the current situation and to imagine how quickly it can spread.

This paper discusses the proposed prediction model of COVID19 that is spreading worldwide using machine learning that has been used in Anaconda distribution. The model steps are discussed in the subset section

II. OBJECTIVE

This research paper seeks to investigate the potential global impact of Novel Coronavirus (COVID-19) by predicting the prevalence of confirmed cases and the analysis of the number of deaths and acquisitions with the help of machine learning using Python. This paper introduces an objective way to predict the progress of cases, outlining the timeline of the forecasting process which has major implications for planning and decision-making.

III. METHODOLOGY
a. BUILDING OF DATASET:
The collection of data in a tabular form represents a dataset. We have collected the data of COVID-19 cases from the website www.kaggle.comwhich is updated on a daily basis by John Hopkins University. The following steps represent the hierarchical structure of program designed to derive data, visualize it and then make necessary predictions from it.
Step 1. import the necessary libraries for data manipulation and visualisation.
Step 2. Next up, we will import the different csv files and start manipulating the data.

```
In [36]: df_deaths.head()
```

```
Out[36]:
```

	Province/State	Country	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	...	5/31/20	6/1/20	6/2/20	6/3/20	6/4/20	6/5/20	6/6/20	6/7
0	NaN	Afghanistan	33.0000	65.0000	0	0	0	0	0	0	...	257	265	270	294	300	309	327	3
1	NaN	Albania	41.1533	20.1683	0	0	0	0	0	0	...	33	33	33	33	33	33	34	
2	NaN	Algeria	28.0339	1.6596	0	0	0	0	0	0	...	653	661	667	673	681	690	698	7
3	NaN	Andorra	42.5063	1.5218	0	0	0	0	0	0	...	51	51	51	51	51	51	51	
4	NaN	Angola	-11.2027	17.8739	0	0	0	0	0	0	...	4	4	4	4	4	4	4	

5 rows × 144 columns

```
In [67]: df.style.background_gradient(cmap='Reds')
```

2861	10/06/20	8:00 AM	Karnataka	0.000000		0.000000	2904	66	5921	0.000000	3251
2862	10/06/20	8:00 AM	Kerala	0.000000		0.000000	848	16	2096	0.000000	1232
2863	10/06/20	8:00 AM	Ladakh	0.000000		0.000000	52	1	108	0.000000	55
2864	10/06/20	8:00 AM	Madhya Pradesh	0.000000		0.000000	6729	420	9849	0.000000	2700
2865	10/06/20	8:00 AM	Maharashtra	0.000000		0.000000	42636	3289	90787	0.000000	44860
2866	10/06/20	8:00 AM	Manipur	0.000000		0.000000	61	0	304	0.000000	243
2867	10/06/20	8:00 AM	Meghalaya	0.000000		0.000000	13	1	43	0.000000	29
2868	10/06/20	8:00 AM	Mizoram	0.000000		0.000000	1	0	88	0.000000	87
2869	10/06/20	8:00 AM	Nagaland	0.000000		0.000000	10	0	127	0.000000	117
2870	10/06/20	8:00 AM	Odisha	0.000000		0.000000	2133	9	3140	0.000000	998
2871	10/06/20	8:00 AM	Puducherry	0.000000		0.000000	52	0	127	0.000000	75
2872	10/06/20	8:00 AM	Punjab	0.000000		0.000000	2167	55	2719	0.000000	497
2873	10/06/20	8:00 AM	Rajasthan	0.000000		0.000000	8328	255	11245	0.000000	2662

```
In [23]: dates = confirmed.keys()
         world_cases = []
         total_deaths = []
         mortality_rate = []
         total_recovered = []

         for i in dates:
             confirmed_sum = confirmed[i].sum()
             death_sum = deaths[i].sum()
             recovered_sum = recoveries[i].sum()
             world_cases.append(confirmed_sum)
             total_deaths.append(death_sum)
             mortality_rate.append(death_sum/confirmed_sum)
             total_recovered.append(recovered_sum)
```

So, we can see that the dataset provides the record till 15 December 2020, indicating the number of cases found in specific state in the country. Moreover, the dataset contains 2666 entries and 9 features. The above dataset indicate the highest no. of recovered cases, confirmed cases and death cases. Also if we do some analysis on dataset results are shown below :

```
In [24]: confirmed_sum

Out[24]: 67073728
```

```
In [25]: death_sum

Out[25]: 1536056
```

```
In [26]: recovered_sum

Out[26]: 43103827
```

In this we have imported the data from .csv file and passed the dates to perform date/time operation.

Now to look at the cases for India. Here df is used for the data-frames and some arguments are passed into the function to get the total infected, recovered and death cases along with the data present in the function.

```
In [43]: # Check for India's data
         df.query('Country=="India"').groupby("Date")[['Confirmed', 'Deaths', 'Recovered']].sum().reset_index()
```

```
Out[43]:
```

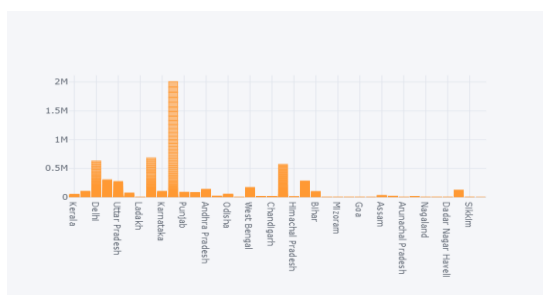|	Date	Confirmed	Deaths	Recovered
0	2020-01-22	0	0.0	0
1	2020-01-23	0	0.0	0
2	2020-01-24	0	0.0	0
3	2020-01-25	0	0.0	0
4	2020-01-26	0	0.0	0
...	...	...	...	...
136	2020-06-06	246622	6946.0	118695
137	2020-06-07	257486	7207.0	123848
138	2020-06-08	265928	7473.0	129095
139	2020-06-09	276146	7750.0	134670
140	2020-06-10	276583	7745.0	135206

141 rows × 4 columns

So, the above graph is plotted between number of cases in World vs India. The green color here indicates the total number of confirmed cases in all over the world and the blue colour indicates the no of confirmed cases in India.

We can see that the cases in all over the world are exponentially increasing as compared to India. At its early stage there is slightly increase in number of confirmed cases, the rate of recovery also increases as there is increase in number of confirmed cases. But the total number of death cases is slightly less as compared to confirmed and recovered cases. At its early trend the number of death cases are very less and then its slightly keep on increasing. But still the death cases are less as compared to number of confirmed cases which is a good sign.

Now we will be plotting the COVID 19 trend in India.

Advancement in Electronics & Communication Engineering (AECE-2020)
Raj Kumar Goel Institute of Technology (RKGIT), Ghaziabad, UP, India

Page | 33

The pace of novel coronavirus has quickened in the past few days. The above graph represents the state wise analysis of India. The rise and fall of the cases in different states can be analyzed from the graph. As per the analysis the states like Maharashtra, Gujarat and Tamil Nadu being highly infected by the virus. As single-handed Maharashtra only accounts for the 34% of the cases, which now has made the country amongst the top five caseload country.

Also the health centers have warned us about the five most affected states which are - Maharashtra, Tamil Nadu, Delhi, Gujarat, Uttar Pradesh. Maharashtra surpassing enormous no of cases in a single day whereas, Karnataka wraps more influenza like illness caseloads, Delhi in need of 90 thousand beds daily by July end, Gujarat with the caseload of 35 thousand in which 70% of state burden is from Ahmedabad.The petrified state in which we are currently living in has left everyone perplexed and paranoid giving Covid19 the stature of "the crisis of the century". The rise of this virus in India is now faster than our neighboring countries.

b. FORECASTING:
Forecasting involves collecting the information on historical data and using them to predict future observations. An important distinction in forecasting is that the future is completely unavailable and it must only be estimated from what had already happened. While using a classical statistic, the primary concern is the analysis of time series. Time series analysis provides a body of techniques to better understand a dataset. Following are the steps involved in forecasting the time series analysis:

1.Visualize the time series to analyse the trends prior to building any kind of time series model.
2.Stantionarize the series
3.Find the optimal parameters.
4.Build the model required.
5.Make predictions

Therefore, to make time series prediction we will using a library called Fbprophet. So Fbprophet is a library created by Facebook, written in python and it allows us to make time series analysis and make some predictions based on the data we have

accumulated over the days. So, we import it. The way prophet works is that it needs 2 columns:
1.DS which stands for date stamp.
2. The other is a variable 'y' which is trying to predict the cases.

We are predicting with 95% interval of confidence
Forecast for confirmed cases: -Firstly we will change the columns names in our confirmed data frame and the data inside the date column /ds is converted into date/time operation using pandas. Then we make future predictions for next 10 days. Look at the last five records shown below.

```
In [44]:  latest_confirmed

Out[44]:  0       47306
          1       42988
          2       88252
          3        7050
          4       15591
                   ...
          266     98038
          267        10
          268      2337
          269     17916
          270     10718
          Name: 12/6/20, Length: 271, dtype: int64
```
Fig 5

Predicting deaths: -Similarly we will predict the number of deaths across the world.So, we extract the data, we change the column names and we create a model and forecast it. The data below in the table and plots predict the death cases for the next 10 days.

```
In [45]:  latest_deaths

Out[45]:  0        1874
          1         905
          2        2516
          3          78
          4         354
                   ...
          266       828
          267         1
          268       639
          269       364
          270       291
          Name: 12/6/20, Length: 271, dtype: int64
```
Fig 6

Predicting recoveries: -For predicting the recoveries, we will perform the same procedure as done for confirmed cases and death cases and figuring out the model. So, the prediction for recoveries is also done for next 10 days Then forecasting it into the future to make visualisation easier. The data and plots are shown below:

```
In [46]:  latest_recoveries

Out[46]:  0       37685
          1       21617
          2       57146
          3        6238
          4        8338
                   ...
          251     72773
          252         8
          253      1549
          254     17173
          255      8880
          Name: 12/6/20, Length: 256, dtype: int64
```
Fig 7

IV. RESULTS

1. By doing predictions for total number of confirmed cases and forecasting the plots we can analyse there is a steady increase in number of cases from figure 8. The graph is rising exponentially. At its early trend it is increasing slowly but from mid-March2020, the curve has taken a sharp turn and its increasing gradually till June 2020

2.Similarly by doing predictions for total number of death cases and forecasting the plots we can analyse that there is a steady increases in the cases figure 9.Also by visualising the trend we can analyse that instead of having lower fatality rate, SARS-CoV-2 has caused thrice the total of deaths when compared to the combined statistics of deaths caused by both MERS and SARS-CoV.

3.Likely by doing the predictions for total number of recovered cases and forecasting the plots we can analyse that the number of recovery is also going to take a rapid increase figure 10 as there are more number of patients get introduced .If we analyse the trend, it looks like the number of patients who have recovered is likely to the active number of cases. So, if total number of confirmed cases increases then the recovery rate also increases.

V.CONCLUSIONS

The proposed methodology predicts the overall number of COVID19 infected cases, overall no of recovered cases, overall no of deaths all over the country. Weekly predictions have also been done for the confirmed, recovered, death cases. Based on these recent trends, the future trends have been predicted and the plots are visualized for the confirmed, recovery, death case, using machine learning. The methodology used here is having 95%accuracy in predicting the confirmed, deaths and recovered cases. The machine learning is useful in forecasting the impact of COVID-19 in different sectors which may help the government to implement the proper policies to overcome the economic crisis [6].Therefore, to empower the government and health care sector, it is necessary to analyse various forecasting and prediction tools. Moreover, the accuracy of prediction tools can be enhanced by the usage of advanced computing intelligent approaches such as ensemble method like bagging, stacking etc., application of optimization techniques, usage of artificial neural networks and higher order neural networks in the screening and prediction of COVID-19 which is considered as further scope of research [6].The public health officials and government should take different preventive measures to control the rapid increase of the COVID-19 [3]. Besides officials, the general public should keep social distancing and use

precautions to ensure their safety and control the disease from further spreading [3].

REFERENCES

[1]   Putra M, Kesavan MM, Brackney K, Hackney DN, Roosa MK. Forecasting the Impact of Coronavirus Disease During Delivery Hospitalization: An Aid for Resources Utilization. American Journal of Obstetrics & Gynecology MFM. 2020 Apr 25:100127.

[2]   Chakraborty T, Ghosh I. Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis. Chaos, Solitons & Fractals. 2020 Apr 30:109850.

[3]   Yousaf M, Zahir S, Riaz M, Hussain SM, Shah K. Statistical Analysis of Forecasting COVID-19 for Upcoming Month in Pakistan. Chaos, Solitons & Fractals. 2020 May 25:109926.

[4]   Parbat D. A Python based Support Vector Regression Model for prediction of Covid19 cases in India. Chaos, Solitons & Fractals. 2020 May 31:109942.

[5]   Panwar H, Gupta PK, Siddiqui MK, Morales-Menendez R, Singh V. Application of Deep Learning for Fast Detection of COVID-19 in X-Rays using nCOVnet. Chaos, Solitons & Fractals. 2020 May 28:109944.

[6]   Rekha Hanumanthu S. Role of Intelligent Computing in COVID-19 Prognosis: A State-of-the-Art Review. Chaos, Solitons & Fractals. 2020 May 29:109947.

[7]   Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. International journal of infectious diseases. 2020 Mar 4.

[8]   Lee H, Park SJ, Lee GR, Kim JE, Lee JH, Jung Y, Nam EW. The relationship between trends in COVID-19 prevalence and traffic levels in South Korea. International Journal of Infectious Diseases. 2020 Jul; 96:399.

[9]   Tuli S, Tuli S, Tuli R, Gill SS. Predicting the Growth and Trend of COVID-19 Pandemic using Machine Learning and Cloud Computing. Internet of Things. 2020 May 12:100222.

[10]   Hasan N. A Methodological Approach for Predicting COVID-19 Epidemic Using EEMD-ANN Hybrid Model. Internet of Things. 2020 May 28:100228.

[11]   Lansbury L, Lim B, Baskaran V, Lim WS. Co-infections in people with COVID-19: a systematic review and meta-analysis. Journal of Infection. 2020 May 27.

[12]   Chintalapudi N, Battineni G, Amenta F. COVID-19 disease outbreak forecasting of

registered and recovered cases after sixty-day lockdown in Italy: A data driven model approach. Journal of Microbiology, Immunology and Infection. 2020 Apr 13.

[13]   Wang L, Li J, Guo S, Xie N, Yao L, Cao Y, Day SW, Howard SC, Graff JC, Gu T, Ji J. Real-time estimation and prediction of mortality caused by COVID-19 with patient information-based algorithm. Science of the Total Environment. 2020 Apr 8:138394.

[14]   Duffey RB, Zio E. Prediction of CoVid-19 infection, transmission and recovery rates: a new analysis and global societal comparisons. Safety Science. 2020 May 28:104854.

[15]   Tagliazucchi E, Balenzuela P, Travizano M, Mindlin GB, Mininni PD. Lessons from being challengedby COVID-19. Chaos, Solitons & Fractals. 2020 May 23:109923