Analysis and forecast of COVID-19 spreading in India using Nonlinear curve fitting model with machine learning

Pawan Pandey (pandey.pawan.cs@gmail.com)
Raj Kumar Goel Institute of Technology
Vikas Katoch
Raj Kumar Goel Institute of Technology
Pawan Kumar
Raj Kumar Goel Institute of Technology

Research Article

Keywords: COVID-19, Nonlinear curve fitting model

Posted Date: October 7th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-87147/v1

Abstract

In this paper, an analysis and forecasting of Indian COVID-19 data is discussed by using scipy optimize curve fitting model of machine learning. We demonstrates the month wise analysis of coming cases, daily recovered cases, death cases and test cases conducted by the Government of India, of COVID-19 from 01st March 2020 to 02nd August 2020, and also forecast for the new cases, recover cases & death cases from 03rd August 2020 to 01st November 2020. Our study show that the total numbers of affected persons due to COVID-19 up to 01st November 2020 will be total cases 13,690,491, recover cases 10,499,593 and death of 129,271.

Introduction

The World Health Organization (WHO) declared COVID-19 (scientifically referred to as the severe acute respiratory syndrome–coronavirus 2 or SARS-CoV-2) a pandemic on 11 March 2020 **[1].** This virus had already spread from China to other Asian countries, Europe and the United States. As of 5th July 2020, cases have been identified in 188 countries or regions **[2].** It is a respiratory disorder in human having cough, fever and breath problems. The global response to the COVID-19 pandemic has led to a sudden reduction of both greenhouse gas (GHG) emissions and air pollutants **[3-5].** This has led to unprecedented enforced and voluntary restrictions on travel and work. On 31st December 2019, China reported to WHO, some persons are infected from pneumonia (caused by a novel coronavirus, currently known as 2019-nCoV) in Wuhan City **[6].** Analysis of mobility data from Google **[7]** and Apple **[8]** shows that mobility declined by 10% or more during April 2020 in all but one of the 125 nations tracked. The mobility declined by 80% in five or more nations.

The first case of COVID-19 in India is found on 30th January 2020, a female student studying at Wuhan city, she belongs to Kerala **[9].** After finding more cases, Indian Government has started the lockdown process with many phases from 25th March 2020, so the preventive methods are used during this lockdown to protect the people from COVID-19, such as wearing the mask, hands sanitizing, frequent hand washing, restrictions in travelling, avoid social gathering, staying at home etc. **[10].** Indian Government also ordered to close all the schools, colleges, markets and cinema halls during this period. People can only move out in emergency conditions by taking the permission from local authorities. India has very slow growth in the initial stage of COVID-19 pandemic but it increases exponentially later on **[11, 12].**

There are many mathematical models and machine learning models given for analysis and prediction of COVID-19 pandemic situation. The spatial distribution and region wise spreading of COVID-19 prediction across the India is given by using Geospatial Approach with the help of GIS Software for distribution and trend analysis till 11th April 2020 **[13].** SIER and regression models are used for forecasting for next two weeks on the basis of analysis collected by Johns Hopkins University from 30th January 2020 to 30th March 2020 **[14].** The RMSLE calculates the error rate of 1.52 for SEIR model and 1.75 for regression model for above analysis. The time series analysis based on ARIMA are also used for forecast. The time series based study indicates that the number of cases increase exponentially **[15].** The linear regression model with machine learning is also used for forecasting and this study used the linear regression, multilayer perceptron and vector autoregression methods for analysis and prediction on the COVID-19 Kaggle dataset **[16].** The containment model also used for COVID-19 in India, with prediction for reduction the number of upcoming cases **[17].**

There are many more research papers are available not only for Indian but also describes the covid 19 pandemic situations of China, Italy, France and United States, which can be helpful for planning and decision making

[18,19,20].

In this paper, a comprehensive nonlinear curve fitting model for analysis and forecast of COVID-19 in India is proposed. This study is divided into two parts; (1) we analyzed the new cases; recover cases, death cases and test cases on daily basis (2) we forecasted the values of news cases, recover cases and death cases weekly by using the nonlinear curve fitting model, The python, Pandas and Scipy optimize curve fitting model are used in our computational work **[21].**

Methodology

The nonlinear regression is a powerful technique to fix a broad range of values in nonlinear manner. The nonlinear regression determines the values of parameters that minimize the sum of squares of the distances of the data points (least square method) to the curve. Generally this method is used when experimental values are Gaussian in nature. The nonlinear regression procedure adjusts these values and produce new values to make curve fitted **.** We use exponential function to fit our data in this model **[22].**

$$Y_n = \alpha * \exp(\beta * x_n), \qquad (1)$$

where α and β are constants and have the values for new cases: 1741.833, 0.233, for recover cases: 854.028, 0.250 and for death cases 163.097, 0.155 respectively. These values are calculated by Scipy optimize curve fitting model of machine learning based on number of weeks (x_n) with respect to previous values available. The function may be defined as **[23]**

scipy.optimize.curve_fit(f, xdata, ydata, p0=None, sigma=None, absolute_sigma=False, check_finite=True, bounds=-inf, inf, method=None, jac=None, **kwargs),

where ‘f’ is the model function, ‘xdata’ is independent array object, ‘ydata’ is dependent array object, ‘p0’ is initial guess for the parameters, ‘sigma’ determines uncertainty in ‘ydata’, ‘absolute_sigma’ is a Boolean value to sense and the estimated parameter covariance, ‘check_finite’ to check the input array for value error etc. , ‘bounds’ represents lower and upper bound on parameters, ‘method’ to used for optimization (least square), ‘jac’ computes the Jacobean matrix of the function with respect to parameters as a dense array like structure, ‘kwargs’ keyword arguments passed to optimize method. This function returns ‘popt’ an array of optimal values for the parameters, ‘pcov’ provides variance of the parameter estimate in 2-dimensional array. Following function used to perform the curve fitting:

popt, pcov = curve_fit(func, xData, yData, initialGuess)

initial Guess = [1.0, 1.0]

Analysis And Forecasting

We have analyzed four points of COVID-19, new cases; recover cases, death cases and test cases daily basis as shown in Figure 1-5. The dataset created by Max Roser, Director, “Our World in Data” and his team members used for COVID-19 new cases, death cases & test cases analysis and forecasting **[11].** For recover cases analysis & forecasting we have used dataset created by Johns Hopkins University Center for Systems Science and Engineering [**12].**

The analysis for date wise new cases are shown in Fig.(1a), (2a), (3a), (4a), (5a) for the month of March, April, May, June and July. Similarly date wise recovered cases are shown in Fig.(1b), (2b), (3b), (4b), (5b) for these months. Fig. (1c), (2c), (3c), (4c), (5c) has date wise death cases and in Fig.(1d), (2d), (3d), (4d), (5d) shows date wise test cases performed by Indian Government.

In month of March, India has 1,248 total new cases, 120 recover cases, 32 death cases and 29,663 tests are performed but in April, 31,799 total new cases, 1,845 recover cases, 1,042 death cases and 795,313 test cases. In month of May, 149,093 total new cases, 82,784 recover cases, 4,090 death cases and 2,906,826 tests are performed with continuous growth. In month of June, 384,697 total new cases, 256,060 recover cases, 11,729 death cases with the highest number of death cases 2,003 on 17th June 2020 and 4,871,627 tests are performed and in month of July, 1,072,030 total new cases, 746,462 recover cases, 18,854 death cases and 10,224,316 tests are performed.

We analyzed the data on week's basis and made the prediction for new cases, recovers cases and death cases as shown in Figs. (6a), (6b) and (6c) respectively. Using the nonlinear curve fitting model these figures show that the forecasting for new cases, recovers cases and death cases are increasing exponentially by using nonlinear curve fitting model. The actual values are represented by the blue dots and predicted values are represented by red line. After inserting the corresponding values in curve fitting model, for the prediction of new cases and recover cases the value of x_n is increased by 0.5 and for death cases it is increased by 0.25.

The calculated values of new cases, recovers cases & death cases are shown in table 1.

week no	date		weekly counting			total counting to date		
	from	to	new cases	recover cases	death cases	total new cases	total recover cases	total death cases
Actual Values								
1-14	3/1/2020	5/31/2020				182,140	91,849	5,164
15	6/1/2020	6/7/2020	64,485	31,996	1,765	246,625	123,845	6,929
16	6/8/2020	6/14/2020	74,294	45,950	2,266	320,919	169,795	9,195
17	6/15/2020	6/21/2020	89,539	67,398	4,059	410,458	237,193	13,254
18	6/22/2020	6/28/2020	118,398	84,527	2,841	528,856	321,720	16,095
19	6/29/2020	7/5/2020	144,306	102,710	3,173	673,162	424,430	19,268
20	7/6/2020	7/12/2020	176,388	129,038	3,406	849,550	553,468	22,674
21	7/13/2020	7/19/2020	228,065	146,616	4,142	1,077,615	700,084	26,816
22	7/20/2020	7/26/2020	307,904	217,481	5,247	1,385,519	917,565	32,063
23	7/27/2020	8/2/2020	365,201	268,635	5,301	1,750,720	1,186,200	37,364
Predicted Values								
24	8/3/2020	8/9/2020	415,897	304,055	5,545	2,166,617	1,490,255	42,909
25	8/10/2020	8/16/2020	467,284	344,539	5,764	2,633,902	1,834,795	48,672
26	8/17/2020	8/23/2020	525,021	390,414	5,991	3,158,922	2,225,209	54,664
27	8/24/2020	8/30/2020	589,891	442,397	6,228	3,748,814	2,667,606	60,892
28	8/31/2020	9/6/2020	662,777	501,302	6,474	4,411,590	3,168,908	67,366
29	9/7/2020	9/13/2020	744,668	568,050	6,730	5,156,258	3,736,958	74,097
30	9/14/2020	9/20/2020	836,677	643,685	6,996	5,992,935	4,380,642	81,093
31	9/21/2020	9/27/2020	940,055	729,390	7,272	6,932,989	5,110,033	88,365
32	9/28/2020	10/4/2020	1,056,205	826,507	7,560	7,989,195	5,936,540	95,925
33	10/5/2020	10/11/2020	1,186,708	936,555	7,858	9,175,902	6,873,095	103,783
34	10/12/2020	10/18/2020	1,333,334	1,061,256	8,169	10,509,237	7,934,352	111,952
35	10/19/2020	10/25/2020	1,498,078	1,202,561	8,492	12,007,314	9,136,913	120,444
36	10/26/2020	11/1/2020	1,683,176	1,362,680	8,827	13,690,491	10,499,593	129,271

Table 1 Indian COVID 19 forecasting from 03rd August 2020 to 01st November 2020 (weekwise)

Conclusion

In this paper, by analyzed the existing COVID-19 data's of 23 weeks (5 months, from 01st March 2020 to 02nd August 2020) and a corresponding model is established, and then the prediction is done for next 13 weeks (3 months, from 03rd August 2020 to 01st November 2020) by using nonlinear curve fitting model. We predicted the total numbers of new cases; recover cases and death cases till 01st November 2020, week wise. Our analysis shows that there will be 13,690,491, COVID-19 cases with the recovery of 10,499,593 cases and the death of 129,271 till 01st November 2020 in India, with the current conditions and having approximate (per week) frequency of 1,683,176 new cases found, 1,362,680 recovery rate and death rate of 8,827 persons. These figures are only mathematical values which are estimated by doing some calculations. But these values can be reduced by using some preventive measures against the COVID-19 disease, such as performing the continuous large number of COVID-19 tests, maintain the social distance to each other, proper wearing the mask etc. The increase in these values will also be stopped when some effective medicine or vaccine of COVID-19 available for use to the effected persons or as needed.

Declarations

Compliance with Ethical Standards

Funding: There are no financial conflicts of interest to disclose.

Conflict of Interest: The authors declare that they have no conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Lixiang L., Zihang Y., Zhongkai D., Cui M., Jingze H., Haotian M., Deyu W., Guanhua C., Jiaxuan Z., Haipeng P., Yiming S., "Propagation analysis and prediction of the COVID-19", "Infectious Disease Modelling", Vol. 5, March 2020.
2. COVID-19 Map (Johns Hopkins Coronavirus Resource Center, accessed 5 July 2020); https://coronavirus.jhu.edu/map.html.
3. Liu, Z. et al. COVID-19 causes record decline in global CO2 emissions. Preprint at http://arxiv.org/abs/2004.13614 (2020).
4. Le Quéré, C. et al. Temporary reduction in daily global CO2 emissions during the COVID-19 forced confinement. Nat. Clim. Change 10, 647–653 (2020).
5. COVID-19 Air Quality Worldwide Dataset (The World Air Quality Project, accessed 5 July 2020); https://aqicn.org/data-platform/covid19/.
6. Phelan A. L., Katz R., Gostin L.O., "The novel coronavirus originating in Wuhan, China: challenges for global health governance", "JAMA", Feb 2020.
7. Google LLC Community Mobility Reports (Google, accessed 5 July 2020); https://www.google.com/covid19/mobility/.
8. Apple LLC Mobility Trends Reports (Apple, accessed 5 July 2020); https://www.apple.com/covid19/mobility.
9. Gupta, R., & Pal, S. K., "Trend Analysis and Forecasting of COVID-19 outbreak in India", med. Rxiv, March 2020 (https://www.medrxiv.org/content/10.1101/2020.03.26.20044511v1).

10. Economic Times, “Lockdown 1.0: Some success in flattening the curve”, 2020. (https://economictimes.indiatimes.com/news/politics-and-nation/lockdown-1-0-some-success-in-flattening-the-curve/articleshow/75136108.cms?from=mdr).
11. Hannah Ritchie, Esteban Ortiz-Ospina, Diana Beltekian, Edouard Mathieu, Joe Hasell, Bobbie Macdonald, Charlie Giattino, and Max Roser, “Mortality Risk of COVID-19”, “Our World in Data”. (https://ourworldindata.org/mortality-risk-covid).
12. Johns Hopkins University Center for Systems Science and Engineering, Dataset (https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases).
13. Bagyaraj Murugesan, Shankar Karuppannan, Alemayehu Tenaw Mengistie, Muthukumarasamy Ranganatha3, Gnanachandrasamy Gopalakrishnan, “Distribution and Trend Analysis of COVID-19 in India: Geospatial Approach”, “Journal of Geographical Studies”, April 2020.
14. Gaurav Pandey, Poonam Chaudhary, Rajan Gupta, Saibal Pal, “SEIR and Regression Model based COVID-19 outbreak predictions in India”, arXiv.org (https://arxiv.org/abs/2004.00958).
15. Hiteshi Tandon, Prabhat Ranjan, Tanmoy Chakraborty, Vandana Suhag, “Coronavirus (COVID-19): ARIMA based time-series analysis to forecast near future”, arXiv.org (https://arxiv.org/abs/2004.07859).
16. Sujath, Jyotir Moy Chatterjee, Aboul Ella Hassanien, “A machine learning forecasting model for COVID-19 pandemic in India”, “Stochastic Environmental Research and Risk Assessment”, May, 2020.
17. Fotios Petropoulos, Spyros Makridakis, “Forecasting the novel coronavirus COVID-19”, “PLOS ONE”, March, 2020.
18. Roosa, Y.Lee, R.Luo, A.Kirpich, R.Rothenberg, J.M.Hyman, P.Yan, G.Chowell, “Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020”, “Infectious Disease Modelling”, 2020.
19. DuccioFanellia, FrancescoPiazza, “Analysis and forecast of COVID-19 spreading in China, Italy and France”, “Chaos, Solitons & Fractals”, Vol. 134, May 2020.
20. Xiaoling Yuan, Jie Xu, Sabiha Hussain, He Wang, Nan Gao, and Lanjing Zhang2, “Trends and Prediction in Daily New Cases and Deaths of COVID-19 in the United States: An Internet Search-Interest Based Model”, “Exploratory Research and Hypothesis in Medicine”, April 2020.
21. Scipy Optimize curve fit, (https://github.com/scipy/scipy/blob/v1.5.1/scipy/optimize/minpack.py#L532-L834)
22. HJ Motulsky, LA Ransnas, “Fitting curves to data using nonlinear regression: a practical and nonmathematical review”, “The FASEB journal, wiley.com”, 198
23. H Motulsky, A Christopoulos, “Fitting models to biological data using linear and nonlinear regression: A practical guide to curve fitting”, “Oxford University Press”, May, 2004.

Figures

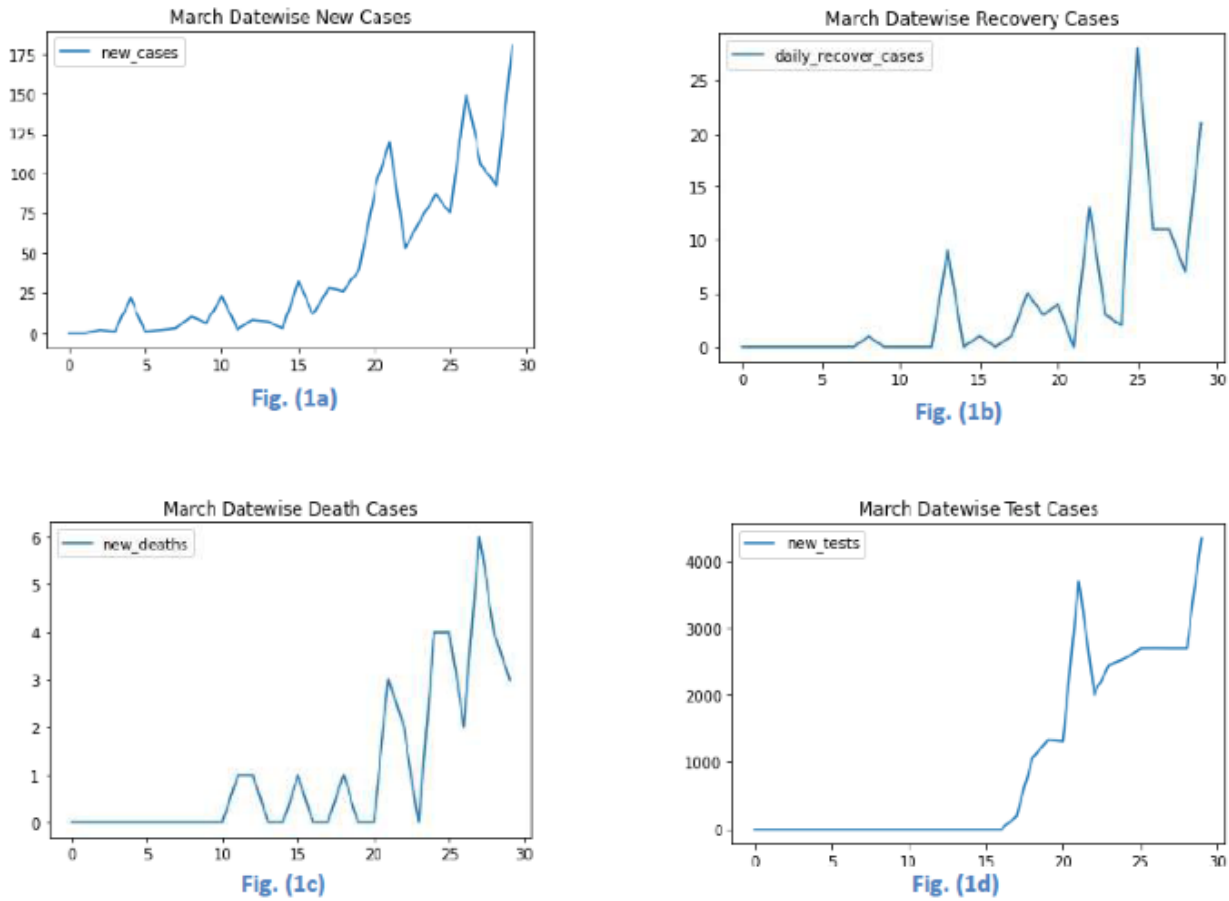

Figure 1

Indian COVID 19 Data Analysis of March

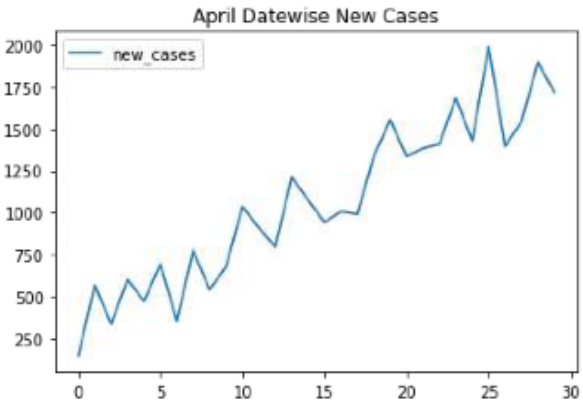

Fig. (2a)

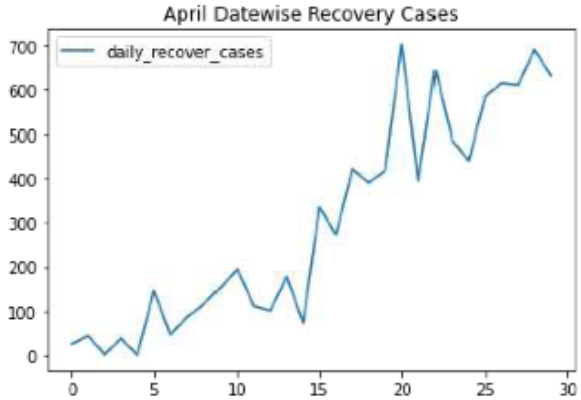

Fig. (2b)

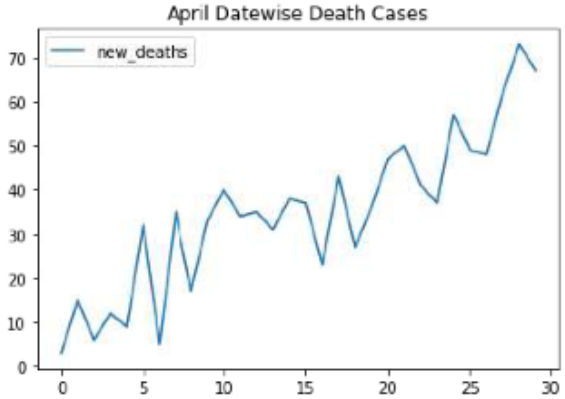

Fig. (2c)

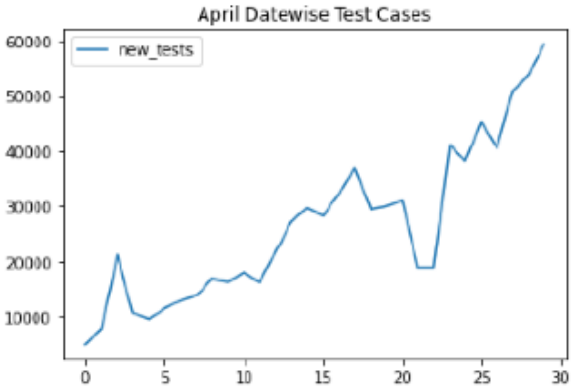

Fig. (2d)

Figure 2

Indian COVID 19 Data Analysis of April

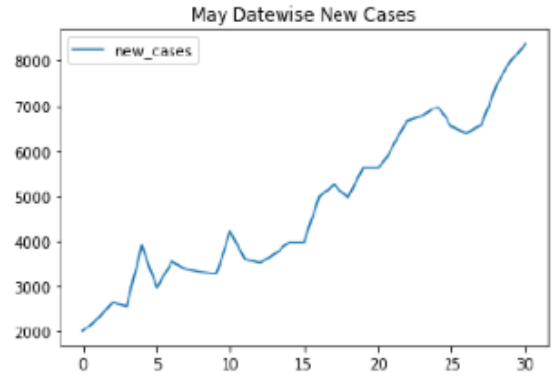

Fig. (3a)

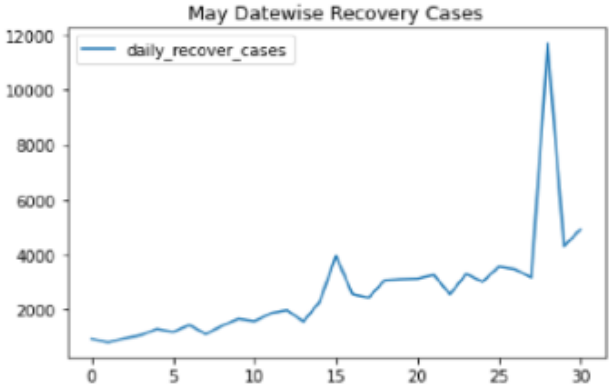

Fig. (3b)

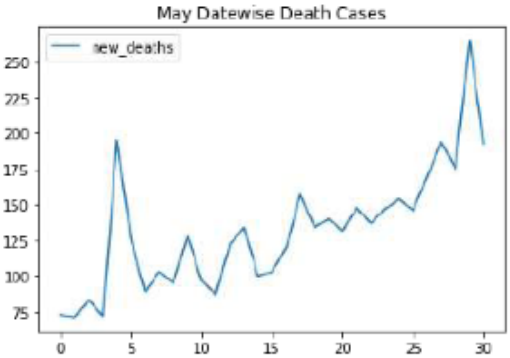

Fig. (3c)

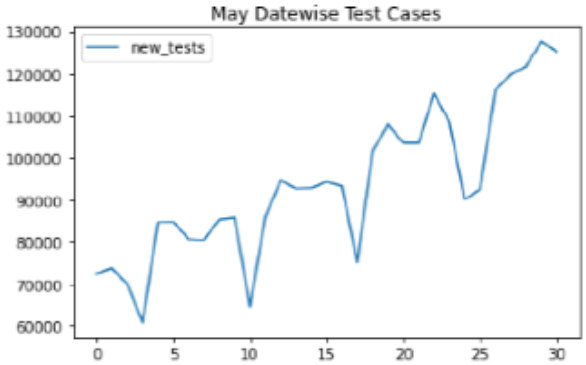

Fig. (3d)

Figure 3

Indian COVID 19 Data Analysis of May

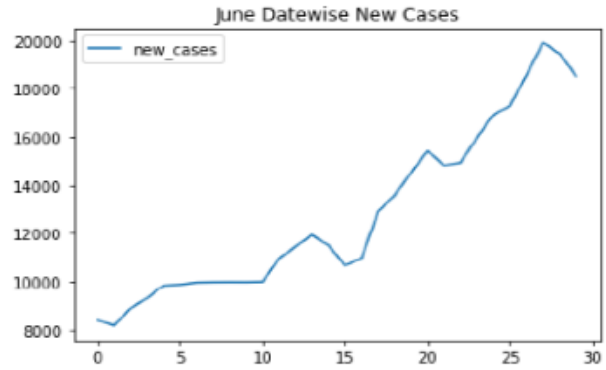

Fig. (4a)

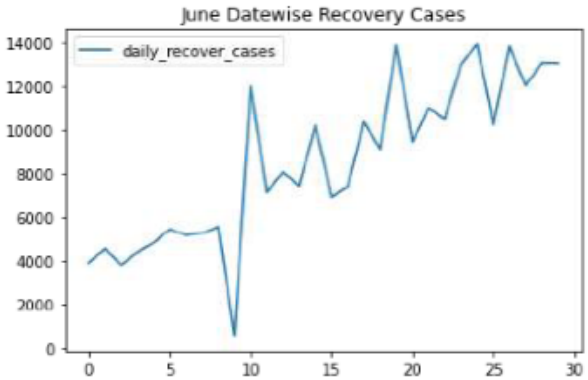

Fig. (4b)

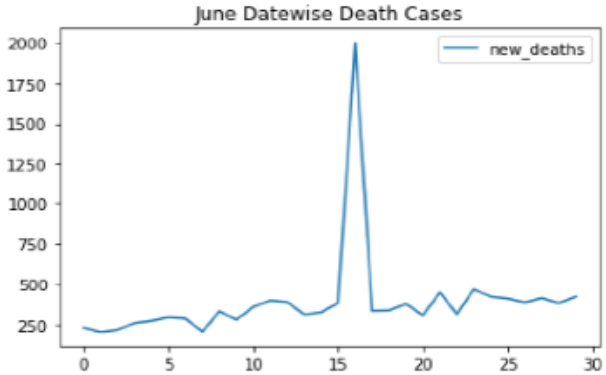

Fig. (4c)

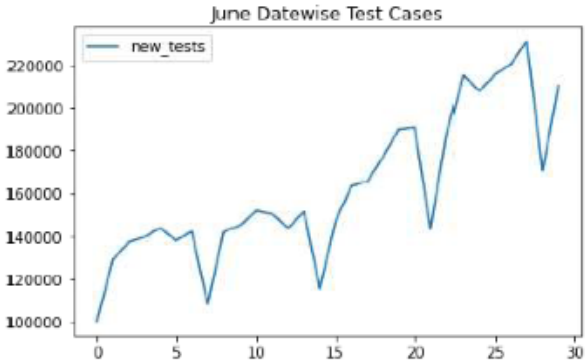

Fig. (4d)

Figure 4

Indian COVID 19 Data Analysis of June

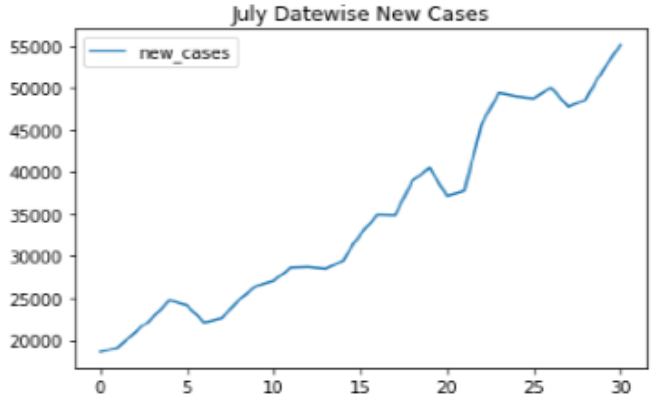

Fig. (5a)

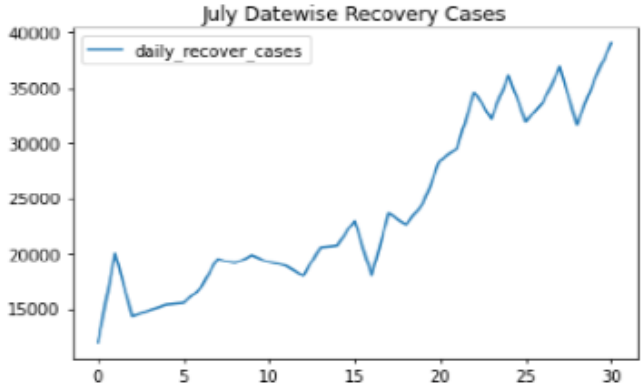

Fig. (5b)

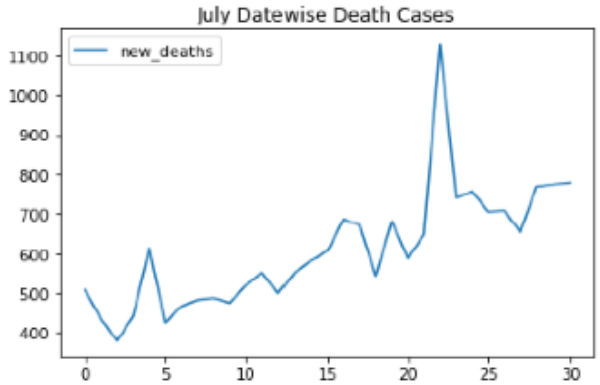

Fig. (5c)

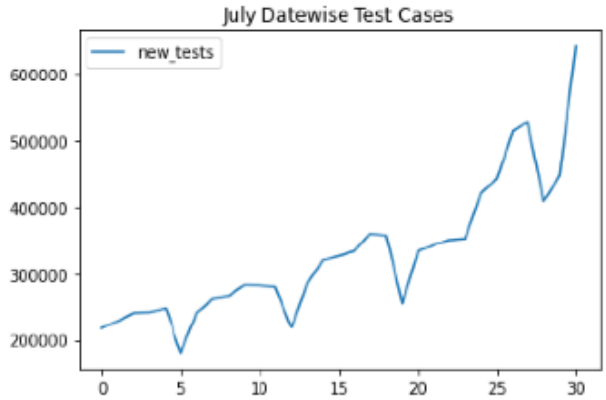

Fig. (5d)

Figure 5

Indian COVID 19 Data Analysis of July

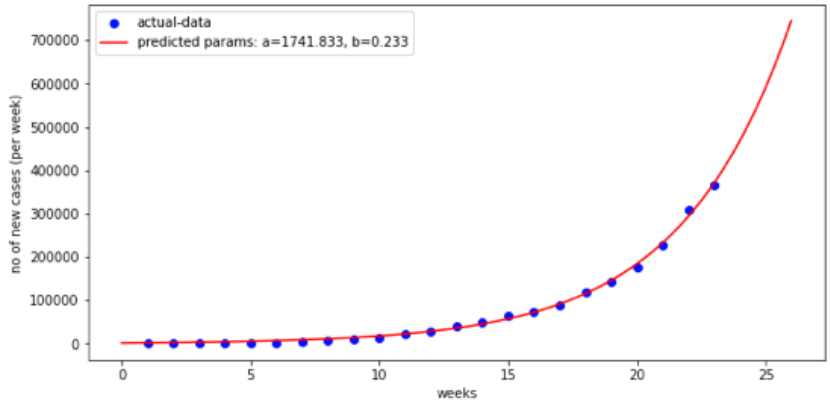

Fig. (6a) Indian COVID 19 forecasting of new cases (weekwise)

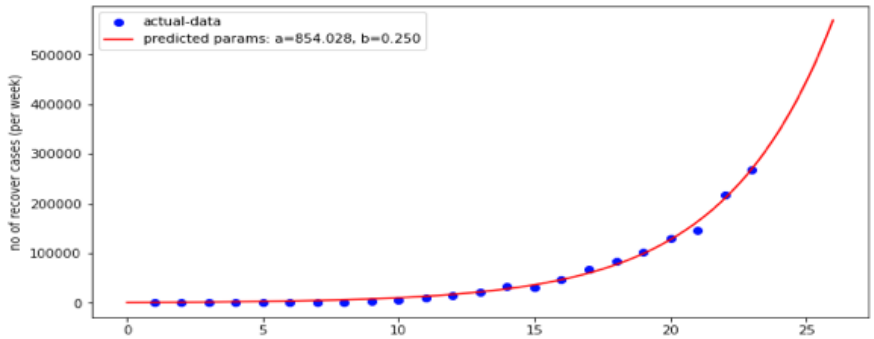

Fig. (6b) Indian COVID 19 forecasting of recover cases (weekwise)

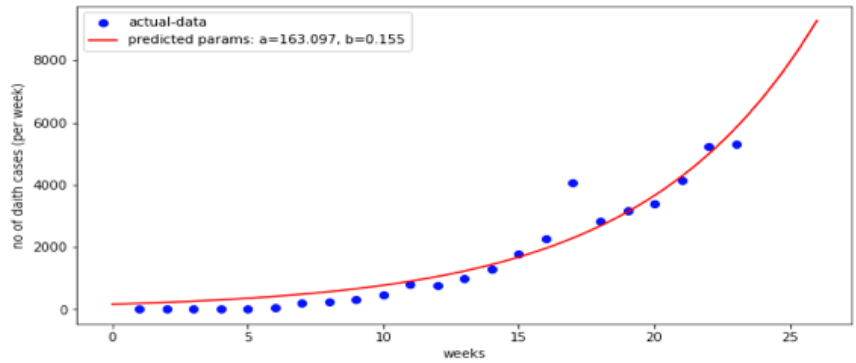

Fig. (6c) Indian COVID 19 forecasting of death cases (weekwise)

Figure 6

(6a) Indian COVID 19 forecasting of new cases (weekwise). (6b) Indian COVID 19 forecasting of recover cases (weekwise). (6c) Indian COVID 19 forecasting of death cases (weekwise)