Assessment of Prediction Models of Confirmed, Recovered and Deceased cases due to COVID-19

P Rakshit, S Debnath, J Mistri and S Kumar

1Department of Computer Science & Engineering, JIS College of Engineering, Kalyani, Nadia, Pin-741235, West Bengal, India
2Data Scientist, Business Analyst & Quality Manager, Tata Consultancy Services Ltd, Kolkata, West Bengal, India
*Corresponding Author: pranati.rakshit@jiscollege.ac.in

Abstract. Pandemic relates to a situation where any disease starts spreading geographically and affects a entire country or the whole world. So when an epidemic becomes pandemic, it really a question of our survival. COVID-19 has become a pandemic as we all know and needs real and underneath research on that. The procession of death is uncountable still now. It can cause significant economic, social, and political disruption. So it’s very necessary to know the impact of it on originating venue so that we can analyze its potential and rate of spreads. So to do this we have applied here some Machine learning algorithm and concepts of regression for prediction. In this present work we have made prediction model of confirmed cases, Recovered and death cases using K-Nearest Neighbour regressor and Gradient Boosting Regressor. The model performance is very good in predicting all the cases. The R squared value is very near to 1.

1. Introduction
Covid 19 is a cursed gift to the nation, due to which a Pandemic situation arises in the whole world. Our life has been changed now a days because our Earth has been infected by virus, COVID-19 - a novel corona virus. A virus which is merely a thing between living and non living has become a super powerful enemy against humanity posing a challenge to their existence [1-2]. The whole human race is looking for a weapon to combat against this invisible enemy.

The pandemic caused by Covid (the Corona Virus Infectious Disease) is the most devastating pandemic the world has faced up till now. This virus is suspected to be originated from a small animal market in Wuhan, China, named as Wuhan virus at first. A number of patients admitted to the hospitals in Wuhan with pneumonia like symptoms but were not responding to the usual treatment and within a few days the virus is found to be extremely contagious and rapidly spreaded in Wuhan, Hubei province of China, in December 2019 [3-4]. Today (15-8-2020) it has spreaded among 213 countries all over the world and total number of cases are more than 21 million and with 763387 numbers of deaths.

After the genome sequencing of the respiratory tract samples of the infected patients, the virus is isolated as new novel-beta-coronavirus and later it is identified as Severe Acute Respiratory Syndrome type 2 coronavirus that is SARS-CoV-2 virus.

Before this SARS-COV-2 pandemic there were two other mentionable coronavirus pandemic or epidemic in the past [5]. One in November 2002 to March 2003 an epidemic was reported due to SARS-CoV-1 virus originated in Guangdong province of south-eastern China. Later it was reported...
that total 8096 people affected in 26 countries and total number of deaths where 774 by the end of May 2004.

The other mentionable case of respiratory tract infection epidemic was in Middle East countries Saudi Arabia Oman UAE in 2012 [6]. This was caused by another type of Corona virus which is named as Middle East Respiratory Syndrome Corona virus or MERS-CoV [7-9].

The global problem of the outbreak is being researched by scientists of different areas, a number methodology to analyze and predict the evolutions of this pandemic are being discussed in different forum [10-12]. Some mathematical model has been made to forecast some transmission dynamics [13-15]. In this paper we are going to make some prediction models which can predict the confirmed, recovered and deceased cases.

2. Methodology and Implementation

Here, we do prediction of effected or Confirmed cases, Recovered cases and Deceased cases. We have considered the total tenure of this epidemic or more specifically pandemic situation caused by COVID 19. Here we have considered the data of 247 days from 31 December, 2019 upto 31st August, 2020. The original dataset is shown in Table 1. It is arranged in total number of confirmed, recovered and deceased cases and specimen part of which is shown in table 2.

| Table 1. Original dataset |
|---------------------------|
| Sno | Date  | Time | State/UnionTerritory | ConfirmedIndianNational | ConfirmedForeignNational | Cured | Deaths | Confirmed |
|-----|-------|------|-----------------------|-------------------------|-------------------------|-------|--------|-----------|
| 1   | 30/01/20  | 6:00 PM | Kerala | 1 | 0 | 0 | 0 | 1 |
| 2   | 31/01/20  | 6:00 PM | Kerala | 1 | 0 | 0 | 0 | 1 |
| 3   | 01/02/20  | 6:00 PM | Kerala | 2 | 0 | 0 | 0 | 2 |
| 4   | 02/02/20  | 6:00 PM | Kerala | 3 | 0 | 0 | 0 | 3 |
| 5   | 03/02/20  | 6:00 PM | Kerala | 3 | 0 | 0 | 0 | 3 |

| Table 2. Prepared dataset to be used for training and testing |
|-----------------------------|
| Sl.No. | Days | Confirm | Recover | Death |
|------|------|--------|---------|-------|
| 0    | 1    | 1      | 0       | 0     |
| 1    | 2    | 1      | 0       | 0     |
| 2    | 3    | 2      | 0       | 0     |
| 3    | 4    | 3      | 0       | 0     |
| 4    | 5    | 3      | 0       | 0     |
Figure 1 is explaining how the Confirmed, Recovered and Deceased cases can be represented as per the pandemic days are going. Figure 2 represents the correlation among different attributes of the dataset. Figure 3 depicts the ground truth or original confirmed cases of COVID-19. Figure 4 depicts the predicted situation of confirmed cases using K-Nearest Neighbour Regressor. Figure 5 depicts the predicted situation of confirmed cases using Gradient Boosting Regressor.
Confirmed Cases:

Figure 3. Confirmed: Original data or Ground truth

Figure 4. Confirmed: Prediction using K-Nearest Neighbour Regressor
Figure 5. Confirmed: Prediction using Gradient Boosting Regressor

Figure 6, Figure 7 and Figure 8 are showing the ground truth, prediction using K-Nearest Neighbour Regressor and Gradient Boosting Regressor respectively of the recovered cases. Figure 9, Figure 10 and Figure 11 are showing the ground truth, prediction using K-Nearest Neighbour Regressor and Gradient Boosting Regressor respectively of the deceased cases.

**Recovered Cases:**

Figure 6. Recovered: Original data or Ground truth
Figure 7. Recovered: Prediction using K-Nearest Neighbour Regressor

Figure 8. Recovered: Prediction using Gradient Boosting Regressor
Deceased Cases:

Figure 9. Deceased: Original data or Ground truth

Figure 10. Deceased: Prediction using K-Nearest Neighbour Regressor
3. Result and Discussion

To measure the performance of our model we have calculated some relevant parameter like MSE and R-Squared metrics.

MSE (Mean Squared Error) is the average of squared differences between prediction and actual observation.

R-squared (Coefficient of determination) represents the coefficient of how well the values fit compared to the original values. The value from 0 to 1 interpreted as percentages. The higher the value is, the better the model is.

The above metrics can be expressed,

\[ R^2 = 1 - \frac{\sum_{x=1}^{N}(y_x - \bar{y})^2}{\sum_{x=1}^{N}(y_x - \bar{y})^2} \]

\( y_x = \text{ActualValue} \)
\( \hat{y} = \text{predictedValue} \)
\( \bar{y} = \text{meanvalue} \)
The above mentioned tables are showing the R squared value and MSE value for all the cases using K-Nearest Neighbour Regressor and Gradient Boosting Regressor. The value of R squared measures are very good which are 0.99 for all the cases. The MSE values are also quite good as in this present study we have considered a long total tenure of this pandemic.

4. Conclusion
In this present work we have made prediction model of Confirmed cases, Recovered and Deceased or death cases using K-Nearest Neighbour Regressor and Gradient Boosting Regressor. The model performance is very good in predicting all the cases. The R squared value is very near to 1. In future some mathematical model Susceptible-Exposed-Infectious-Recovered (SEIR) based forecasting can be done and also using some different machine learning model we will do the prediction.

References
[1] W.H.O, Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel Coronavirus (2019-nCoV). 2020. https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-Coronavirus-(2019-ncov). Accessed 18 Aug 2020.
[2] Coronavirus disease (COVID-19) Weekly Epidemiological Update and Weekly Operational Update. Situation reports May 2020. Situation report-113. https://www.who.int/emergencies/diseases/novel-Coronavirus-2019/situation-reports. Accessed 18 Aug 2020.
[3] WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020, https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020. Accessed 18 Aug 2020.
[4] Todt, D.; Pfaender, S.; Steinmann, E.: Persistence of Coronaviruses on inanimate surfaces and its inactivation with biocidal agents. J Hospital Infection. 105, 587 (2020)
[5] Rothe, C.; Schunk, M.; Sothmann, P.; Bretzel, G.; Froschel, G.; Wallrauch, C.; et al.: Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. N Engl J Med. 382, 970-971 (2020)
[6] Fang, Y.; Nie, Y.; Penny, M.: Transmission dynamics of the covid-19 outbreak and effectiveness of government interventions. A data-driven analysis. Journal Medical Virology. 92, 645-659 (2020)
[7] Kim, Y.; Lee, S.; Chu, C.; Choe, S.; Hong, S.; Shin, Y.: The Characteristics of Middle Eastern Respiratory Syndrome Coronavirus Transmission Dynamics in South Korea. Osong Public Health and Research Perspectives. 7, 49-55 (2016)
[8] Hethcote, H. W.: The mathematics of infectious diseases. SIAM Review. 42, 599-653 (2000)
[9] Hou, C.; Chen, J.; Zhou, Y.; Hua, L.; Yuan, J.; He, S.; et al.: The effectiveness of quarantine of Wuhan city against the Corona Virus Disease 2019 (COVID-19). A well-mixed SEIR model analysis. J Medical Virology. 92, 841-848 (2020)
[10] Nishiura, H.; Jung, S.; Linton, N. M.; Kinoshita, R.; Yang, Y.; Hayashi, K.; et al.: The extent of transmission of novel Coronavirus in Wuhan, China. J. Clin. Med. 9, 330 (2020)
[11] Batista, M.: Estimation of the final size of the Coronavirus epidemic by the logistic model. Slovenia. University of Ljubljana. (2020)
[12] Li, L.; Yang, Z.; Dang, Z.; Meng, C.; Huang, J.; Meng, H.: Propagation analysis and prediction of the COVID-19. Infectious Disease Modelling. 5, 282-292 (2020)
[13] Huang, Y.; Yang, L.; Dai, H.; Tian, F.; Chen, K.: Epidemic situation and forecasting of COVID-19 in and outside China. Bull World Health Organ. 10, (2020)
[14] Petropoulos, F.; Makridakis, S.: Forecasting the novel Coronavirus COVID-19. PloS one. 15, e0231236 (2020)
[15] López, L.; Rodo, X.: A modified SEIR model to predict the COVID-19 outbreak in Spain and Italy. simulating control scenarios and multi-scale epidemics. SSRN 3576802 (2020)